Design and Implementation of an Optical Inspection System for Analyzing D-Periods in Collagen Fibril Images

M. Zervakis and M. Tzaphlidou

Abstract— **This paper considers an approach for analyzing fibrillar collagen structures based on digital signal processing. It focuses on the quantitative comparison between collagen structural data (electron-optical data) and chemical data. Theoretical models in the form of sequence-generated histograms are used as reference for extracting and analyzing the structural unit in images from collagen fibrils. In this respect, collagen provides a valuable model system for studying the chemical basis of ultrastructure and the mechanisms of various treatments on a protein, as well as detecting the alterations in its structure produced by a disorder. The algorithms developed in this study can be applied to any fibrous proteins. Some application examples are presented.**

I. Introduction and Problem Definition

One of the most important components of connective tissues is collagen, which is responsible for structural integrity of tissues and for providing a physical arena for important cell biological processes. The debilitating human diseases in which abnormalities of collagen are an essential and characteristic feature have their origins in the failure of connective tissues to perform their functions properly.

The most striking feature of the long fibrils, as seen in the electron microscope, is the regular transverse banding with an axial periodicity D. D-periodicity in fibrils arises because the long rod-like molecules occur in near-parallel array, regularly staggered with respect to one another by a constant axial displacement D (=234 residues) or integral multiples thereof [1]. The fibril banding receives contributions from residues in 5 staggered molecules to every 4 molecules. As each D-period is identical in axial projection, it is sufficient for the purposes in the present study to treat the image of a fibril as 'D-periods'.

The manual process of selecting periods and correlating them with a theoretical model is quite cumbersome. Considering the recent advances in computer hardware and software, it becomes feasible to develop an automated algorithm for performing the tasks of period selection and merging into an average D-period reflecting the structure of a microscope fibril image. The development of an automated fibril analysis tool involves two major concepts; a) an automated process for selecting periods according to a specific criterion and b) an analysis methodology aiming to quantify the existence of collagens and their ratio into the fibril

This paper forms an initial attempt to address these two problems. It primarily aims to prove the feasibility of such an analysis concept rather than to develop a fully automated procedure for fibril analysis. Thus, both steps of period selection/extraction and D-period analysis are implemented as semi-automated procedures where the user's experience and the reference to a sequence generated histogram or theoretical model are still critical. Regarding period selection and extraction, we propose two algorithms exploiting different signal processing methodologies. Both of them are based on the correlation between image periods and theoretical model periods. The first one, operating on 1-D digital signals, selects periods on the major lines crossing the image's fibrils. The second one operates on the fibril image through a small 2-D model template. In this case, the search for periods is done not on isolated lines but on image regions matching the 2-D structure of the model template. Regarding the second problem, the D-period analysis, we develop a framework and an algorithm for fibril analysis, which indicates the relative strength of a collagen model in the fibril rather than derives the exact analogy of different collagen models. Therefore, this analysis technique requires the generation of a theoretical model, which will guide and influence the analysis results.

II. Preprocessing

Both algorithms proposed for D-period extraction share a common preprocessing procedure aimed mostly at understanding and representing the topology of the image. More specifically, it aims at extracting a main line of the image crossing a fibril, localizing all major lines crossing fibrils, and estimating the length of D-periods to be expected in this specific image. The major lines of the image, as well as its main line, are extracted by exploiting the properties of the Hough transform operating on a segmented version of the initial image. The D-period length is initially estimated based on the correlation of different length versions of a

Manuscript received June 30 , 2006.

M. Zervakis is with Dept of Electronic and Computer Eng. Technical University of Crete Chania 73100, GREECE (e-mail: michalis@danai.systems.tuc.gr)
M. Tzaphlidou is with Medical Physics Lab Medical School University of Ioannina Ioannina 45110, GREECE (e-mail: mtzaphli@uoi.gr)

theoretical model with the main line of the image.

III. Algorithms For D-period Extraction

In this section we present two algorithms for semi-automatic D-period extraction from positively stained collagen fibril images. These algorithms operate under the assumption that the orientation of D-periods in an image remains unchanged among the image fibrils.

A. 1-D Search Algorithm

The first of the proposed algorithms uses a 1 dimensional pattern recognition scheme working on image straight lines that have been identified from the preprocessing step as the most likely to cross collagen fibrils. Its basic characteristic is that it stretches image lines so that prospective D-periods reach the length of a quite long version of a reference pattern (i.e. model I). This interpolation step turns out to be very important for the robustness of the algorithm, as it allows the use of model I that retains most of the original model information. Recall that the typical sizes of theoretical models are in the order of 260 pixels, whereas the size of a D-period in a digitized fibril image may be as small as 45-50 pixels. The reverse approach, i.e. the downsampling of the model to the size of the image D-period results in severe downsampling and a significant loss of information through aliasing.

The basic processing steps of the algorithm are: Line stretching and correlation profile, Correlation coefficient threshold - segmentation of the correlation profile table**,** Final line scanning and D-period sampling, Average image D-period generation

B. 2-D Search Algorithm

In order to increase its robustness, the second algorithm employs a two-dimensional pattern recognition scheme and searches for template rather than for linear patterns within the image. Thus, instead of identifying image lines that may contain consecutive D-periods, it attempts to localize small fibril regions, referred to as *cells*, which may contain several parallel D-periods. In order to detect such cells it correlates portions of the image with a small 2-D pattern realization of the theoretical model (i.e. model I), referred to as *driver image*. The cells of the image that highly correlate with the driver cell are treated as small templates where the algorithm searches for D-periods in a way similar to that of the 1-D search algorithm.

The basic processing steps of this algorithm are: Driver image construction and D-period orientation check, Driver image length and angle correction, Correlation with driver image along the valid image lines–Creation of correlation table, Correlation coefficient threshold–segmentation of correlation table, Sampling D-periods from acceptable cells, Average image D-period construction

C. Comparison of the D-period extraction algorithms.

As a general comment we could say that the 2-D search algorithm is more efficient than the 1-D search in selecting valid periods from the fibrils. Its efficiency boost comes from the use of a 2-D pattern, i.e. the driver image, to locate well-balanced image regions. The 2-D search algorithm deals better with noisy or blurred image fibrils, since a 2-D pattern is affected less by noise and other degradations than an individual 1-D vector. On the contrary, the 1-D search algorithm tends to sample a small number of unacceptable D-periods.

The robustness of the 2-D search algorithm allows for the implementation of efficient and fast heuristics that increase the autonomy and auto-correction ability of the algorithm. Efficient heuristics, such as automatic D-period orientation and angle correction, work well within the context of this algorithm, whereas they proved to be too unstable when implemented for the 1-D search algorithm.

IV. Analysis Of Average D-periods

The second part of this work addresses the analysis of D-periods to derive information about the relative content tendency of type I over type III collagen in fibrils. Towards this direction we develop a framework for comparing D-periods with linear combinations of theoretical models I and III. The linear combination of types I and III is justified by the averaging of detected periods in the fibril into just one average D-period, where noisy periods of all types are indeed linearly combined. We emphasize that the proposed framework does not derive the exact analogy of types I and III in the fibril but it rather derives a tendency indicator that signifies the portion of reference (type I) D-periods in the fibril under consideration.

The core of the method is based on comparing an average D-period with several linear combinations of the theoretical models I and III, referred to as *mixture models*. The comparison is based on the Pearson's r correlation coefficient. To encounter for neighborhood dependencies in the D-period, which may be introduced by misalignment of the individual periods extracted, the comparison is preceded by a least squares fit of each mixture model to the D-period using the Singular Value Decomposition (SVD) technique.

Using either one of the two methods developed in section III we end up with the image's average D-period. As in the tissues used the predominant component, type I collagen, is associated with type III, the average D-period has to be a mixture of the two types. An exact estimation of the relative amounts of these two types based only on the average D-period, may be impossible. Considering also the fact that the proposed algorithms for D-period extraction favor the sampling of periods that resemble a specific theoretical model, it becomes obvious that we do not intend to identify the exact ratio of type I and III collagen. Nevertheless, we can detect slight but noticeable differences in the content of

type I over type III collagen comparing the sequence generated D-period with the averaged experimental one. As a consequence, we can also infer that the content of type I collagen in this fibril is higher than in the other.

V. Application Examples

We now present specific application examples that demonstrate the use of the preceding tools. The following table summarizes the results deriving from three different tissues, i.e. bone, liver and skin. The selection and preparation of data are described in a previous publication [2]. When using the 1-D search algorithm as period extractor, the relative type I tendency for the bone – liver – skin sets is respectively 99%, 95% and 93%, while the corresponding results for the 2-D search algorithm are 99%, 96% and 93%. The two algorithms sample quite different D-periods from the same images. The proximity of the results of the two D-period extraction algorithms a) indicate that the proposed algorithms are quite robust and b) enhance the validity of the derived tendency in type I content. Our results are in good agreement with those derived from biochemical methods confirming that in bone tissues type I collagen is the main constituent while adult bone contains only this type [3]. Liver contains an assortment of collagen types, usually with type I predominating, though this is not always the case. Skin has a different overall profile from other tissues. Two collagen types, I and III, are the main constituents accompanied by other minor types [3].

VI. Conclusions

In this work we explore the potential for automatic analysis of microscopic images from periodic-fibrillar structures. Quantitative analysis of electron-optical banding patterns has been shown to be an extremely sensitive technique capable of yielding information at a resolution obtainable by no other method [4]. The analysis described in this article is capable of revealing whether or not more than one collagen type is present within the same collagen fibril.

This work forms a first attempt to design algorithms for both selection and analysis of periods. At this stage, the algorithms need tuning from experts, but their performance indicate that consistent results can be obtained and encourages further study and development towards this direction. Furthermore, it develops an algorithm for the analysis and comparison of average D-periods with sequence generated histograms. This algorithm aims in deriving the relative content of type I protein, but it can be easily modified to deal with other types of fibrillar proteins providing that their sequence data are known. The proposed algorithms are applied to three different tissues, bone, liver and skin, which all contain mainly type I collagen associated with type III. The amount of type III depends on the tissue. This property was well detected by the proposed method.

Towards a fully automated procedure, both the selection and analysis algorithms must be further developed. For the selection of periods, we recognize the need for isolating individual fibrils from a microscopic image and the adaptive selection of the algorithm threshold depending on the contrast of each fibril. In the level of analysis of average D-periods, the proposed algorithm must be expanded as to enable the estimation of the actual ratio of each protein in the fibril. The aforementioned assessment of the type of each selected period would highly assist the derivation of protein content in the same fibril rather than the derivation of content tendencies of a single model when comparing different fibrils, which is the current stage of development. Alterations in the relative amounts of collagen types (present in tissues), upon disorders, could be the key for the study and perhaps ultimately the cure of a number of diseases.

ACKNOWLEDGEMENTS

This work was partially supported by Biopattern, IST EU funded project, Contract no.: 508803 as well as by HERPs Center, Leonardo da Vinci Project Ref: RO/04/B/P/PP 17 5006.

References

[1] A.J. Hodge, J.A. Petruska, “Recent studies with the electron microscope on ordered aggregates of the tropocollagen macromolecule”, In: Aspects of Protein Structure, Ed. Ramachandran G.N., London and New York, Academic Press, pp. 289-300, 1963.

[2] M.E. Zervakis, V. Gkoumplias and M. Tzaphlidou, “Analysis of Fibrous Proteins from Electron Microscopy Images”, Medical Engineering and Physics, vol. 27, pp. 655-667, 2005.

[3] J. Weiss, “Collagens and collagenolytic enzymes”, In: Connective Tissue Matrix, Ed. Hukins D.W.L., Verlag Chemie, pp. 17-53, 1984.

[4] J.A. Chapman, M. Tzaphlidou, K.M. Meek, K.E. Kadler, “The collagen fibril - A model system for studying the staining and fixation of a protein”. Electron Microscopy Reviews, vol. 3, pp. 143-182, 1990.

TABLE 1. BONE – LIVER – SKIN SETS. CHARACTERISTICS OF AVERAGE D-PERIODS EXTRACTED BY THE 1-D SEARCH ALGORITHM.

Bone 9-weeks-old mice - Normal				**Liver** 8-months-old rabbits- Normal				**Skin** 6-months-old rats - Normal			
Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient	Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient	Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient
B1	91	0.90	0.89	L1	45	0.96	0.93	S1	103	0.86	0.95
B2	100	0.99	0.90	L2	29	0.94	0.95	S2	61	0.98	0.94
B3	35	0.86	0.89	L3	32	0.99	0.94	S3	78	0.99	0.95
B4	63	0.99	0.88	L4	100	0.91	0.95	S4	108	0.75	0.92
B5	86	0.99	0.87	L5	63	0.98	0.96	S5	69	0.99	0.90
								S6	66	0.88	0.87
Total Set	375	**0.99**	0.90		269	**0.95**	0.97		485	**0.93**	0.93

TABLE 2. BONE – LIVER – SKIN SETS. CHARACTERISTICS OF AVERAGE D-PERIODS EXTRACTED BY THE 2-D SEARCH ALGORITHM.

Bone 9-weeks-old mice - Normal				**Liver** 8-months-old rabbits- Normal				**Skin** 6-months-old rats - Normal			
Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient	Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient	Image ID	Number of sampled periods	Type I proportion for best fit mixture model	Correlation Coefficient
B1	47	0.99	0.87	L1	24	0.98	0.81	S1	67	0.83	0.93
B2	32	0.96	0.83	L1_2	12	0.93	0.86	S2	41	0.96	0.93
B3	13	0.87	0.85	L2	31	0.95	0.92	S3	50	0.96	0.91
B4	19	0.99	0.87	L3	20	0.96	0.93	S4	45	0.83	0.93
B5	36	0.99	0.81	L4	90	0.90	0.96	S5	19	0.99	0.90
				L5	46	0.93	0.95	S6	57	0.90	0.85
Total Set	180	**0.99**	0.87		223	**0.96**	0.95		279	**0.93**	0.92