Scientific Research Publishing

JBiSE

ISSN: 1937-6871

Vol. 2 No. 1 February 2009

Journal of Biomedical Science and Engineering

A Segmentation Result on Left Kidney Using 3D Semi-Variogram Based SRG

www. scirp. org/journal/jbise

Journal Editorial Board

ISSN 1937-6871 (Print) ISSN 1937-688X (Online)
Http://www.scirp.org/journal/jbise

Journal of Biomedical Science and Engineering (JBiSE)
Published Online February 2009 in SciRes (http://www.scirp.org/journal/jbise).

TABLE OF CONTENTS

Volume 2, Number 1, February 2009

Journal of Biomedical Science and Engineering (JBiSE)

SUBSCRIPTIONS

The ***Journal of Biomedical Science and Engineering*** (Online at Scientific Research Publishing, www.scirp.org) is published quarterly by Scientific Research Publishing, Inc. 5005 Paseo Segovia, Irvine, CA 92603-3334, USA.
E-mail: jbise@scirp.org

Subscription Rates: Volume 2 2009

Printed: $50 per copy.

Electronic: freely available at www.scirp.org.

To subscribe, please contact Journals Subscriptions Department at jbise@scirp.org.

Sample Copies: If you are interested in obtaining a free sample copy, please contact Scientific Research Publishing, Inc at jbise@scirp.org.

SERVICES

Advertisements

Contact the Advertisement Sales Department at jbise@scirp.org.

Reprints (a minimum of 100 copies per order)

Contact the Reprints Co-ordinator, Scientific Research Publishing, Inc. 5005 Paseo Segovia, Irvine, CA 92603-3334, USA.
E-mail: jbise@scirp.org

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: jbise@scirp.org

J. Biomedical Science and Engineering, 2009, 2, 1-8

Texture feature based automated seeded region growing in abdominal MRI segmentation

Jie Wu[1], Skip Poehlman[1], Michael D. Noseworthy[2], Markad V. Kamath[3]

[1]Department of Computing and Software. [2]Brain-Body Institute, St. Joseph's Healthcare, Hamilton, Ontario, Canada. [3]Department of Medicine, McMaster University, Hamilton, Ontario, Canada. Correspondence should be addressed to Wu Jie ({wuj7, skip, nosewor, kamathm}@mcmaster.ca)

Received July 2nd, 2008; revised July 31st, 2008; accepted October 14th, 2008

ABSTRACT

A new texture feature-based seeded region growing algorithm is proposed for automated segmentation of organs in abdominal MR images. 2D Co-occurrence texture feature, Gabor texture feature, and both 2D and 3D Semi-variogram texture features are extracted from the image and a seeded region growing algorithm is run on these feature spaces. With a given Region of Interest (ROI), a seed point is automatically selected based on three homogeneity criteria. A threshold is then obtained by taking a lower value just before the one causing 'explosion'. This algorithm is tested on 12 series of 3D abdominal MR images.

Keywords: Image Segmentation, Seeded Region Growing, Texture Analysis

1. INTRODUCTION

Nowadays when radiologists face hundreds of images every day, automatic analysis of medical images becomes of particular interest to researchers, as it is an effective support tool for diagnosis or quantitative analysis. Medical image segmentation, a critic step for most subsequent image analysis tasks, is to delimit the image areas representing different anatomies. Segmentation of the abdomen, in particular, is often a challenging task due to the considerable overlap of soft tissues [4].

Since intensity-based methods have met with limited success for abdominal segmentation, texture segmentation, which makes use of statistical textures analysis to label regions based on their different textures, has attracted our attention. In this approach, low-level features based on texture information, that is expected to be homogenous and consistent across multiple slices for the same organ, are mostly used to perform automatic image analysis in the medical imaging field [3].

Among various image segmentation methods, the Seeded Region Growing (SRG) algorithm, originally proposed by Adams and Bischof [6], is a fast, robust, parameter-free method for segmenting intensity images given initial seed locations for each region. In SRG, individual pixels that satisfy some neighborhood constraint are merged if their attributes, such as intensity or texture, are similar enough. The seed location, an optimal threshold value and a similarity measure need to be determined either manually or automatically.

The goal of the approach presented in this paper is to achieve automatic texture based segmentation of organs in MR images of the abdomen. We first extract texture features for each pixel in the ROI. Three texture features are examined: co-occurrence, semi-variogram and Gabor texture feature. co-occurrence[5] is a well-known and successful texture feature in medical image analysis, while semi-variogram [1] is a widely used measure of dissimilarity in geostatistics. Co-occurrence and semi-variogram are both statistical texture features, while the Gabor filter, on the other hand, is one of the most popular signal processing based approaches for texture features. In this paper we also investigate extensions of semi-variogram methods to volumetric data. Volumes are often processed as a series of 2-D images. 2-D texture features are computed for pixels in each slice. Unfortunately, by processing volumes as a series of separate 2-D slices, texture information across slices is ignored. Our methods for computing volumetric texture features have been developed to include this extra texture information. Secondly our automatic SRG algorithm is run on the feature spaces. The seed is determined by minimizing a cost function with three factors. The threshold is obtained by taking a lower value just before the one causing 'explosion'. Some improvements are made to avoid under-segmentation, over-segmentation and to speed up the calculation. SRG is then applied and a right kidney is extracted in the experiment.

The contributions of this paper are as follows:

Semi-variogram texture feature is, for the first time, used on abdominal organ segmentation and extended for examining volumetric MR images

A novel automated SRG algorithm is proposed and successfully applied on abdominal MR images

2. BACKGROUND

There has been some research in the field of texture analysis in medical image segmentation: In [10] 3D extended, multisort co-occurrence matrices have been ap-

plied on MRI brain datasets. Karkanis *et al*. [11] applied a multilayer feed-forward neural network based on second order gray level statistics to classify cancer regions in colonoscopic images. In [12] statistical, gradient and Gabor filter features are used to segment prostatic adenocarcinoma. Among all, co-occurrence matrices are the most widely used texture feature.

In the medical imaging field, SRG has been successfully used to segment medical images for different purposes, for cervical cancer [7], extraction of cerebral blood vessels [8] and breast cancer detection [9]. There are also some trials on automating the SRG algorithm: Whitney *et al*. [14] overcomes the need to manually select threshold values by analyzing the histogram of voxel similarity to automatically determine a stopping criterion, but they still require the user to choose a seed point. In [2] Law *et al*. proposed a Genetic Algorithm based seed selection method and a threshold value optimization method, but their algorithm has the problem of possible under-segmentation and speed can also be an issue although they did not address them.

3. METHODOLOGY

Our system consists of two stages: i) After MR images are loaded into the system, each pixel in the ROI is processed and three features: co-occurrence and semi-variogram are extracted respectively; ii) Automated SRG algorithm is applied on texture feature space and in the end a region grown out of the seed is obtained. **Figure 1** illustrates the system diagram.

4. FEATURE EXTRACTION

Pixel-level feature extraction is used to discover the similarities between pixels. The intensity-based method is very straightforward: the gray-level of each pixel is its feature. But texture feature extraction is more complicated.

While there are generally two directions on texture analysis: **"all-pairs"** approach [13], where the local texture is calculated by all pixels in the neighborhood; and **"direction distance pairs"** approach [3], where local texture is calculated for every direction and distance, we adopted **"direction distance pairs"** approach, as it takes various permutations of pixels into consideration. **Figure 2** shows how distance and directions are defined.

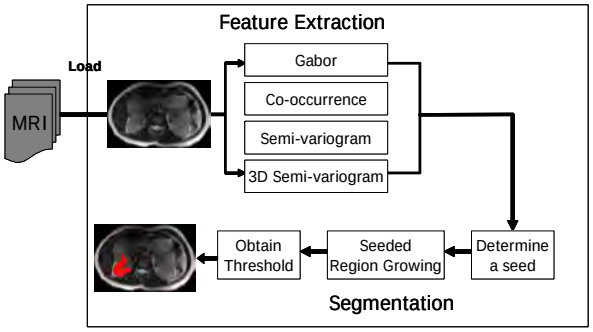


Figure 1. system diagram

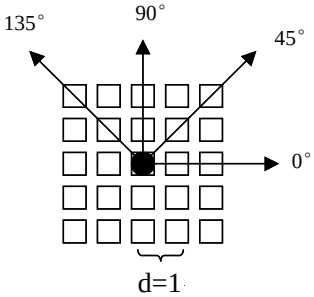


Figure 2. distance and directions

4.1. Co-occurrence

Grey level co-occurrence texture features were proposed by Haralick [5] in 1973 to extract second order statistics from an image. The grey level co-occurrence matrices (GLCM) was defined as a matrix of frequencies at which two pixels (in specified direction and distance) occur in the image. This matrix is square with dimension N_g, where N_g is the number of grey levels in the image.

In MR images, the number of grey levels is very big compared to normal pictures. An image with 256 grey levels will have a 256*256 co-occurrence matrix, while one with 1400 gray levels will have a 1400*1400 co-occurrence matrix. *Binning strategy* thus becomes of great importance on the reduction of grey levels in a MR image in terms of its ability of combining several intensities into a single intensity level, or a bin.

We adopted *clipped binning strategy* where one large bin is allocated for low intensity gray levels (0-I_1), one for high intensity gray levels (I_2 and above) and 30 equal bins are allocated for the remaining intensity gray levels.

The low intensity gray level I_1 and high intensity gray level I_2 are determined by the image histogram.

The low intensity gray level I_1 is set to be the value of first valley, as pixels with gray levels from 0 to it are mostly background pixels. The high intensity gray level I_2 is chosen as the first one with very few pixels, as these pixels are trivial.

After binning a GLCM is created for a 3*3 neighborhood of each pixel instead of an image. Four directions

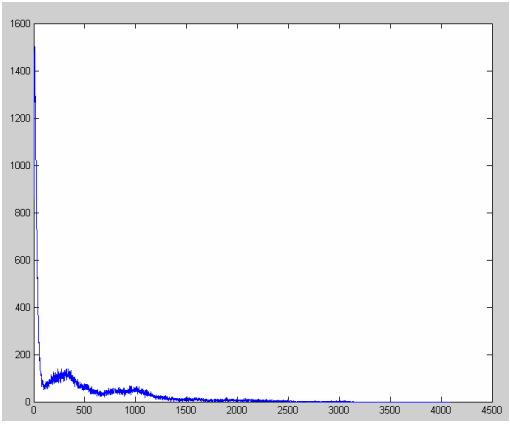


Figure 3. a sample histogram

(0°, 45°, 90°, 135°) and distance=1 are used to find pairs of pixels.

Once the GLCM has been created, Haralick then described 14 statistics that can be calculated with the intent of describing the texture of the image:

These 14*4=56 features consist of the co-occurrence texture feature space for each pixel.

4.2. Gabor

Gabor wavelets are one of the most popular signal processing based approaches for texture feature. The 2-D Gabor functions are Gaussians modulated by complex sinusoids as follows:

$$\psi(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x^2}{2\sigma_x^2}-\frac{y^2}{2\sigma_y^2}\right)\exp(2\pi Wjx)$$

where W is the modulation frequency, and σx and σy are the standard deviations of the two-dimension Gaussian distribution along x and y directions. The Gabor filter masks work as orientation and scale tunable detectors. The statistics of these microfeatures in a given region can be used to capture the underlying texture information.

Zhang *et al* in [15] introduced a Gabor wavelet based texture representation for content based image retrieval. They attempted to find images or regions with the same texture and achieved satisfactory results. In our study, for each pixel we first apply Gabor filters with different scale and orientations on its 3*3 neighborhood window:

$$E(m,n)=\sum_x\sum_y\left|G_{mn}(x,y)\right| \tag{9}$$

where E(m,n) is the summation of Gabor wavelet transform of each pixel in the image with different scale m and orientation n.

A mean and a standard deviation of the magnitude of the transformed coefficients are then calculated to represent the homogeneous texture features:

$$\mu_{mn}=\frac{E(m,n)}{W*H} \tag{10}$$

$$\sigma_{mn}=\frac{\sqrt{\sum_x\sum_y\left(G_{mn}(x,y)-\mu_{mn}\right)^2}}{W*H} \tag{11}$$

In our experiment 1 scale and 4 orientations are used, where scale is 1 and orientations are 0, 45, 90 and 135 respectively. A feature vector is thus generated with its mean and standard deviation.

Table 1. Haralick Texture Features

Feature	Formula
Energy	$\sum_i\sum_j p(i,j)^2$
Correlation	$\frac{\sum_i\sum_j (i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x\sigma_y}$ where μ_x, μ_y, σ_x, σ_y are the means and standard deviations of p_x, p_y, the partial probability density functions
Inertia	$\sum_i\sum_j (i-j)^2 p(i,j)$
Entropy	$-\sum_i\sum_j P(i,j)\log(P(i,j))=f_4$
Inverse Difference Moment	$\sum_i\sum_j \frac{1}{1+(i-j)^2}p(i,j)$
Sum Average	$\sum_{i=2}^{2N_g} ip_{x+y}(i)=f_6$ where x and y are the coordinates(row and column) of an entry in the co-occurrence matrix, and p_{x+y} is the probability of co-occurrence matrix coordinates summing to $x+y$
Sum Variance	$\sum_{i=2}^{2N_g}(i-f_6)^2 p_{x+y}(i)$
Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i)\log\{p_{x+y}(i)\}$
Difference Average	$\sum_{i=0}^{N_g-1} ip_{x-y}(i)=f_9$
Difference Variance	$\sum_{i=0}^{N_g-1}(i-f_9)^2 p_{x-y}(i)$
Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)\log\{p_{x-y}(i)\}$
Information measure of correlation 1	$\frac{f_4-HXY1}{\max(HX,HY)}$
Information measure of correlation 2	$(1-\exp[-2(HXY2-f_4)])^{1/2}$ where HX and HY are the entropies for p_x, p_y, HXY1=$-\sum_i\sum_j P(i,j)\log(p_x(i)p_y(j))$ HXY2=$-\sum_i\sum_j p_x(i)p_y(j)\log(p_x(i)p_y(j))$
Maximal Correlation Coefficient	Square root of the second largest eigenvalue of Q, where $Q(i,j)=\sum_K \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$

4.3. Two Dimensional Semi-variogram

We extracted 2D semi-variogram feature from every pixel in the region of interest (ROI) with the following steps[1]:

First, for every pixel, a 3*3 pixel neighbouring window was considered. Four directional variograms (0°, 45°, 90°, 135°) and distance d=1 were computed for all combinations in the window. The semi-variogram was then computed using the *mean Square-Root Pair Difference*

$$r(h)=\frac{1}{N}\sqrt{\sum_{i=1}^{N}\left[G(x,y)-G(x',y')\right]}$$

where G(x, y) is the gray level for the pixel (x, y).

Example: Given the following 3*3 neighbouring window:

$$\begin{matrix}1&1&2\\3&2&3\\0&1&1\end{matrix}$$

For d =1, direction =0°,

$$=\sqrt{(|1-1|+|1-2|+|3-2|+|2-3|+|0-1|+|1-1|)}/6$$
$$=\sqrt{(0+1+1+1+1+0)}/6$$
$$=2/6=0.33$$

Second, directional features were transformed to rotation-invariant features: the mean, standard deviation and sum of perpendicular ratios

$$\log[\gamma_0/\gamma_{90}+\gamma_{90}/\gamma_0+\gamma_{45}/\gamma_{135}+\gamma_{135}/\gamma_{45}]$$

4.4. Semi-variogram for Volumetric Data

We present a new approach for calculating Semi-variogram texture features for volumetric images that will allow capturing the characteristic of the texture in 3D image data.

As opposed to two dimensional Semi-variogram texture features based on the spatial dependence of gray levels within a specific slice, Semi-variogram for volumetric data is based on the spatial dependence of intensities from the current slice and its two neighboring slices (one above and one below).

Now a GLCM is created for a 3*3*3 neighborhood of each pixel in a specific slice. Four directions (0°, 45°, 90°, 135°) and distance 1 are used to find pairs of pixels. The semi-variogram was then computed using the mean Square-Root Pair Difference of not only the pair of pixels in the same slice but also in the ones a slice above and a slice below the current slice. For example, when direction=0° and distance=1, the mean Square-Root Pair Difference is calculated as follows:

$$\gamma(h)=\sqrt{\sum_{i=1}^{N}\begin{matrix}[G(x,y,z)-G(x+1,y+1,z)]^2+\\ [G(x,y,z)-G(x+1,y+1,z-1)]^2+\\ [G(x,y,z)-G(x+1,y+1,z+1)]^2\end{matrix}}$$

Following this directional features were transformed to rotation-invariant features just like in the two dimensional approach.

5. AUTOMATED SEEDED REGION GROWING

As illustrated in **Figure 4**, there are three important steps in our automated SRG algorithm. A seed is a perquisite and we need to automatically select a seed point replacing the selection through user interaction. With the given seed SRG can start to grow, but a threshold value has to be determined to only cover the reasonable pixels.

5.1. Selection of a Seed Point

The proposed seed point determination method is based on a cost-minimization approach. An ideal candidate seed point should have these properties:

i. It should be inside the region and near the center of the region

ii. Assume most of the pixels in the ROI belong to the region (i.e. ROI is not too big compared to the region), the feature of this seed point should be close to the region average

iii. The distances from the seed pixel to its neighbors should be small enough to allow continuous growing

According to these criteria, a cost function is built by adding three sub-functions corresponding to the three criteria respectively:

i. The spatial distance from the pixel to the center point of the ROI

ii. The Euclidean distance on feature space from the pixel to the centroid of the ROI

iii. The sum of the Euclidean distance on feature space from the pixel to its neighbors

We want to give equal weight to the three sub-functions in the cost function. However, as they have different quantities, we need to multiply three weights: w1, w2 and w3 to balance them, as follows:

$$f(x,y)=w_1\times g_1(x,y)+w_2\times g_2(x,y)+w_3\times g_3(x,y)$$

Thus, cost function is applied to every pixel in the ROI and the one with a minimum value is chosen as our seed. Some previous researchers have used Genetic Algorithm (GA) methods to minimize their fitness function and obtain the seed, but we choose not to do so. First our ROI is not big and our pixel number can be kept reasonably small. Second the SRG itself is a robust algorithm that is not very sensitive to the choice of the seed pixel. Third the GA algorithm is much more complicated than our straightforward algorithm. In a small sample space, on a not very critical problem, the faster and simpler way maybe better.

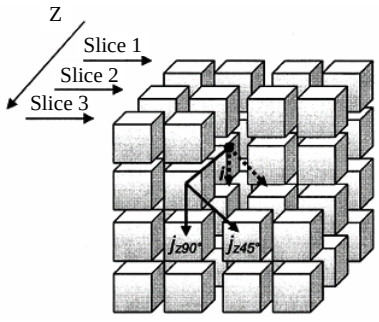


Figure 4. volumetric data

5.2. Seeded Region Growing

Given a seed point, the region growing method searches the seed point's neighbors to determine whether they belong to the same region. If they are determined to be so, their neighbors are searched. The process is recursively executed until no more new neighbors can be added to the region.

Now it comes to the question of "How to determine whether a neighbor pixel belongs to the same region?" The criterion is when the distance is lower than a threshold value a neighbor point is added. So we need to determine the distance measure, linkage strategy, connectivity strategy and a threshold value (elaborated in next section).

Distance measure: we use the Euclidean distance in the feature space as the distance measure. For example, with the semi-variogram texture feature, the distance between two pixels is the Euclidean distance of their semi-variogram feature vectors.

Linkage strategy: we tried both single linkage strategy and centroid linkage strategy. Single linkage, in which pairs of neighboring pixels are compared for merging, is one of the conceptually simplest approaches. While in

centroid linkage, a pixel's value is compared with the mean of an already existing but not necessarily completed region. We chose single linkage strategy over centroid linkage because it is faster and more memory-efficient considering the calculation of texture features and recursively running SRG requires much memory.

Connectivity strategy: in 2D region growing, there are two connectivity strategies that people use: four-neighbor and eight-neighbor. Four-neighbor region growing checks only the vertically and horizontally connected four neighbors, while eight-neighbor region growing checks vertically, horizontally and diagonally connected eight neighbors. The choice of neighbor connection strategy is usually case dependent. By our visual inspection of the segmented images from these two methods, the shapes delineated by four-neighbor methods are closer to the actual shapes than eight-neighbor methods. It is noted that four-neighbor usually produces more conservative or restrictive shapes than eight-neighbor. In our case, eight-neighbor explodes faster because the region grows more aggressively. When we try to determine the threshold value that is just before explosion, it is prone to result in under-segmentation. Thus we use four-neighbor region growing approach.

5.3. Optimization of Threshold Value

An optimal threshold value is the value that can make a stop to the region growing and the obtained region is optimal. It is desirable that the threshold value is high enough to extract the whole region; however if the threshold value is higher than the optimal one, the extracted region may grow over the actual region boundary and grow to a much larger region. This case is called 'explosion' [2].

Our idea is to find the highest threshold value just before this explosion. So our algorithm starts from a low threshold value and increases it by 1. With each threshold value we perform a SRG algorithm and evaluate its result. This is the first pass. When it comes to a value causing explosion, we retrieve the last value not causing explosion, and from that value to the explosion value, by a step of 0.1, perform another pass, and retrieve the value just before explosion. That value is our optimal threshold value. But in this algorithm we have a few issues to solve:

i. *How to quantify an explosion*? If we plot the 'threshold value' vs 'number of pixels in the region', the explosion value must have a big slope because the threshold value always increases by the same amount whereas the number of pixels in the region has increased significantly. Thus this slope value: (#pixels in the region/threshold value change) is used to check explosion. If the slope value is larger than a big value, explosion has occurred.

ii. *How to avoid under-segmentation*? The algorithm may stop if an explosion occurs before it is supposed to stop. In this case, the resulting region is smaller than the actual region and this is an under-segmentatation. To avoid stopping region growing too early, the algorithm does not stop immediately when it finds an explosion. It finds all the explosions and does not stop until their #pixels are more than the total number of pixels in the ROI. By then, we pick the last explosion with the most #pixels as the actual explosion.

iii. *How to avoid over-segmentation*? To avoid over-segmentation, two stop criteria are added to the algorithm. One is the #pixels in the resulting region cannot exceed the #pixels in the ROI. Second is the leftmost, rightmost, uppermost and lowermost pixels in the resulting region should not exceed the spatial location of these four pixels of ROI by some extent. We allow them to exceed 20 pixels, but this is flexible and case dependent.

iv. *How to solve the speed problem?* One of our biggest concerns about this algorithm is the speed since it has to do two-pass scans and for every threshold value SRG has to be performed and on texture feature space the process of texture feature extraction and distance calculation is much more complex than simple intensity features. However, we noticed a fact that can save calculation time: when threshold value increases, the resulting region is always the super set of the previous resulting region. The reason is obvious: the pixels that can be added to the region with a lower threshold value definitely can be added with a higher threshold value. In other words, with every new threshold value, we do not need to perform a SRG from scratch. We can perform a SRG based on the resulting region obtained from previous lower threshold value. From the resulting region, we extract their boundaries and start to grow on these boundary pixels with the new threshold value. In this way calculations are saved and speed is much improved.

With these issues explained and solved, we can now proceed with the application of the algorithms.

6. EXPERIMENTS AND DISCUSSION

6.1. Data

The algorithms are run on 3D abdominal MR images obtained from a GE 3T scanner at the Brain Body Institute of St. Joseph's Healthcare. The image set contains 12 series of 512*512 gray-scale images in DICOM fomat. No preprocessing is applied.

6.2. Segmentation

The algorithms are implemented with matlab and **Figure 5** shows a screenshot of our segmentation system. In Manual Mode users have the options to choose a specific threshold or give a range of threshold for the program to look for the optimal threshold value. In Auto Mode users only need to select a rectangle ROI. When a ROI is selected, the system can start segmentation of organs automatically without any user intervention.

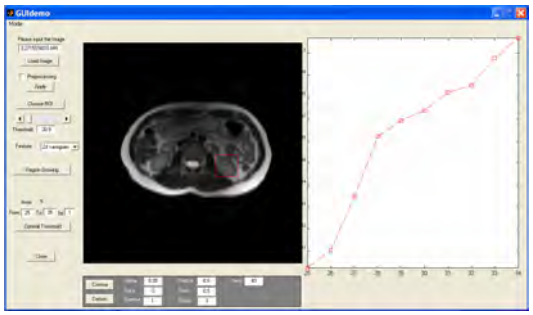


Figure 5. the user interface implemented in the algorithm (Manual Mode)

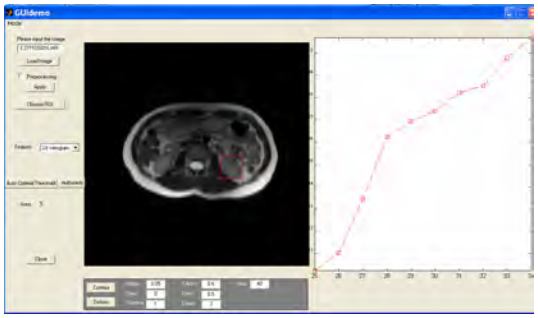


Figure 6. the user interface implemented in the algorithm (Autol Mode)

In the seed point retrieval algorithm, every pixel in the ROI is evaluated by the cost function and the one with minimum cost function is our seed point. The resultant point is examined and found to be within the region, near center and have similar feature values with most of the pixels in the region, which is required.

Figure 6 is an example of the chosen seed pixel. In the threshold determination algorithm, we test incremental threshold values from a starting value, which is given the minimum value of the distances between the seed point and its neighbors since it has been minimized through the cost function minimization process. Following this, we apply the threshold deter- mination algorithm described in 5.3. **Figure 7** is a sample plot of this optimization process. X-axis is the threshold value and Y-axis is the count of pixels in the resultant region. We

apply SRG on every incremental threshold value and when it reaches 34, an explosion is detected. The algorithm then does a second pass from 33 to 34 by 0.1, and the value 33.8 is the one just before the explosion and thus chosen to be the threshold value.

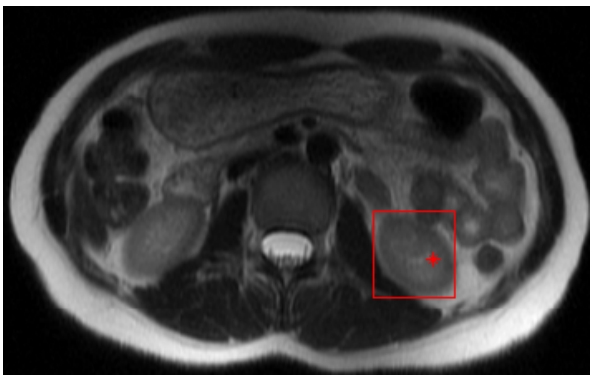


Figure 7. a sample seed pixel selected by the algorithm

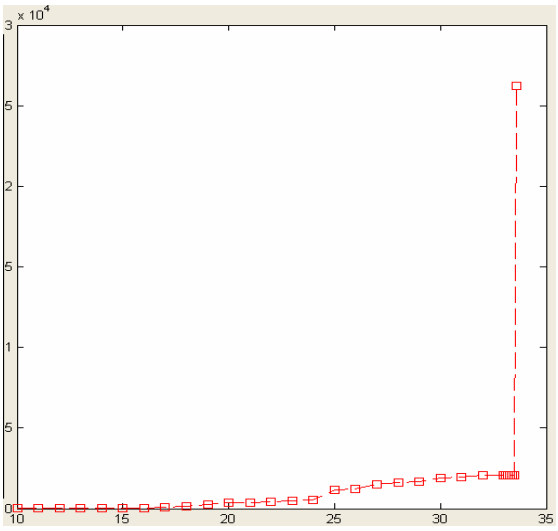


Figure 8. a sample threshold optimization plot

Then we ran our algorithms on left kidneys of abdominal MR images and obtained segmentation results for both co-occurrence and semi-variogram SRG respectively. **Figure 8** shows a sample result for 2D semivariogram, **Figure 9** for co-occurrence, **Figuer 10** for Gabor and **Figure 11** for 3D semivariogram.

If we compare these four methods, from the performance point of view, based on the results from our images, we find them comparable to each other.

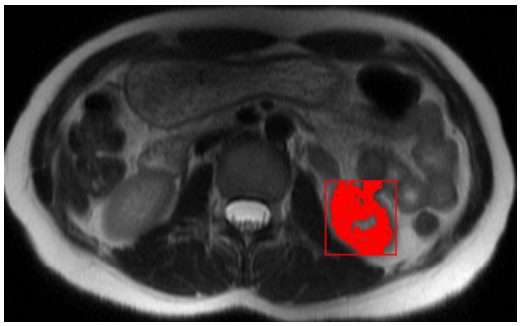


Figure 9. a segmentation result on left kidney using 2D semi- variogram based SRG

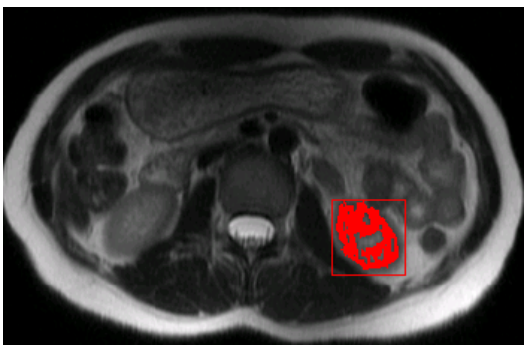


Figure 10. a segmentation result on left kidney using co-occurrence based SRG

Figure 11. a segmentation result on left kidney using Gabor based SRG

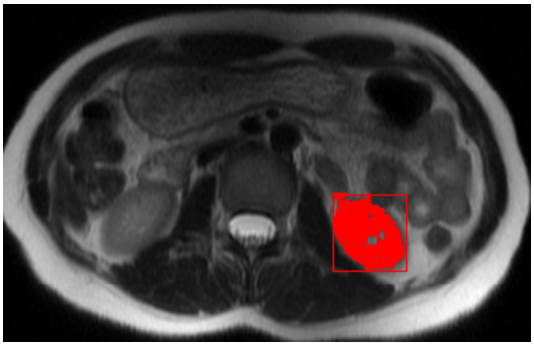


Figure 12. a segmentation result on left kidney using 3D semi-variogram based SRG

From the speed point of view, 2D semi-variogram based SRG performs much faster than co-occurrence based SRG. This is reasonable because for every pixel a co-occurrence matrix has to be built first, followed by Haralick's fourteen statistics. However semi- variogram features can be extracted from the neighboring window directly. Gabor filter method performs faster than co-occurrence and slower than 2D semi-variogram method. 3D semi-variogram method needs to process multiple slices and thus needs more computation time than 2D semi-variogram.

7. CONCLUSION

In this paper we proposed our texture feature-based automated SRG algorithm on abdominal organ segmentation. The benefit of this algorithm is obvious, as it provides a parameter-free production environment to allow minimum user intervention. This can be especially helpful for batch work or to novice computer users.

We also proposed the usage of 2D and 3D semi-variogram as a texture feature in medical organ segmentation. They are compared with the co- occurrence method and Gabor filter method and found to be a feasible texture feature in organ segmentation.

But our approach does have drawbacks. Texture feature based methods all have the assumption that the region should have texture homogeneity. For organs with complex texture like the heart, this approach might not work well [14]. In our experiment the segmentation results leave some pixels inside kidney out also because these pixels have different texture homogeneity with the others. This, on the other hand, can help detect cysts or tumors inside organs.

Our future work includes investigation of other 2D and 3D texture features and evaluation of their performance. We are also combining this approach with other edge detecting or deformable model approaches to get a better boundary.

ACKNOWLEDGEMENT

This work is in part supported by Natural Sciences and Engineering Research Council of Canada (NSERC) granted to Dr. Markad V. Kamath.

REFERENCES

[1] P. Maillard, (2001) "Developing methods of texture analysis in high resolution images of the Earth", X Simpósio Brasileiro de Sensoriamento Remoto, São Paulo-SP: Fábrica da Imagem. 1−11.

[2] T.-Y. Law and P. A. Heng, (2000) "Automated extraction of bronchus from 3-D CT images of lung based on genetic algorithm and 3-D region growing", Proc. SPIE 3979, Medical Imaging 2000: Image Processing, 906−916.

[3] R. Susomboon, D. S. Raicu, and J. D. Furst, (2006)"Pixel-Based Texture Classification of Tissues in Computed Tomography", CTI Research Symposium, Chicago, April 2006.

[4] J. E. Koss, F. D. Newman, T. K. Johnson, D. L. Kirch, (1999) "Abdominal organ segmentation using texture transform and a Hopfield neural network", IEEE Trans. Medical Imaging, Vol.18, 640−648.

[5] R. M. Haralick, K. Shanmugam, and I. Dinstein, (1973) "Texture Features for Image Classification", IEEE Trans. On Systems, Man, and Cybernetics, Vol. Smc-3, No.6, 610−621.

[6] R. Adams, L Bischof, (1994) "Seeded region growing". IEEE Trans. Pattern Anal. Machine Intell. 16, 641-647.

[7] N. A. Mat-Isa, M. Y. Mashor & N. H. Othman, (2005) "Seeded Region Growing Features Extraction Algorithm; Its Potential Use in Improving Screening for Cervical Cancer". International Journal of the Computer, the Internet and Management. (ISSN No: 0858-7027). Vol. 13. No. 1. January-April.

[8] Y. Tuduki, K. Murase, M. Izumida, H. Miki, K. Kikuchi, K. Murakami & J. Ikezoe (2000). "Automated Seeded Region Growing Algorithm for Extraction of Cerebral Blood Vessels from Magnetic Resonance Angiographic Data". Proceedings of The 22nd Annual International. Conference of the IEEE Engineering in Medicine and Biology Society. 3. 1756−1759.

[9] P. A. Venkatachalam, U. K.Ngah, A. F. M. Hani& A. Y. M. Shakaff, (2002). "Seed Based Region Growing Technique in Breast Cancer Detection and Embedded Expert System". Proceedings of International Conference on Artificial Intelligence in Engineering and Technology. 464−469.

[10] V. A. Kovalev, F. Kruggel, H.-J Gertz, and D.Y. von Cramon. (2001) "Three-dimensional texture analysis of MRI brain datasets." IEEE Trans. on Medical Imaging, 20(5): 424−433.

[11] S. A. Karkanis, *et al.*, (1999) "Detecting abnormalities in colonoscopic images by texture descriptors and neural networks," Proc. of the Workshop Machine Learning in Med. App., 59−62.

[12] A. Madabhushi, M. Feldman, D. Metaxas, D. Chute, and J. Tomaszewski. (2003) "A novel stochastic combination of 3D texture features for automated segmentation of prostatic adenocarcinoma from high resolution MRI." Medical Image Computing and Computer-Assisted Intervention, volume 2878 of Lecture Notes in

Computer Science, pp. 581−591. Springer-Verlag.

[13] M. Kalinin, D. S. Raicu, J. Furst, and D. S. Channin, "A classification Approach for anatomical regions segmentation", IEEE Int. Conf. on Image Processing, 2005.

[14] B. W. Whitney, N. J. Backman, J. D. Furst, D. S. Raicu, (2006) "Single click volumetric segmentation of abdominal organs in Computed Tomography images", Proceedings of SPIE Medical Imaging Conference, San Diego, CA, Februar.

[15] D. Zhang, A. Wong, M. Indrawan, G. Lu, (1990) Content-Based Image Retrieval Using Gabor Texture Features, IEEE Transactions PAMI, 12:7, 629−639.

[16] J. Wu, S. Poehlman, M. D. Noseworthy, M. Kamath, (2008) Texture Feature based Automated Seeded Region Growing in Abdominal MRI Segmentation, 2008 International Conference on Biomedical Engineering and Informatics, Sanya, China, May 27−30.

J. Biomedical Science and Engineering, 2009, 2, 9-15

Assessment of depth of anesthesia using principal component analysis

Mina Taheri[1], Behzad Ahmadi[2], Rassoul Amirfattahi[3] and Mojtaba Mansouri[4]

[1,2,3]Digital Signal Processing Research Lab, Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, 84156-83111, Iran
[1,2,3,4]Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran. Correspondence should be addressed to Mina Taheri (mh.taheri@ec.iut.ac.ir)

Received June 9th, 2008; revised October 12th, 2008; accepted November 13th, 2008

ABSTRACT

A new approach to estimating level of unconsciousness based on Principal Component Analysis (PCA) is proposed. The Electroencephalogram (EEG) data was captured in both Intensive Care Unit (ICU) and operating room, using different anesthetic drugs. Assuming the central nervous system as a 20-tuple source, window length of 20 seconds is applied to EEG. The mentioned window is considered as 20 nonoverlapping mixed-signals (epoch). PCA algorithm is applied to these epochs, and largest remaining eigenvalue (LRE) and smallest remaining eigenvalue (SRE) were extracted. Correlation between extracted parameters (LRE and SRE) and depth of anesthesia (DOA) was measured using Prediction probability (P_K). The results show the superiority of SRE than LRE in predicting DOA in the case of ICU and isoflurane, and the slight superiority of LRE than SRE in propofol induction. Finally, a mixture model containing both LRE and SRE could predict DOA as well as Relative Beta Ratio (RBR), which expresses the high capability of the proposed PCA based method in estimating DOA.

Keywords: Bispectral Index, Depth of Anesthesia, Eignevalue Decomposition, Principal Component Analysis

1. INTRODUCTION

To provide optimal working conditions for surgeons in the operating room as well as ensuring patient's safety, an anesthesiologist's effort is absolutely essential. However, patient awareness during surgery with the rate of 1:1000 [1] and over dosing with anesthetic agents is of major clinical concerns of anesthesia. Therefore, the necessity to assess and monitor the depth of anesthesia (DOA) is obvious. In conventional methods, DOA is measured based on the monitoring of several physiological signals such as respiration pattern, blood pressure, body temperature, tearing, sweating and heart rate [1], even though these signals are affected indirectly by anesthetic agents. On the other hand, these agents have significant effects on the electroencephalogram (EEG) waveform.

A large amount of information can be extracted from EEG waveform based on different signal processing methods. Ability of this information to predict DOA depends on the variation of its value in different levels of anesthesia. In general, the goal is to produce a unit-less EEG-based index that monotonically quantifies DOA. Several methods are available that have recently been reviewed by Freye *et al.* [2] and Jameson *et al.* [3].

The earliest methods were based on the FFT analysis of EEG signals. These approaches tend to find parameters that describe spectrum characteristics. Peak power frequency (PPF), median power frequency (MPF), and spectral edge frequency (SEF) have been the first descriptors in this field. Another parameter extracted from spectrum was the ratio of power in two empirically derived frequency bands [4]. In a work presented by Traast *et al.* [5] the power of EEG in different frequency bands was determined and the results indicate pronounced changes in EEG during emergence from propofol/sufentanil total intravenous anesthesia.

Zikov *et al* [6] proposed a wavelet based anesthetic value for central nervous system monitoring (WAV_{CNS}) that quantifies the depth of consciousness between awake and isoelectric state. Their proposed technique is based on the analysis of the single-channel (frontal) EEG signal using stationary wavelet transform (SWT). The wavelet coefficients calculated from the EEG are pooled into a statistical representation which is then compared to well-defined awake and isoelectric states. Presenting a clinical study, they compared this technique with BIS monitor (Aspect Medical Systems, MA) as a reference and showed that they are well correlated (r=0.969). Furthermore, WAV_{CNS} had a faster algorithm than BIS and was well suited for use as a feedback sensor in advisory systems and closed-loop control schemes.

Ferenets *et al* [7] analyzed the performance of several new measures based on the regularity and complexity of the EEG signal. These measures consist of spectral en-

tropy (SpEn), approximate entropy (ApEn), and Higuchi fractal dimension (HFD) and Lempel-Ziv complexity (LZC). Their results show superior ability of the mentioned measures to predict DOA. Due to the arguments presented in their paper it is not feasible to point out "the best" EEG measure for the assessment of the depth of sedation, their results indicate that the measures sensitive to both the power spectrum as well as the amplitude distribution, i.e., the ApEn, LZC and HFD, perform slightly better than the other two tested measures. In the case of their tested measures, they recommend window length of 20 s.

Application of neural networks (NN) in estimating DOA is reviewed by Robert *et al* [8]. They examined a large number of EEG derived parameters as NN inputs including spectral, entropy, complexity, bicoherence, wavelet transformation derived, autoregressive modeling and hemodynamic parameters as well as a great NN topology such as MLP and Self-Organizing networks. Finally, they recommended a two hidden layers MLP model or an ART model in which their weights are continuously updated after training phase. Moreover the use of qualitative parameters, besides quantitative ones, as network inputs is recommended. In a recent work by Lalitha *et al* [9] non-linear chaotic features and neural network classifiers are used to detect anesthetic depth levels. Chaotic features consist of correlation dimension (CD), Lyapunov exponent (LE) and Hurst exponent (HE) are used as features and two neural network models, i.e., multi-layer perceptron network (feed forward model) and Elman network (feedback model) are used for classification. Their experimental results show that the Lyapunov exponent feature with Elman network yields an overall accuracy of 99% in detecting the anesthetic depth levels.

According to various mentioned methods, different EEG monitors have been developed. The Narcotrend™ monitor (Monitor Technik, Bad Bramsted, Germany) that is based on pattern recognition of the raw EEG and classifies the EEG into different stages, introduces a dimensionless Narcotrend™ index from 100 (awake) to 0 (electrical silence). The algorithm uses parameters such as amplitude measures, autoregressive modeling, fast Fourier transform (FFT) and spectral parameters [10]. The SEDLine™ EEG monitor capable of calculating of PSI™ index uses the shift in power between the frontal and occipital areas. The mathematical analysis includes EEG power, frequency and coherence between bilateral brain regions [11]. Datex-Ohmeda™ s/5 entropy Module uses entropy of EEG waves to predict DOA [3] and finally BIS™ (Aspect Medical Systems, Newton, MA), that is the first monitor in the marketplace and has become the benchmark comparator for all other monitors, introduces the BIS™ index (that is a unit-less number between 100 and 0) as a DOA indicator based on combination of spectral, bispectral and temporal analysis [4]. Approximately 450 peer-reviewed publications between 1990 and 2006 have examined the effectiveness, accuracy and usefulness, both clinical and economical, of the BIS™ monitor [3].

The aim of Principal Component Analysis (PCA) is to find source signals which are gaussian and uncorrelated. PCA can be interpreted in terms of blind source separation methods inasmuch as PCA is like a version of ICA in which the source signals are assumed to be gaussian. In other words, PCA finds a matrix which transforms the signal mixtures into a new set of uncorrelated signals. Extracted signals are ordered via PCA according to their variances (variance can be equated with power or amplitude). Consequently, the function of PCA is more than simply finding a transformation of the signal mixtures [12].

PCA has been widely used in pattern recognition and signal processing. The major applications and examples are engineering and scientific disciplines, e.g., in data compression, feature extraction, noise filtering, signal restoration, and classification [13]. PCA is used widely in data mining as a data reduction technique. In image processing and computer vision, PCA representations have been used for solving problems such as face and object recognition, tracking, detection, background modeling, parameterizing shape, appearance, and motion [14, 15]. In [16], the noise sensitivity, specificity and accuracy of the PCA method is evaluated by examining the effect of noise, base-line wander and their combinations on the characteristics of ECG for classification of true and false peaks.

The most important biomedical application of ICA and PCA is identifying different types of generators of the EEG as well as identifying its magnetic counterpart (MEG) [17]. MEG measurements give basically very similar information to EEG, but with a higher spatial resolution. MEG is mainly used for basic cognitive brain research. Another contribution is noise cancellation for brain signals such as electroencephalograms and magnetoencephalograms (EEG/MEG). References [18,19] introduced a new method to separate brain activity from artifacts using ICA. The approach is based on the assumption that the brain activity and the artifacts, e.g. eye movements or blinks, or sensor malfunctions, are anatomically and physiologically separate processes, and this separation is reflected in the statistical independence between the magnetic signals generated by those processes. In addition, ICA has been applied to problems in fields as diverse as speech processing, brain imaging (e.g., fMRI and optical imaging [20]), electrical brain signals (e.g., EEG signals), to extract features from a special array of electroencephalographic electrodes. The ICA framework can also be used for feature extraction from other kinds of data, for example, color and stereo images [21,22].

Additionally, EEG from patients undergoing surgery was collected. We introduce a novel method based on PCA. Our concentration would be on eigenvalues and eigenvectors. Finally, based on proper statistical methods and our data bank, the correlation between the extracted parameters and BIS index is observed. The reminder of paper is organized as follows: In part 2, methods and materials is described. The results and discussion are presented in section 3, and section 4 contains the final conclusion.

2. METHODS AND MATERIALS

In this section, the experiment, the data acquisition, and the data analysis are described.

2.1. Patients

Following the approval of the ethical committee of the medical school, eight coronary artery bypass graft surgery candidates were selected (6 males, 2 females, of average age 56.2 years and the average weight of 68.3kg) and written informed consents were obtained from all selected subjects. Inclusion criteria were absent of neurological disorders such as cerebrovascular accidents and convulsions. Preoperative neurological complications (such as cerebral emboli and convulsion) caused exclusion from the study. The anesthesiologist performed Preoperative evaluation on the day before surgery. For anxiolysis, the patients were premedicated by intramuscular morphine 0.1 mg/kg and promethazine 0.5 mg/kg, 30 minutes before transfer to operating room. After arrival in operating room, electrocardiogram, pulse oxymetry, depth of anesthesia, and invasive blood pressure monitoring was established. The BIS-QUATTRO sensorTM (Aspect Medical Systems, Newton, MA) applied to the forehead of the patients before induction of anesthesia. Then 8 patients after preoxygenation with O2, were anesthetized in the same manner by intravenous thiopental sodium (5mg/kg), pancuronium bromide (0.1 mg/kg), fentanyl (5μg/kg), and lidocaine (1.5 mg/kg). After the induction of anesthesia and until cardiopulmonary bypass beginning, anesthesia continued by administration of isoflurane (1 MAC), morphine (0.2 mg/kg) and O2 (100%). During coronary artery bypass grafting under CPB, patients were anesthetized by propofol (50-150 μg/kg/min) under BIS control (40-60) and O2 (80%). For organ protection during CPB, patients were undergone mild hypothermia (31-33°C). After coronary artery bypass grafting and patients rewarming and obtaining standard CPB separation criteria, the patients gradually were weaned from CPB. After separation from CPB, anesthesia was continued by isoflurane (1 MAC) and O2 (100%) administration to the end of surgery. After surgery, patients were transported to ICU under portable monitoring and manual ventilation. In the ICU mechanical ventilation with 60% fractioned inspired oxygen and standard homodynamic monitoring were continued. In ICU and until complete recovery, the sedative regimen was intravenous morphine (2mg) if needed. In this study the raw EEG data and relative BIS index were collected during whole period of operation from operative room arrival to complete recovery in the intensive care unit.

2.2. Data Acquisition

The EEG signal was collected by using a BIS-QUATTRO Sensor™ that was composed of self- adhering flexible bands holding four electrodes, applied to the forehead with a frontal-temporal montage.

The used EEG lead was Fpz-At1, and the reference lead was placed at FP1. The sensor was connected to a BIS-X-P Monitor and all binary data packets containing raw EEG data wave signals and BIS index which is converted to binary format using an A/D converter operating with 128 Hz sampling frequency were recorded via an RS232 interface on a laptop using a Bi-spectrum analyzer developed with C++ Builder by Satoshi Hagihira [23]. The algorithms that are presented in this study were tested on these raw EEG signals.

The sensor was attached to the patient's forehead at the beginning of anesthesia and the data were collected continuously until he/she awoke at ICU. Therefore, in this study a large amount of EEG data with their BIS index was collected for each patient. Although DOA is an index beyond BIS index and BIS index needs to be validated and processed, in this paper the BIS index is considered as DOA for simplicity. Some other events such as changes of anesthesia regimen, intubations and applying CPB and transferring to ICU were recorded. Because of short acting time of thiopental sodium (approximately 15-20 sec), this part of EEG data was not analyzed.

2.3. Principal Component Analysis

PCA is a well-known technique in multivariate analysis and data mining. One of the properties of PCA is Eigenvalue Decomposition. The aim of PCA is to derive a relatively small number of decorrelated linear combinations (principal components) of a set of random zero-mean variables while also retaining the signal information as much as possible [24].

Principal Components Analysis has the applications of dimensionality reduction, determination of linear combinations of variables, feature selection, multidimensional data visualization, and identification of underlying variables.

Often components with the smallest variances called minor components (MCs) are regarded as unimportant or associated with noise, while those within which the input data have the largest variances are regarded as important. However, in some applications, the MCs are of the same importance as the PCs, which is noteworthy here. In the proposed algorithm the MCs reveal meaningful information. In the case of feature extraction and dimension reduction, PCA proposes a method based on the eigen structure of data covariance matrix. If signals are zero-mean, the covariance and correlation matrices are identical. Applying the PCA or equivalently Karhunen-Loeve transform (KLT) as a technique for eigenvectors and eigenvalues computation, the algorithm could be formulated as follows. Let X the signal to be analyzed, then

$$\hat{R}_{XX} = E\{X(K)X^T(K)\} = V\Lambda V^T \in \Re^{m\times m} \quad (1)$$

Where $\hat{R}_{XX} = E\{XX^T\}$ is the covariance matrix of zero-mean signal X and E is the expectation operator. Also, $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_m\}$ is a diagonal matrix containing m eigenvalues and $V = [v_1, v_2, ...v_m] \in \Re^{m\times m}$ are principal eigenvectors. Applying KLT as a linear transformation, principal and minor components could be

extracted as follows

$$y_P = V_S^T X \quad (2)$$

Where $X = [x_1(k), x_2(k), ..., x_m(k)]^T$ is the zero-mean input vector and $y_P = [y_1(k), y_2(k), ..., y_n(k)]^T$ is the output vector called the vector of principal components (PCs) and $V_S = [v_1, v_2, ..., v_n]^T \in \Re^{m \times n}$ is the set of signal subspace eigenvectors, with the orthonormal vectors $v_i = [v_{i1}, v_{i2}, ..., v_{im}]^T$. The vectors v_i are eigenvectors of the covariance matrix, while the variances of the PCs y_i are the corresponding principal eigenvalues. Minor components are

$$y_M = V_N^T X \quad (3)$$

Where $V_N = [v_m, v_{m-1}, ..., v_{m-n+1}]$ consists eigenvectors associated with the smallest eigenvalues [21]. The basic problem is the standard eigenvalue problem which can be formulated by the equations

$$R_{XX} v_i = \lambda_i v_i, \qquad i = 1,2,...,n \quad (4)$$

Where v_i are the eigenvectors and λ_i are the corresponding eigenvalues. Note that the above equation can be written in matrix form $V^T R_{XX} V = \Lambda$.

In the standard numerical approach for extracting the principal components, first the covariance matrix $R_{XX} = E\{XX^T\}$ is computed and then its eigenvectors and corresponding eigenvalues are extracted by one of the known numerical algorithms. However, if the input data vectors have a large dimension, then the covariance matrix R_{XX} becomes very large and it may be difficult to compute the required eigenvectors [24].

A neural network approach with adaptive learning algorithms enables us to find the eigenvectors and the associated eigenvalues directly from the input vectors without a need to compute or estimate the very large covariance matrix $\Re_{XX}$. Such an approach will be especially useful for nonstationary input data, i.e., in cases of tracking slow changes of correlations in the signals or in updating eigenvectors with new samples.

Every neuron inside the human brain acts like a small electric generator when it is active. If large numbers of neurons become simultaneously active it is possible to measure the resultant electrical effects at the scalp using an array of electrodes. Our virtual assumption is to simulate the central nervous system (CNS) as a 20-tuple source, which generate 20 signals. The EEG sensor attached to patient forehead collect different 20-tuple mixtures of these sources. Our aim is to track small changes, but due to time-domain nature of our analysis, small window lengths are more preferable. Fortunately, this would reduce the dimension. Nevertheless, large data vectors made the covariance matrix R_{XX} very large. Different window and epoch (each mixed signal is named as an epoch) lengths have been investigated and at the end, window length of 20 seconds was selected for further analysis. After that, each window is divided into 20 equal and nonoverlapping epochs. So, epoch length is equal to one second. The mentioned epochs are considered as 20 mixed signals. Then, PCA analysis and especially eigenvalue decomposition are applied to the electroencephalogram (EEG). Thus, the covariance matrix $R_{XX} = E\{XX^T\}$ is computed and then its eigenvectors and associated eigenvalues are determined. The extracted eigenvalues presented an acceptable behavior in different depths of anesthesia. Our concentration was put specifically on largest remaining eigenvalue (LRE) and smallest remaining eigenvalue (SRE). The correlation between DOA and LRE were measured with regression analysis. The same was done for SRE and DOA.

2.4. Statistical Analysis

The coefficient of determination (R^2) was calculated to evaluate the performance of different parameters and their combinations to predict DOA. Statistical significance was assumed at probability levels of $P \leq 0.05$. Our aim was to maximize the correlation between the measured sub- parameters (LRE and SRE) and BIS index, i.e., it is equivalent to nonlinear regression with ordinary least squares. Also, the correlation between BIS index and the extracted sub-parameters was investigated with the model-independent Prediction Probability (Pk) [25]. As a nonparametric measure, the Pk is independent of scale units and does not require knowledge of underlying distributions or effort to linearize or otherwise transform scales. A Pk value of 1 means that the predicting variables (LRE and SRE) always predict the value of the predicted variable (e.g., BIS index) correctly. Pk value of 0.5 means that predictors predict no better than only by chance. The Pk values were calculated on a spreadsheet using the Excel 2003 software program and the PKMACRO written by Warren Smith [25]. In the case of inverse proportionality between indicator and indicated parameters, the actual measured Pk value is 1-Pk.

Another statistical analysis used in this study was ordinal logistic regression. This regression examines the relationship between one or more predictors and an ordinal response. The index that determines the efficiency of this regression model is called "Concordant", which shows the percentage of values predicted successfully with the model.

3. RESULTS AND DISCUSSION

The results were classified in drug groups (isoflurane and propofol) and Intensive Care Unit (ICU). The correlation between the extracted parameters and BIS index (Bispectral index) is measured by means of the statistical methods described in the previous section and the results are presented.

3.1. SRE

First and foremost, it can be concluded that SRE is directly proportional to the BIS index. The scatterplots depicted in **Figure 1** and **Figure 2** show the above assert clearly.

Figure 3 compares the efficiency of SRE and LRE in predicting depth of anesthesia in different groups (ICU and drugs). In this figure, concordant was used as a statistical measure.

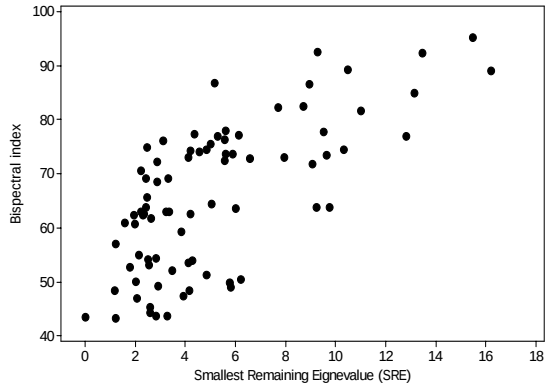


Figure 1. BIS index versus Smallest Remaining Eigenvalue (SRE) in ICU

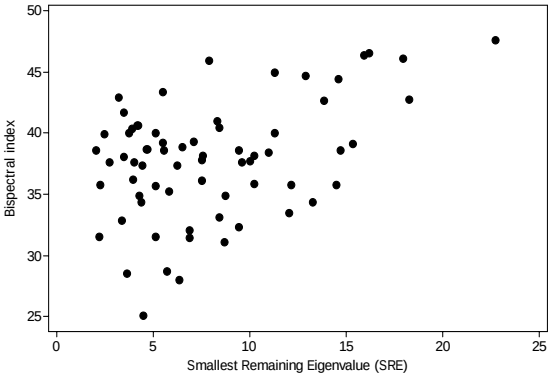


Figure 2. BIS index versus Smallest Remaining Eigenvalue (SRE) in propofol

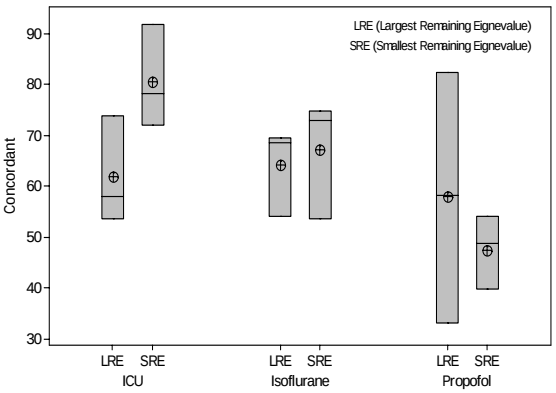


Figure 3. Concordant for different groups versus SRE and LRE

The results show the higher capability of SRE when it is used as a measure of DOA in ICU, rather than when it is used as measure of DOA in isoflurane and propofol groups. For further analysis, prediction probability (PK) was used. PK values are presented as mean ± STD in **Table 1**.

It should be noted that the P_k values are calculated for the whole BIS index range and without being divided into predetermined groups. This is the reason of the existence of smaller values in comparison with concordant values. The values in **Table 1** corroborate the results of **Figure 3**.

Table 1. Prediction probabilities of different group for SRE

Different Groups	PREDICTION PROBABILITY
ICU	65.5±8 %
Isoflurane	63±8 %
Propofol	58.5±10 %

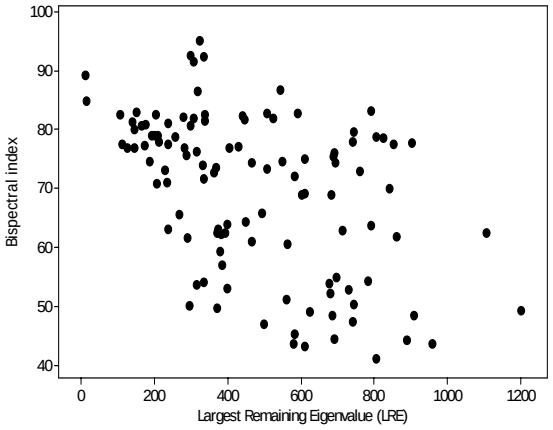


Figure 4. BIS index versus Largest Remaining Eigenvalue (LRE) in ICU

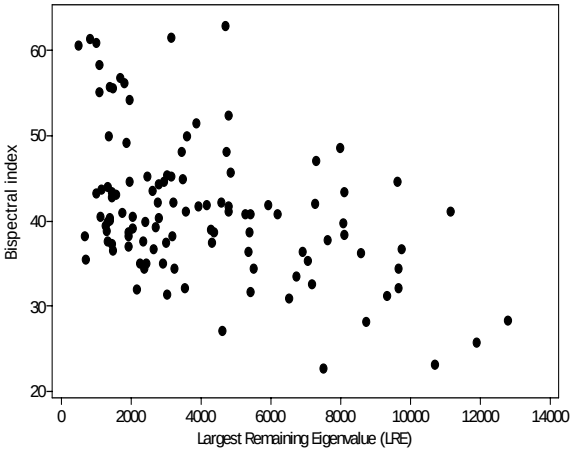


Figure 5. BIS index versus Largest Remaining Eigenvalue (LRE) in isoflurane

Table 2. Prediction probabilities of different group for LRE

Different Groups	Prediction Probability
ICU	60±6.97 %
Isoflurane	60.5±8 %
Propofol	62.5±9 %

3.2. LRE

LRE is inversely proportional to the BIS index, that is, LRE increases with the increasing depth of anesthesia. Scatterplots shown in **Figure 4** and **Figure 5** could prove the above claim.

In order to compare the efficiency of LRE algorithm in different groups we should refer to **Figure 3**. This figure indicates that there is no obvious superiority in any of these groups.

In this case prediction probability (P_K) provides us with a more precise insight. P_K values are presented as mean±STD in **Table 2**. The measured P_K values presented in **Table 2** confirm the result of **Figure 3** for LRE. The only extra information extracted from **Table 2** is that in the case of propofol induction the results are slightly better.

Finally, **Figure 3** expresses the superiority of SRE than LRE in predicting DOA in the case of ICU and isoflurane induction, and the slight superiority of LRE than SRE in propofol induction.

3.3. Relative Beta Ratio

Finally a mixture model including both LRE and SRE is compared to the model containing Relative Beta Ratio (RBR). RBR is calculated as

$$RBR = \log \frac{P_{30-47\,Hz}}{P_{11-20\,Hz}} \quad (5)$$

Where, $P_{30\text{-}47Hz}$ and $P_{11\text{-}20Hz}$ indicate the power spectral density in frequency ranges of 30−47 Hz and 11−20 Hz, respectively. A mixture model is a model in which all of the model parameters are involved in predicting the desired index. For instance, in our mixture model, BIS index is predicted using both LRE and SRE parameters. RBR is said to be the main parameter in calculating the BIS index and is referred to as an effective and critical parameter in predicting depth of anaesthesia [4]. **Figure 6** reveals a high similarity between the proposed mixture model containing both LRE and SRE parameter and the main parameter of BIS monitor which is called RBR.

In this figure, the group containing isoflurane induction is omitted. Both of the parameters (LRE and SRE) perform more satisfactorily in ICU than propofol induction.

4. CONCLUSION

A method based on PCA is proposed to estimating DOA. The principal components are extracted and the related eigenvalues are calculated as well. The smallest and largest eigenvalues express a meaningful behavior due to the changes of BIS index. So, the above parameters are selected for estimating DOA.

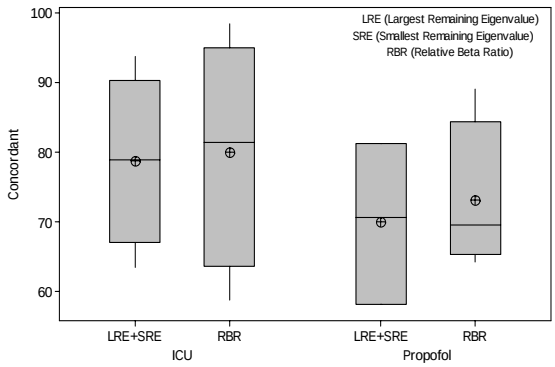


Figure 6. Comparison of RBR and proposal mixture model based on concordant values

The main result of applying the foregoing algorithm is that SRE is directly proportional to the BIS index. On the other hand, LRE is inversely proportional to the BIS index. All in all, the results show an acceptable correlation between the extracted parameters and DOA. The LRE and SRE are extremely capable of estimating DOA, especially in ICU. This is due to the ability of PCA in calculating the changes in signal energy and the changes in signals complexity as well. On the other hand, it is shown [7] that the EEG signal complexity changes meanwhile patients level of consciousness vary. Thus, PCA could be a powerful tool for predicting BIS index.

Except in propofol, the SRE parameter could predict the BIS index better than LRE. Consequently, the mixture model containing both LRE and SRE is approximately equal to a model containing RBR in predicting BIS index.

Another point that should be mentioned is that the original BIS is in fact much more than its components. The elaborate artifact rejection algorithms as well as the nature of the nonlinear function to combine the components have an important impact on the original BIS value, which were not considered in this study. Consequently, in order to improve the accuracy of the depth of anesthesia estimation, comparison against sedation scales (such as OAA/S) and drug levels is needed. The reason is that BIS is not equal to depth of anesthesia but needs to be validated for DOA assessment itself.

The work reported is preliminary. Although the results are significant, wide patient population is necessary for better evaluation. In conclusion, the approach used in this work based on the application of PCA could propose the use of PCA in estimating DOA.

ACKNOWLEDGMENT

The authors would like to thank Dr. Satoshi Hagihira due to his notification and problem solving in the application of Bispectrum Analyzer and Dr. Warren Smith that made PKMACRO available.

REFERENCES

[1] R. D. Miller, (2005) Miller's Anesthesia, Sixth edition, Elsevier Churchill Livingstone, 1227−1264.

[2] E. Freye and J. V. Levy, (2005) "Cerebral monitoring in the operating room and the intensive care unit: An introductory for the clinician and a guide for the novice wanting to open a window to the brain. Part I: The electroencephalogram", J Clin Monit Comput, 1−76.

[3] L. C. Jameson and T.B. Sloan, (2006) "Using EEG to monitor anesthesia drug effects during surgery", J Clin Monit Comput, 445−472.

[4] I. J. Rampil, (1998) "A primer for EEG signal processing in anesthesia", Anesthesiology, 980−1002.

[5] H. S. Traast and C. J. Kalkman, (1995) "Electroencephalographic Characteristics of emergence from propofol/sufentanil total intravenous anesthesia", Anesth Analg, 366−371.

[6] T. Zikov, S. Bibian, G. A. Dumont, M. Huzmezan, and C. R. Ries, (2006) "Quantifying cortical activity during general anesthesia using wavelet analysis", IEEE Trans. Biomed. Eng., Vol. 53, No. 4, 617−632.

[7] R. Ferenets, T. Lipping, A. Anier, V. Jäntti, S. Melto, and S. Hovilehto, (2006) "Comparison of entropy and complexity meas-

ures for the assessment of depth of sedation", IEEE Trans. Biomed. Eng., Vol. 53, No. 6, 1067−1077.

[8] C. Robert, P. Karasinski, C. D. Arreto, and J. F. Gaudy, (2002) "Monitoring Anesthesia using neural networks: A survey", J Clin. Monit. Comput, Vol. 17, 259−267.

[9] V. Lalitha and C. Eswaran, (2007) "Automated Detection of Anesthetic Depth Levels Using Chaotic Features with Artificial Neural Networks", J Med Syst, 445−452.

[10] R. Bender, B. Schultz, and U. Grouven, (1992) "Classification of EEG signals into general stages of anesthesia in real time using autoregressive models", Conf Proc of the 16th Annual Conference of the Gesellschaft fur Klassifikatione, University of Dortmund.

[11] D. R. Drover, H. J. Lemmens, E. T. Pierce, *et al*, (2002) "Patient State Index: titration of delivery and recovery from propofol, alfentanil, and nitrous oxide anesthesia", Anesthesiology, 82−89.

[12] J. V. Stone, (2004) Independent Component Analysis: A Tutorial Introduction, Bradford Book.

[13] K. I. Diamantaras and S. Y. Kung, (1996) Principal Component Neural Networks. Theory and Applications, Adaptive and Learning Systems for Signal Processing, Communications and Control, John Wiley & Sons Inc., New York.

[14] M. Turk and A. Pentland, (1991) "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 71−86.

[15] S. Y. Kung, K. Diamantaras, and J. Taur. (1991) "Neural networks for extracting pure/constrained/oriented principal components. In J. R. Vaccaro, editor", SVD and Signal Processing El-sevier Science, Amsterdam, 57−81.

[16] M. P. S. Chawla, (2008) "A comparative analysis of principal component and independent component techniques for electrocardiograms", Neural Computing & Applications.

[17] Hyv¨arinen, A., J. Karhunen, and E. Oja, (2001) Independent Component Analysis., John Wiley & Sons Inc., New York.

[18] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja, (1998) "Independent component analysis for identification of artifacts in magnetoencephalographic recordings", In Advances in Neural Information Processing Systems, MIT Press, Vol. 10, 229−235.

[19] S. Makeig, A. J. Bell, T. P. Jung, and T. Sejnowski, (1996) "Independent component analysis of electroencephalographic data", Advances in Neural Information Processing Systems, MIT Press, Vol. 8, 145−151.

[20] G. D. Brown, S. Yamada, and T. J Sejnowski, (2001) "Independent components analysis (ica) at the neural cocktail party", Trends in neuroscience, Vol. 24, 54−63.

[21] P. O. Hoyer and A. Hyv¨arinen, (2000) "Independent component analysis applied to feature extraction from colour and stereo images", Network: Computation in Neural Systems, Vol. 11, 191−210.

[22] Parra, L., C. D. Spence, P. Sajda, A. Ziehe, and K.-R. M¨uller, (2000) Unmixing hyperspectral data. In Advances in Neural Information Processing Systems 12, MIT Press, 942−948.

[23] S. Hagihira, M. Takashina, T. Mori, T. Mashimo, and I. Yoshiya, (2001) "Practical issues in bispectral analysis of electroencephalographic signals", Anesth Analg, Vol. 93, 966−970.

J. Biomedical Science and Engineering, 2009, 2, 16-19

Modified technique for volumetric brain tumor measurements

Yasser M. Salman[1]

[1]Egyptian Armed Forces, Correspondence should be addressed to Yasser M. Salman (yass32005@yahoo.com)

Received March 22nd, 2008; revised November 24th, 2008; accepted December 8th, 2008

ABSTRACT

Quantitative measurements of tumor response rate in three dimensions (3D) become more realistic with the use of advanced technology imaging during therapy, especially when the tumor morphological changes remain subtle, irregular and difficult to assess by clinical examination. These quantitative measurements depend strongly on the accuracy of the segmentations methods used. Improvements on such methods yield to increase the accuracy of the segmentation process. Recently, the essential modification in the Traditional Region Growing (T-RG) method has been developed and a "Modified Region Growing Method" (MRGM) has been presented and gives more accurate boundary detection and holes filling after segmentation. In this paper, the new automatic calculation of the volumetric size of brain tumor has been implemented based on Modified Region Growing Method. A comparative study and statistical analysis performed in this work show that the modified method gives more accurate and better performance for 3D volume measurements. The method was tested by 7 fully investigated patients of different tumor type and shape, and better accurate results were reported using MRGM.

Keywords: Region Growing, Modified Region growing, and Volumetric Brain Tumor Measurements

1. INTRODUCTION

More recent studies have shown that 3D quantitative imaging-based method of tumor size assessment using MRI is highly accurate in determining actual tumor size [1,2] and may be superior to clinical palpation in predicting local tumor control [3,4,5].

Manual region of interest (ROI) volumetry method is a standard approach of 3D quantitative measurements which is very precisely to detect tumor contours and it is considered to be the "gold standard" because region of interest (ROI) is segmented manually by the expert radiologists. The disadvantage of this method is, it requires intensive time because of its dependency on manual segmentation process. Segmentation of ROI in volumetric medical images is still a challenging problem, and solutions usually have been based on either model-based deformation of templates or intensity thresholding such as region growing method [6,7]. Recent studies prove that the region growing is an effective approach and less computation intensive for image segmentation especially for the homogenous regions [8,9,10,7,11]. The primary disadvantage of region growing method is the partial volume effect [12,13]. Partial volume effect limits the accuracy of MR brain image segmentation. It blurs the intensity distinction between tissue classes at the border of the two tissues types, because the voxel may represent more than one kind of tissue types [14,15]. S. Lakare [12] *et. al*, developed effective modifications in region growing technique. This modification called Modified region growing method (MRGM) used to remove the partial volume effects and to incorporate gradient information for more accurate boundary detection and filling holes occurred after segmentation.

The software implemented in this paper involves the proposed quantitative measurement of brain tumors based on MRGM segmentation and the visualization tool to monitor and reconstruct the brain tumor in 2D and 3D space. For testing and validation, the proposed MRGM method has been compared with traditional region growing method against experts' manual trace method, and the statistical analysis was performed to evaluate the proposed method against TRG and golden tracing method by experts.

In this paper, section 2 describes the settings of MR image acquisition, the details of patient population and noise reduction technique as data pre-processing. In section 3, we present the segmentation and calculation used for assessment of the brain tumor measurements, also this section describes the statistical and data analyses used to evaluate and validate the proposed method. Section 4 describes the results of brain tumor visualization in 2D and 3D spaces, the brain volume calculation using the proposed method compared with other method, and the result of statistical and data analysis. Section 5 presents the merits and demerits of the proposed method

compared to each others and concludes allover the work done in this study.

2. MATERIALS

2.1. MRI Image Acquisitions

MRI images were acquired on a 1.5T using T1-weighted contrast images. A resolution of 256x256x60 with a voxel resolution of 0.93mm x0.93mm with slice thickness of 3mm was set.

2.2. Patient Population

The study group consists of seven patients scanned with 228 MRI axial slices with biopsy histologically proved Glioblastoma Multiforme (GBM) and Low Grade Astrosytoma brain tumors types.

2.3. Pre-processing

Noise presented in the image can reduce the capacity of the region growing filter to grow large regions, or may result in a fault edges. When faced with noisy images, it is usually convenient to pre-process and enhance the image by using a noise reduction filter. Gaussian smoothing filter [16] is commonly used as an approach for noise reduction. The size of the neighborhood mask can be set by the user. The quality of the enhanced Gaussian filtered images is much better as the contrast between tumor and surrounding tissue is high as well tumors studied are of homogenous borders (regular convex shapes).

3. METHODS

3.1. Traditional Region Growing Segmentation Method

The Traditional region growing algorithm based on extraction of a connected set of pixels whose pixel intensities are consistent with the pixel statistics of a seed point. The mean and variance across 8-connected neighborhood are calculated for a seed point [16,17].

3.2. Modified Region Growing Segmentation Methods (MRGM)

To understand the basic principles behind MRGM, we first reviewed the S. Lakare, *et al*. [12]. In their work, MRGM provided for object segmentation has been implemented. This implementation allowed stable boundary detection when the gradients suffer from intersection variations and gaps.

3.3. Volume Calculation

3.3.1. Manual ROI Volumetry

Figure 1 shows the area inside the outline that was manually segmented, labeled, calculated, and multiplied by the MR slice thickness plus the interslice gap to calculate a per-slice tumor volume. The total tumor volume was obtained by summing the volume calculations for all slices.

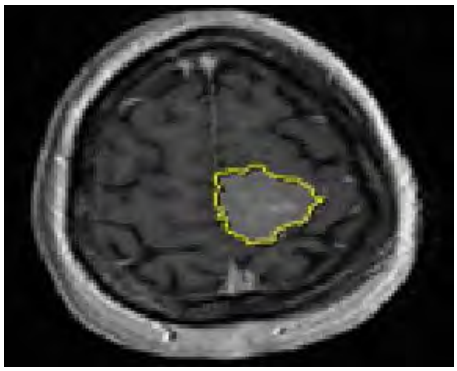


Figure 1. Manual trace method

3.3.2. TRG and MRGM Calculations for Growing Tumor

After tumor region has been segmented using both T-RG and MRGM segmentation techniques, the tumor volume calculations are performed in this segmented region. To calculate the volume of segmented tumor region, the automatic labeling of the entire volumetric tumor region has been done slice by slice and by calculating the total number of pixels into the labeled regions. Areas of the labeled region were calculated and multiplied by the MR slice thickness plus the interslice gap to obtain a per-slice tumor volume. The total tumor volume was then obtained by summing the tumor-bearing slices.

3.4. Statistical Consideration and Data Analysis

The comparative study has been done using T-RG, MRGM and Manual ROI volumetry methods. Two observers (expert radiologist and a none radiologist) independently rated each MR image twice by using manual and the two automated measurement methods. Observers have performed the comparison between inter and intra-observer reliability and image processing computational times for both methods were reported. To compare the intra and inter-observer reliability of the three measurement methods, we used the following agreement index [21], AI as follow:

$$AI = 1 - \frac{|X_a - X_b|}{(X_a - X_b)/2} \quad (1)$$

For inter-observer agreement calculation, X_a was the measurement obtained by observer 1 and X_b, the measurement obtained by observer 2 with the same technique on the same case. For intra-observer agreement calculation, X_a is the measurement made during the first trial, and X_b is the measurement made during the second trail by the same observer with the same technique on the same case. Intra-observer and inter-observer agreement indexes were calculated for each image, to increase sensitivity performance. A value of "1" indicates the perfect agreement and value of "0" indicates no agreement.

4. RESULTS

Results showed that the proposed quantitative measure-

ment of brain tumors based on MRGM has a higher accuracy and precision against traditional region growing method compared to expert manual computation. This yields to better effect in the assessment of brain tumor measurements.

4.1. PC Based Package

We improve our previous work [10] PC based software package implemented using three programming development environment, as VTK[18], ITK [20] and Visual C++, to segment, visualize and calculate the tumor volume at different instants of tumor growing or shrinking. **Figure 2** shows the result of tumor segmentation using T-RG and MRGM. **Figure 3** shows the 3D reconstruction of segmented tumor region for MRGM method, using surface reconstruction algorithm [19].

4.2. Tumor Volume Measurement Accuracy

The Relative Error (RE) for tumor volume can be calculated as Where P_q tumor volume using 3D quantitative methods (Traditional and modified region growing methods), P_m is tumor volume calculated using expert

$$RE(\%)=\left(\frac{P_q - P_m}{P_m}\right)\times 100 \quad (2)$$

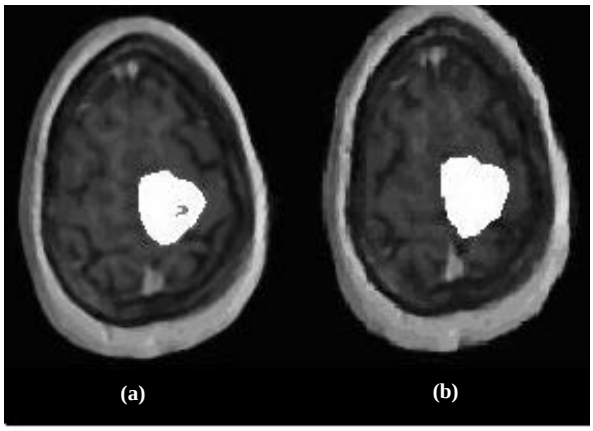


Figure 2. (a) Results of T-RG segmentation; (b) Results of MRGM segmentation

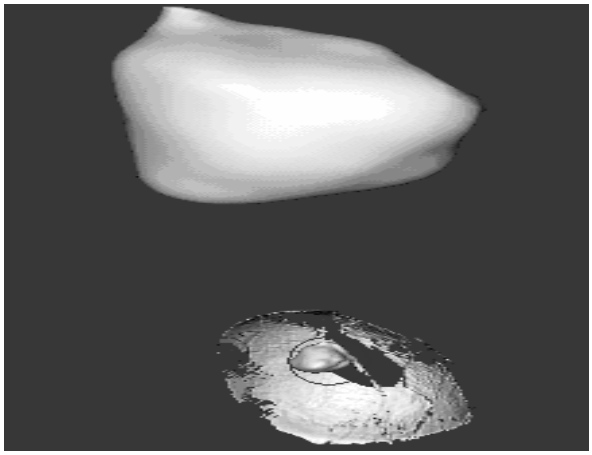


Figure 3. Extracted tumor in 3D space

manual tracing method. **Table 1** shows calculation results and their relative errors for the different quantitative methods compared with the gold standard manual method. These results had been obtained by *observer* 1 and **Figure 4** summarizes these relative errors in chart.

Also, **Table 2** shows calculation results and their relative errors for the different methods compared with the standard manual method. These results obtained by *observer* 2 and **Figure 5** summarizes these relative errors in chart.

4.3. Observer Agreements

The intra and inter-observer agreement indexes for the two observers are summarized in **Table 3**. As shown in the table, there are no significant differences in mean intra and inter-observer agreement between the manual method, traditional and modified region growing methods.

Table 1. Volumetric calculation for brain tumor in cm^3 using the different calculation methods and the relative errors for each method compared with manual segmentation method (data rated using first observer)

Cases	Volume in cm^3			Relative Error%	
	T-RG	MRGM	Manual	T-RG	MRGM
Case#1	5.587	6.281	6.125	8.7836	2.546
Case#2	8.782	9.845	10.234	14.188	3.801
Case#3	11.406	12.066	11.234	1.5310	7.406
Case#4	13.687	12.755	12.987	5.3900	1.786
Case#5	14.014	14.284	14.123	0.7717	1.139
Case#6	15.223	16.116	16.987	10.384	5.127
Case#7	16.122	16.695	16.354	1.4186	2.0851

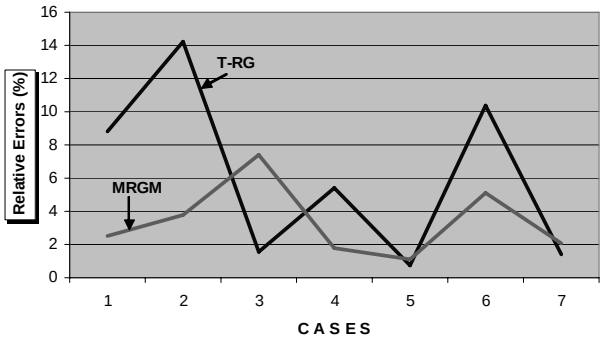


Figure 4. Relative errors for T-RG and MRGM compared with manual segmentation method (data rated using first observer)

Table 2. Volumetric calculation for brain tumor in cm^3 using the different calculation methods and the relative errors for each method compared with manual segmentation method (data rated using second observer)

Cases	Volume in cm^3			Relative Error%	
	T-RG	MRGM	Manual	T-RG	MRGM
Case#1	5.763	6.223	5.987	3.7414	3.941
Case#2	9.088	9.637	10.345	12.150	6.843
Case#3	10.836	12.963	11.897	8.9182	8.960
Case#4	13.962	12.597	12.235	14.115	2.958
Case#5	13.509	14.108	14.678	7.9643	3.883
Case#6	14.73	16.122	16.544	10.964	2.550
Case#7	14.831	16.363	16.786	11.646	2.519

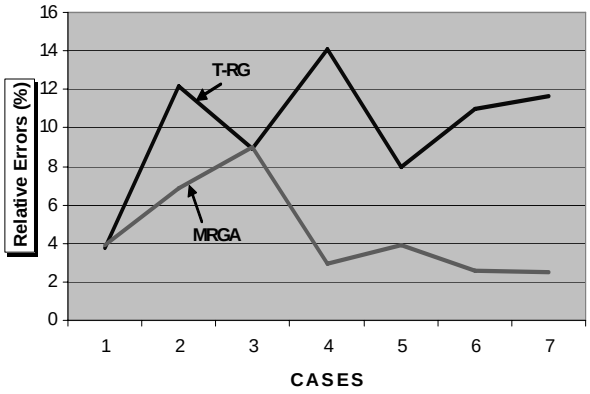


Figure 5. Relative errors for T-RG and MRGM compared with manual segmentation method (data rated using second observer)

Table 3. Results of intra- and inter observer agreement index

Index	Observer	MRGM	T-RG	Manual
Intra-Observer Agreement	1	0.9899 ±0.019	0.9374 ±.0.058	0.9665 ±0.0183
	2	0.9773 ±0.026	0.9728 ±0.0215	0.9744 ±0.0122
Inter-Observer Agreement	1,2	09689 ±0.0342	0.9587 ±0.0207	0.9559 ±0.0237

5. DISCUSSION AND CONCLUSION

Recent attention has been given to improve the segmentation methods in order to increase brain tumor measurements accuracy [8]. In this paper the essential modifications has been implemented in the region growing algorithm and presented as "Modified Region Growing Method" (MRGM). These modifications overcome the partial volume effect artifacts. Hence, applying MRGM to segment brain tumors will increase the accuracy of the volumetric measurements of the brain tumors. A volumetric measurement based on T-RG, MRGM and manual segmented ROI volumetry method have been applied in MR volumetric data total of 228 MRI slices of Glioblastoma Multiforme (GBM) and Low Grade Astrosytoma brain tumors scanned from 7 patients. The results of comparative study showed that the MRGM produces lower relative errors than T-RG method. Also, it has a no significant differences in inter- and intra-observer agreement index. These results ensure that the accuracy of the volumetric measurements of the brain tumor have been improved using MRGM which yields to great effects in many applications in tumor prognosis and therapy such as early signs of treatment failure in radiotherapy and chemotherapy to avoid unneeded higher doses of radiation to patient, tumor growth rate and early detection of tiny changes in tumor size in which it is difficult to be detected in traditional visual metric and clinical examination measurements.

ACKNOWLEDGMENT

We would like to acknowledge ISBR. They support us to obtain the data. Data was provided by the Center for Morphometric Analysis at Massachusetts General Hospital.

REFERENCES

[1] H. Hricak, C. G. Lacey, L. G. Sandles, Y. C. F. Chang, M. L. Winkler, J. L. Stern. (1988) Invasive cervical carcinoma: Comparison of MR imaging and surgical findings, Radiology 166: 623−631.

[2] K. Hatano, Y. Sekiya, H. Araki, *et al*. (1999) Evaluation of the therapeutic effect of radiotherapy on cervical cancer using magnetic resonance imaging, Int. J. Radiat. Oncol. Biol. Phys. 45: 639−644.

[3] N. A. Mayr, W. T. C. Yuh, *et al*. (1997) Tumor size evaluated by pelvic examination compared with 3-D MR quantitative analysis in the prediction of outcome for cervical cancer, Int J Radiat. Oncol. Biol. Phys. 39: 395−404.

[4] H. Hricak, J. Quivey, *et al*. (1993) Phillips T. Carcinoma of the cervix: Predictive value of clinical and magnetic resonance (MR) imaging assessment of prognostic factors, Int. J. Radiat. Oncol. Biol. Phys. 27: 791−801.

[5] N. A. Mayr, E. T. Tali, *et al*. (1993) Cervical cancer: Application of MR imaging in radiation therapy, Radiology 189: 601−608.

[6] M. Kass, A. Witkin, and D. Terzopoulos, (1987) "Snakes: Active contour models," Int. J. Computer Vision, Vol. 1, No. 4, 321−331.

[7] R. Adams and L. Bischof, (1994) Seeded Region Growing, IEEE Transactions on Image processing, Vol.16, No.6, 641−47.

[8] F. Vincent, H. Chong, J. Y. Zhou, B. James, K. Khoo, J. Huang, T. K. Lim. (2004) Tongue carcinoma: Tumor volume measurement, Int. J. Radiation Oncology Biol. Phys., Vol. 59, No. 1, 59−66.

[9] Y. M. Salman, A. M. Badawi, M. A. Assal, S. M. Alian, (2005) New automatic technique for tracking brain tumor responsem , International Conference on Biological and Medical Physics, UAE.

[10] Salman Y. M., A. M. Badawi, (2005) Validation Techniques for Quantitative Brain Tumor Measurements, The 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, China, pp. 7048−7051.

[11] S. C. Zhu and A. Yuille, (1996) Region competition: Unifying snakes, region growing and Bayer/MDL for multiband image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 9, pp. 884−900.

[12] M. Sato, S. Lakare, M. Wan, and A. Kaufman, (2000) A Gradient Magnitude Based Region Growing Algorithm For Accurate Segmentation, In Proc. International Conference on Image Processing, Vol.3, 448−451.

[13] S. Lakare, (2000) 3D Segmentation Techniques for Medical Volumes, Center of Visual Computer, state university of NY, Stony Brooks.

[14] Jauhiainen, Tommi, Jarvinen, *et al*., (1998) MR Gradient Echo Volumetric Analysis of Human Cardiac Casts: Focus on the Right Ventricle, Journal of Computer Assisted Tomography, Vol.22, No.6, 899−903.

[15] J. M. Links , L. D. Beach , B. Subramaniam, (1998) Edge complexity and partial volume effects, Journal of Computer Assisted Tomography, Vol. 22, No.3, 450−458.

[16] R. C. Gonzalez and R. E woods, (1992) Digital Image Processing, Addison-Wesley, USA.

[17] Y. L. Chang, X. Li, (1994) Adaptive Image Region-Growing, IEEE Trans. on Image Processing, Vol. 3, No. 6, 868−872.

[18] W. Schroeder, K. Martin, and B. Lorensen B, (1998) the Visualization Toolkit. New Jersey, Prentice Hall.

[19] W. E Lorensen, H. E Cline, (1987) Marching cubes: a high resolution 3D surface construction algorithm, Computer Graphics (SIGGRAPH 87 Proceedings), Vol. 21, 1693−169.

[20] L. Ibanes, W. Schroeder, L. Ng, (2003) Insight Segmentation and Registration Toolkit (ITK) Software Guide.

[21] B. N. Joe, *et al*., (1999) Brain Tumor Volume Measurement: Comparison of manual and semi automated methods, Radiology, No.212, 811−816.

J. Biomedical Science and Engineering, 2009, 2, 20-23

Researches on a practical system for concentration detection of human exhaled CO_2 gas

Rong-Guo Yan[1], Bin Ge[1], Hai-Ming Xie[1]

School of medical instrument and food engineering, University of Shanghai for Science and Technology, Shanghai 200093. Correspondence should be addressed to Rong-Guo Yan (yanrongguo@usst.edu.cn), Tel: +86 21 55271115.

Received July 30th, 2008; revised November 12th, 2008; accepted November 19th, 2008

ABSTRACT

In clinics, especially in the emergency department, carbon dioxide (CO_2) is considered as the sixth vital sign in evaluating a patient's medical status. However, its wide applications in developing countries are dissatisfactory due to their high cost compared to their economic growth. In order to develop a practical system for concentration detection of human exhaled CO_2 gas for our own, we studied the detection principle based upon the non-dispersive infrared (NDIR) measuring technique and related Lambert-Beer's equation carefully, gave out our functional circuit design of the system, and provided the corresponding graphical user interface (GUI) for CO_2 detection. Finally, the experiment shows that it may be a practical system, and will give benefits for the development of our medical care in the future.

Keywords: CO_2 Concentration Detection, NDIR, Lambert-Beer's Equation, Detection Module, Instrument Module

1. INTRODUCTION

As we know, carbon dioxide (CO_2), as a gas byproduct produced by human cellular metabolism and finally exhaled out into the air through our noise, can be measured to reflect the systematic functions of metabolism, circulation and ventilation [1,2,3]. Especially, through the noninvasive monitoring of end-tidal carbon dioxide ($ETCO_2$), which is also known as capnography measuring the maximum carbon dioxide level reached at the end of each breath [4,5,6], it can provide the doctor or the nurse the sixth vital sign, as important as five other signs, i.e., respiration, heart rate, temperature, blood pressure and pain in evaluating the medical status of the subjects. In America, capnography, other than monitoring pulse oxygen saturation, has already been considered as one of the "basic standards" for monitoring for all the subjects receiving anesthesia.

The most common principles for CO_2 detection are based upon non-dispersive infrared (NDIR) measurement or based upon chemical reactions. Commonly, chemical CO_2 gas sensors with sensitive layers have the principal advantage of very low energy consumption and can be reduced in size to fit into microelectronic based systems. However, compared with the NDIR measurement principle [7], chemical CO_2 sensors face major obstacles like short-and long term drift effects and a rather low overall lifetime [8,9].

Although the technology carried out for monitoring expired CO_2 gas based upon NDIR is not a new technology [10,11,12], its wide applications, such as in mechanical ventilator, anesthesia machine, and etc., in developing countries are still left in a dissatisfactory condition due to its high cost compared to their economic growth. However nowadays, with the development of integrated circuits and non-dispersive infrared (NDIR) sensing technology, its utilization by clinicians outside the operating room or anesthesia realm has also grown, especially in the emergency department.

Thus, this paper will present the principle for CO_2 gas concentration measurement; introduce its functional circuit design that we applied; and provide its current experimental result to show its practicability for use.

2. PRINCIPLE FOR DETECTION

The method we used for concentration detection of CO_2 gas is based upon the NDIR measuring technique and a so-called well-known Lambert-Beer's equation [13]. The equation is given as follow:

$$I = I_0 * e^{-kcl} \quad (1)$$

Where,

I_0 = the intensity of light incident on the sample;

I = the measured intensity of light after the sample;

k = the absorption coefficient of the analyte gas at the characteristic wavelength (cm^2);

c = the concentration of the analyte gas ($1/cm^3$);

l = the path-length or the distance that the light traverses the sample gas (cm).

Let

The project is supported by the research award fund for selectively culturing excellent young teachers in higher educations in shanghai (grant No. slg-07045).

$$A = -\ln(\frac{I}{I_0}) \quad (2)$$

Then according to equation (1), there exists,

$$A = kcl \quad (3)$$

In the equation (3), A is expressed in terms of an absorbance. As we know, the parameter l is already determined when the detection system is designed, and the parameters A and k can be measured in the experiment, then CO_2 gas concentration c can be calculated accordingly.

As **Figure 1** shows, the chamber is made of aluminum; the left-side inner surface adopts arc shape, being specially polished and shot-blasted to increase the intensity onto the right-side detector.

And, basically the concentration detection system for CO_2 gas consists of the following four components including: (1) a pulsed infrared (IR) light source at the left end of the chamber or sample room, whose excitation frequency is about 1Hz to provide a non-dispersive infrared flashing source; (2) an airway chamber having l path length with a gas inlet and a gas outlet at the up side edge as **Figure 1** shows. The gas inlet is connected to our nose with a nose-oxygen-tube like pipe, while the gas outlet gives a way for CO_2 gas flowing out of the chamber, and the chamber will be full of CO_2 gas when performing detection; and (3) an IR detector at the right end of the chamber. The detector is in fact a dual channel PerkinElmer thermopile sensor housing with two window openings (channel T1 and T2)-each window specialized to be equipped with band-pass filters for gas detection [14]. One optical window opening (T2) senses light at a specific wavelength of 4.26μm, a predominant absorption band of CO_2 gas (see **Figure 2**), while the other channel (T1) actually carries a band-pass wavelength centered at 4.0μm, other than CO_2 gas absorption band at 4.26μm, acting as reference [14]. Thus except ground signal, three other signals that the thermopile detector gives are a reference signal (Ref_sig. in **Figure 1**) for reference use, a real measured signal (Msd_sig. in **Figure 1**), and a temperature signal (Tmp_sig. in **Figure 1**) giving the real temperature inside the thermopile sensor for later temperature compensation use. Basically, these millivolts (mV) level signals should be amplified, conditioned via a detection module, and then the actual CO_2 concentration can be calculated according to the Lambert-Beer's equation by the MCU in an instrument module. Functions of the detection module and the instrument module are described in section 3.

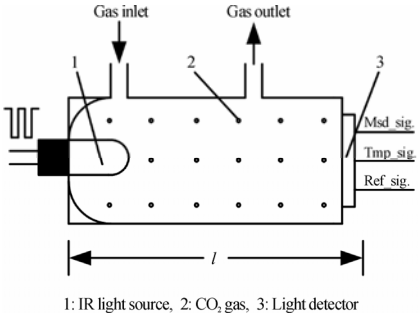


Figure 1. NDIR optical detection system

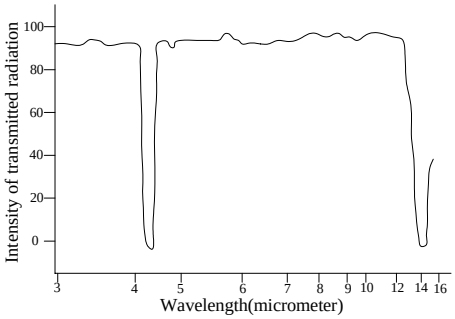


Figure 2. Optical spectral absorption bands of CO_2, 4.26μm is its predominant absorption peak, and the IR light in between this narrowband can pass the interference filter

3. FUNCTIONAL CIRCUIT DESIGN

To accomplish CO_2 gas concentration detection as the principle indicates, we designed a practical detection system including two functional modules: the detection module and the instrument module.

3.1. The Detection Module

The detection module performed signal amplification, conditioning and voltage following of four signals that are needed to estimate gas concentration of exhaled CO_2 as the Lambert-Beer's equation indicates, including one room temperature signal (Room_sig.) and three other signals from the thermopile detector, i.e., a reference signal (Ref_sig.), a real measured gas concentration output (Msd_sig.) and a temperature output (Tmp_sig.) for later temperature compensation (See section 2). Finally, these four signals will be provided for the instrument module for calculation.

3.2. The Instrument Module

The instrument module mainly performs (1) analog-to-digital (A/D) conversion of four signals including a Room_sig., a Ref_sig., a Msd_sig., and a Tmp_sig. from the detection module; (2) calculation gas concentration according to the Lamber-Beer's equation; (3) concentration display on the liquid crystal display (LCD) of the acquired CO_2 gas at the sample time; (4) giving warnings and alarms; (5) configuration by a master PC using a RS-232C universal serial interface. Thus, the main components in the instrument module are: (1) an ATmega16L micro controller unit (MCU), which is a high-performance, low-power AVR® 8-bit microcontroller with an advanced reduced instruction set computer (RISC) architecture and 16KB in-system programmable flash program memory [15]. The MCU is good for programming and debugging using C language; (2) a master PC performing to finish configuration of the system, such as serial number information, warning limits (upper and lower limit) of gas concentration, calibration and temperature compensation information, and etc., through the RS-232C interface; (3) a DS1307 real time clock (RTC), which is a low-power, full binary-coded decimal

clock/calendar. Address and data are transferred serially through a two-wire interface (TWI), bidirectional bus that Atmega16L provides. The current time can be set through the RS-232C based graphical user interface written in computer advanced languages, Visual C++, MATLAB, and etc., for example; (4) a beeper and a HF12232F-based LCD, which dynamically gives current time, gas concentration information at current time, and warnings, as well as other information that can be visually prompted on the LCD. The LCD interfaces with the MCU through the serial peripheral interface (SPI) interface that the Atmeag16L provides. (5) an A/D conversion of four signals from the detection module, including one room temperature signal (Room_sig.) and three signals from the thermopile, i.e., a reference signal (Ref_sig.), a real measured output (Msd_sig.) and a temperature output (Tmp_sig.) for later temperature compensation use. The functions of the instrument module can be seen in **Figure 3**.

4. RESULTS

Figure 4 is a graphical user interface (GUI) for CO_2 concentration detection, which is written in MATLAB. MATLAB is selected in the paper for its simple programming, rich graphic facilities, built-in functions, and extensive toolboxes. This is especially suitable for engineering professionals like us. From the GUI, we can select serial port (COM1, COM2, and etc.) and baud rate (9600bps, 115200bps, and etc.) to establish communications between the master PC and the Atmeaga16L MCU to perform gas concentration detections using the button *SendFF*, where hexadecimal 0*xFF* is used as a handshaking string between two CPUs. Moreover, the button *Load Data* can load any related saved data. In the graphical window, the real measured output (corresponding to Msd_sig.) and its reference output (corresponding to Ref_sig.) in an experimental test are plotted with different signs. Their minimal value, maximal value, and averaged peak-to-peak difference in the sample period are given in the right frame box.

In order to eliminate errors from intensity variation of the light, and to decrease sensor drift due to temperature variations, the ratio between peak-to-peak difference of the real measured output (Msd_sig.) and the reference output (Ref_sig.) is calculated, and this value is proportional to the current gas concentration. Actually, a relation curve between this ratio and the gas concentration has already established when performing calibration of the system at a discrete concentration step, and has been stored

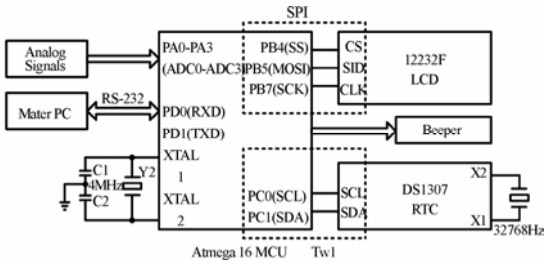


Figure 3. The instrument module

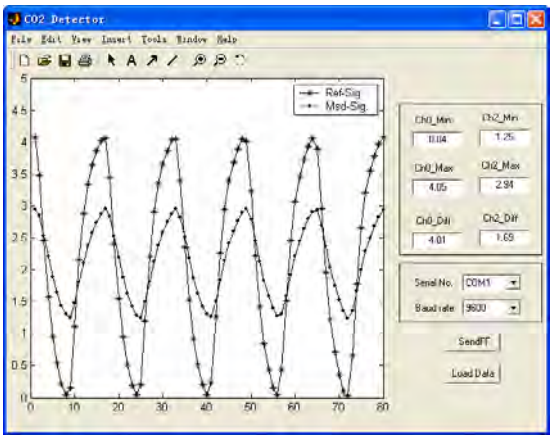


Figure 4. GUI for CO_2 detection written in MATLAB

in the EEPROM inside the MCU. The measured CO_2 gas concentration can then be calculated using the ordinary least-square interpolation method in between two neighboring discrete concentration values designed in calibration period.

As a human machine interface, the calculated real concentration is also shown on the LCD.

5. FUTURE WORKS

In the paper, we introduced the method of concentration detection for human exhaled CO_2 gas, described the functional module (the detection module and the instrument module), and provided the GUI for CO_2 detection written in MATLAB.

The experiment result shows that it is a practical system for medical CO_2 concentration detection whose measurement range is from 0 to 3000 ppm (part per million; 1000ppm =0.1%) with ±5% of reading accuracy as desired, and it will give benefits for the development of our medical care in the future, as expected.

However, there are still remaining challenges ahead of us before real practical use. Firstly, the characteristics of the system need to be verified for relatively long time. Secondly, a face mark for the patients receiving supplemental oxygen should be considered in the mechanical design. To provide additional functions, such as remote control of the system are also needed to be considered in the future for specific environments.

REFERENCES

[1] J. S. Gravenstein, D. A. Paulus, T. J. Hayes (1989) Clinical indications, em: Gravenstein JS, Paulus DA, Hayes TJ-Capnography in clinical practice. Butterworth Publishers, Boston.

[2] M. S. Bhende (1999). Capnography in the pediatric emergency department. Pediatric Emergency Care, 15:64−69.

[3] A. B. Sanders (1989). Capnometry in emergency medicine. Annals of Emergency Medicine, **18**:1287−1290.

[4] R. N. N. Gilboy, R. Michael, C. R. N. A. Hawkins, MS (2006). Noninvasive monitoring of end-tidal carbon dioxide in the emergency department. Advanced Emergency Nursing Journal, 28(4): 301 −313.

[5] M. S. Bhende (2001). End-tidal carbon dioxide monitoring in pediatrics: concepts and technology. Journal of Postgraduate Medicine. 47(2): 153−156.

[6] A. B. Sanders, K. B. Kern, C. W. Otto, M. M. Milander and G. A. Ewy (1989). End-tidal carbon dioxide monitoring during cardiopulmonary resuscitation: a prognostic indicator for survival. The Journal of the American Medical Association. 262 (10): 1347−1351.

[7] Th. Lang, H. D. Wiemhöfer and W. Göpel, Conf. Proc. Eurosensors IX, Stockholm (S) (1995); Carbonate based CO_2 sensors with high performance, Sensors and Actuators B, 34, 1996, 383−387.

[8] http://en.wikipedia.org/wiki/CO2_sensor.

[9] Reliable CO_2 sensors based with silicon-based polymers on quartz microbalance transducers, R. Zhou, S. Vaihinger, K. E. Geckeler and W. Göpel, Conf. Proc. Eurosensors VII, Budapest (H) (1993); Sensors and Actuators B, 18−19, 1994, 415−420.

[10] T. Ahrens & C. Sona, (2003). Capnography application in acute and critical care. AACN Clinical Issues, 14 (2), 123−132.

[11] K. Bhavani-Shankar, (2006). Capnography: A comprehensive educational website. Retrieved July 23, 2006.

[12] M. Eichbrecht, (2000, December/2001, January). Capnography: The ventilation vital sign. Retrieved August 2, 2006.

[13] J. D. J. Ingle and S. R. Crouch, Spectrochemical Analysis, Prentice Hall, New Jersey.

[14] http://www.perkinelmer.com/.

[15] http://www.atmel.com/.

J. Biomedical Science and Engineering, 2009, 2, 24-29

In vivo testing of a bone graft containing chitosan, calcium sulfate and osteoblasts in a paste form in a critical size defect model in rats

Jerome G. Saltarrelli Jr.[1], Debi P. Mukherjee[1]*

Louisiana State University Health Sciences Center Department of Orthopaedic Surgery, 1501 Kings Highway, Shreveport, Louisiana 71130, * Corresponding author to Debi P. Mukherjee: (dmukhe@lsuhsc.edu)

Received September 26th, 2008; revised November 12th, 2008; accepted November 13th, 2008

ABSTRACT

Bone loss associated with musculoskeletal trauma or metabolic diseases often require bone grafting. The supply of allograft and autograft bones is limited. Hence, development of synthetic bone grafting materials is an active area of research. Chitosan, extracted from chitin present in crawfish shells, was tested as a delivery vehicle for osteoblasts in a 2-3 mm size defect model in rats. Twenty-seven male Lewis rats, divided into three groups with sacrifice intervals of 3, 6 & 9 months were used. In the experimental samples, a critical size defect was filled with chitosan bone graft paste and fixed with a plate, while in the operated control group, a critical size defect was repaired only by a plate (no paste was applied). An unoperated control group was also included. Bone growth was evaluated histologically by examining undecalcified and decalcified stained sections. The femurs were also examined non-destructively by micro-computed tomography (μCT). Defects filled with chitosan bone graft paste demonstrated superior healing across all time periods compared to unfilled defects as examined by histology and micro-computed tomography. Crawfish chitosan has successfully been used as a cell delivery system for osteoblasts for use as a synthetic bone graft material.

Keywords: Chitosan, Synthetic Bone Graft, Cell Delivery, Histology, Animal Model Running Head: Chitosan Based Synthetic Bone Graft Material

1. INTRODUCTION

Bone loss associated with musculoskeletal trauma or metabolic diseases often require bone grafting. Autograft or allograft bones are limited in supply. Therefore, development of synthetic bone graft materials is an active area of research.

Chitin is a polysaccharide that exists in fungi, exoskeleton of insects and the outer shell of crustaceans [1,2,3]. It is biocompatible [4], osteoconductive [5], antimicrobial [6,7], biodegradable [8], non-toxic [9], haemostatic [10], fungicidal [11], and the second most abundant natural polysaccharide on earth [12]. Removal of the acetyl group from chitin forms chitosan. Chitosan is more useful due to the presence of amino groups that impart a positive charge to the molecule. Chitosan has been investigated in a number of biomedical applications due to its purity coupled with a positive charge [13,14,15,16]. This positive charge interacts with cells or can act as a binding site for other functional groups thereby expanding the role of the chitosan molecules.

Recently, a new patented process has been developed to purify chitosan from crawfish shells [17]. The objective of this current study was to evaluate this crawfish chitosan as a delivery system for osteoblasts to promote bone growth in a 2-3mm defect in rats. For this purpose, prepared crawfish chitosan was compounded with calcium sulfate to form a paste. The bone growths at 3, 6 & 9 months were estimated by histology and microcomputed tomography (μCT). It was hypothesized that chitosan provides a delivery system, keeping the cells in the defect area for a longer period of time allowing defect repair.

2. MATERIALS AND METHODS

2.1. Chitosan Extraction

Chitosan was extracted from crawfish shells following a patented process [17]. Briefly, shells were first washed, and dried in an 80°C oven for 48 hours. Following drying, the shells were immediately quenched in liquid nitrogen then treated by: (1) 3.5% NaOH to remove proteins; (2) 1N HCl to remove minerals; and (3) 50% NaOH to remove acetyl groups. The extracted chitosan was purified through a 12,000−14,000 Dalton dialysis membrane and dried.

2.2. Preparation of Osteoblasts

Stromal osteoblast cells were obtained from the marrow

of young adult male (125−149g) Lewis rats. Following euthanasia by CO_2 asphyxiation, femora were aseptically excised, cleaned of soft tissue, washed in DMEM+ antibiotic-antimycotic (concentration of antibiotic- antimycotic was 10 times the normal amount used in cellular media). The metaphyseal ends were cut off and the marrow flushed from the midshaft with 5ml of media (DMEM+10% FBS+antibiotic-antimyctotic) using a syringe equipped with a 22-gauge needle and collected in a sterile test tube. Cell clumps were broken up by repeated pipetting of the cell suspension. The cells were centrifuged at 1200 rpm for 10 minutes. The cell pellet was resuspended in media (DMEM+10% FBS + antibiotic-antimycotic+10% IL-3) and seeded in a flask. On the following day, the media was removed and the cells were washed with 10x concentration of antibiotic- antimycotic in PBS; and complete media was introduced into the flask. Cells were fed every two to three days until confluent, and then trypsinized.

2.3. Paste Formulation

The ratio of chitosan to calcium sulfate was 1:4, ideally 0.125g of crawfish chitosan to 0.5g calcium sulfate. The actual average paste contents were 0.5030g of calcium sulfate and 0.1234g of crawfish chitosan with an average osteoblast concentration of 4.06 x 10^5/ml. The calcium sulfate ($CaSO_4$) and extracted chitosan, after sterilization under dry heat, was mixed with 1ml of osteoblast cell suspension (10^6 cells/ml) to form a paste. This was accomplished immediately prior to implantation under the sterile field. The same procedure was used in a previous study repairing a cranial critical size defect in rats [18].

2.4. Experimental Design and Animal Surgical Procedure

Twenty seven male Lewis rats were divided into 3 groups: (1) operated control (2) experimental and (3) non-operated control. Animals in the operated control and experimental groups went through a surgical procedure which created a segmental defect of 2-3mm which is a modification of the 8mm critical size segmental defect model [19,20]. This modification was necessary because a four-hole 23mm long plate was used, which had a solid section of about 4mm between the second and third hole covering the femoral defect. This modification in defect size was necessary for proper femoral repair. The operated control group animals received only plate fixation. The experimental group received both plate fixation and bone graft paste. The unoperated group was a control. Each group, of three animals per time period, was studied for 3, 6 and 9 months.

The surgical procedure is as follows: animals were anesthetized with 1.5% Isoflurane, shaved and cleaned with 70% alcohol and Betadine. An incision was made dorsally to the femur and a 4-hole titanium plate was applied. In both operated groups (experimental and operated control), once the plate was installed a 2-3 mm size defect was applied to the femur using a burring bit on a Dremel tool. In the operated control group, the incision was closed. In the experimental group, the chitosan-bone graft paste was applied to the defect. Post surgery, animals were housed in individual cages and monitored for surgical complications (foot drop, lethargy, loss of use and pain). Upon euthanizing by CO_2 asphyxiation, operated and contra-lateral control femurs were removed and stored frozen until testing in a 0.9% NaCl solution.

2.5. Undecalcified Histological Sample Preparation and Staining

All undecalcified histological samples were fixed in 10% formalin for at least 2 weeks. Following formalin fixation, the samples were taken through an increasing gradient of 2-hydroxyethyl-methacrylate and nanopure water then a transition of 2-hydroxyethyl-methacrylate and Technovit 7200 and finally 100% Technovit 7200. Samples were embedded in resin by submersion in fresh Technovit 7200 and exposure to an ultraviolet light (λ=450nm). Ultraviolet light causes polymerization of Technovit 7200. Sample blocks were attached to slides using Technovit 7210LVC, cut using an Exakt band saw and ground into 10-20µm thick slices using an Exakt surface grinder (Exakt Technologies, Oklahoma City, OK).

All undecalcified histological sections were stained using Goldner's trichrome method whereas bone stains green, muscle stains orange, and fibrous tissue stains red.

2.6. Non-destructive Evaluation by Micro-computed Tomography (µCT)

µCT was preformed (courtesy of Animal Resources, Louisiana State University Health Sciences Center, Shreveport, Louisiana) on a MicroCAT II; model MCII-UAAI following standard operating procedure. Each femur, along with the contra-lateral side, was run for 512 scans along the defect site. This data was compiled into 2-D and 3-D images and qualitative information on bone growth was collected.

3. RESULTS

3.1. Undecalcified Histological Sectioning

Three, six and nine months data for operated control and experimental samples are shown in **Figure 1**, A through I and an example of unoperated mature bone is shown in **Figure 1-J**. Goldner's Trichome stain was used which stains bone green and muscle and fibrous tissues red. All photographs shown were taken at 4x magnification under a light microscope.

3.2. Histological Quantitative Data

From bone pixelation, bone percentages from each group can be compared (**Figure 2**). In each time interval, more bone was present in the experimental group compared to

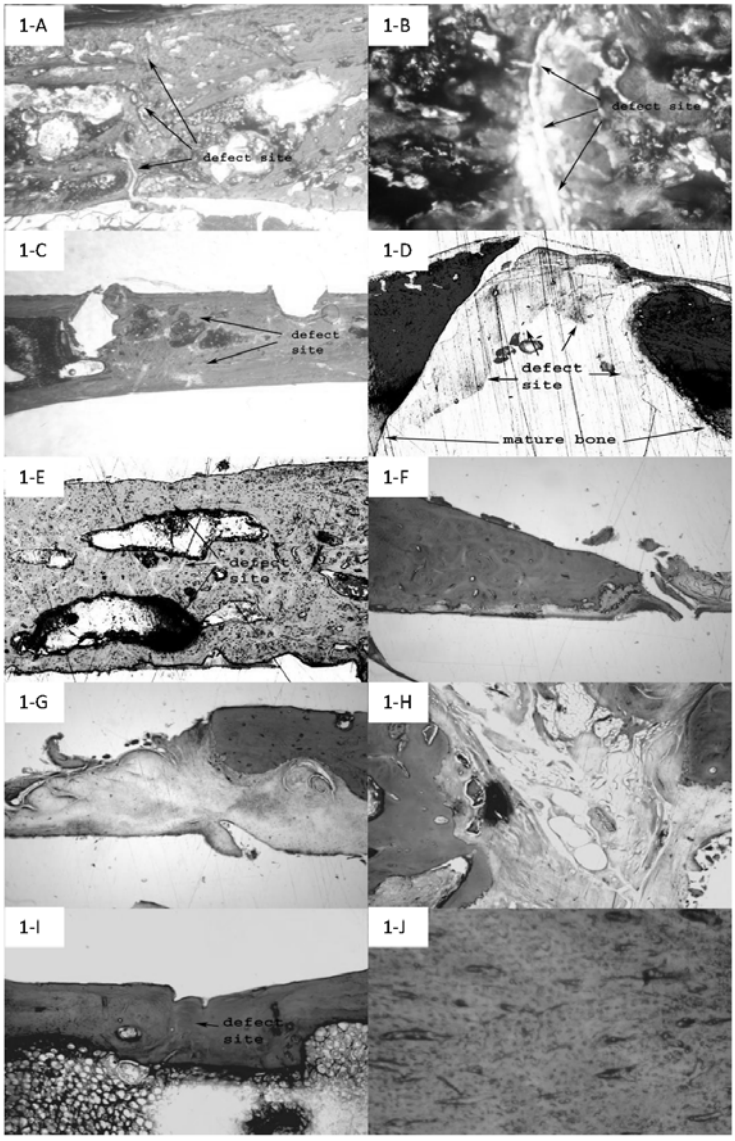


Figure 1. Undecalcified Histology. (Figure 1-A & 1-B shows the presence of a defect after 3 months without chitosan/ Plaster of Paris bone graft paste (operated control group) (1-A=5x; 1-B=20x). Figure 1-C shows experimental samples after 3 months, the defect was repaired with beginning of intermedullary canal. After 6 months, significant fibrous growth remained in the operated control group (Figure 1-D = 20x). In the 6 month experimental group, rapid repair has occurred (vs. operated control) with the presence of large voids (Figure 1-E = 20x). Figure 1-F, G & H (20x) show the 9 month operated control group with infiltration of fibrous tissue which inhibits bone growth. Figure 1-I (5x) is the experimental group after 9 months with near complete bone growth. Figure 1-J is an unoperated control (20x).)

the operated control group. During the three month interval, 40.23% bone growth was calculated in the experimental group while only 29.43% was calculated in the operated control group. At the six month time interval, 35.58% bone growth was calculated in the experimental group and 30.09% bone growth was recorded in the operated control group. As expected, nine month time interval demonstrated the highest percentage of bone growth in both groups. The experimental group experienced 42.82% bone growth while the operated control group experienced 40.16%. In all the sacrifice intervals the degree of bone growth was slightly higher for the experimental group than that of the operated control group. After three months, the experimental samples demonstrated 26.85% more bone in the defect site than the operated control samples. After six months, the experimental samples demonstrated 15.42% more bone and after nine months, 6.2% more bone was evident in the defect site. In depth statistical analysis of the data was not possible due to limited number of animals.

3.3. Non-destructive Testing by Microcomputed Tomography (µCT)

µCT was preformed on a MicroCAT II; model MCII-UAAI following standard operating procedure. Each femur, along with the contra-lateral side, was run

for 512 scans along defect site. This data was compiled into 2-D and 3-D images. We were unable to extract the quantitative information like bone volume or connectivity density of trabecular bone from the µCT images since we did not possess the appropriate software. Therefore we qualitatively looked at the 3-D rendered µCT images for bone repair. The 3, 6 and 9 month data are shown in **Figures 3A-3F** while **Figure 3-G** displays unoperated control data. MicroCT data showed gradual bone growth even at the 3 month period as seen with the computerized volumetric rendering data (3-B). At six and

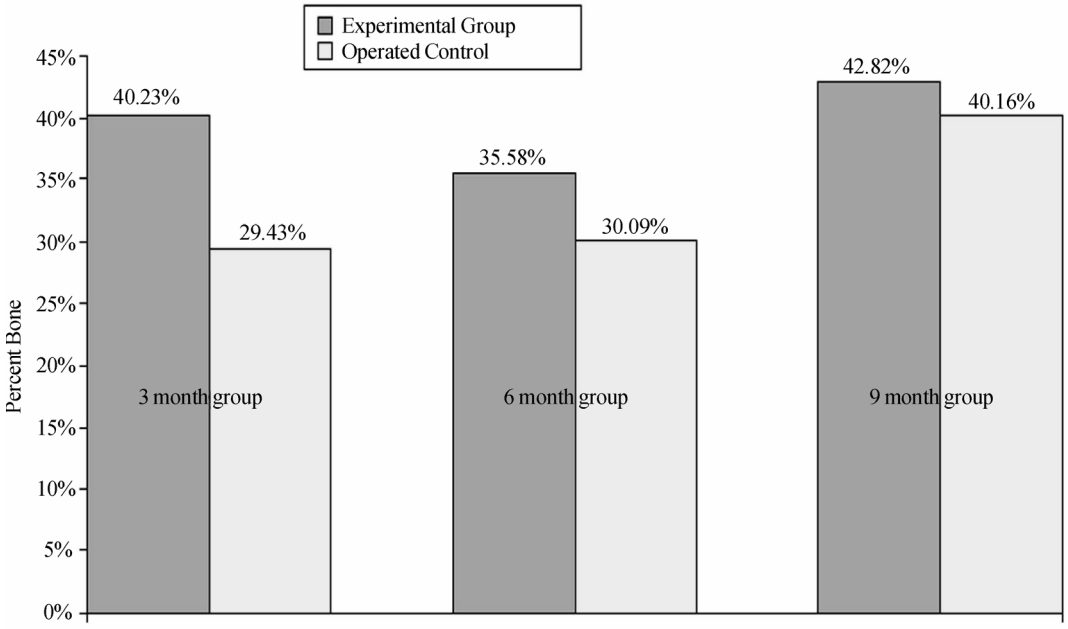


Figure 2. Bone Growth Fractions. (Bone growth fractions were calculated by pixel quantification within the defect site. In all time periods, more bone is present in the experimental samples versus the operated control samples, as represented here. Error bars represent variation between specimens.)

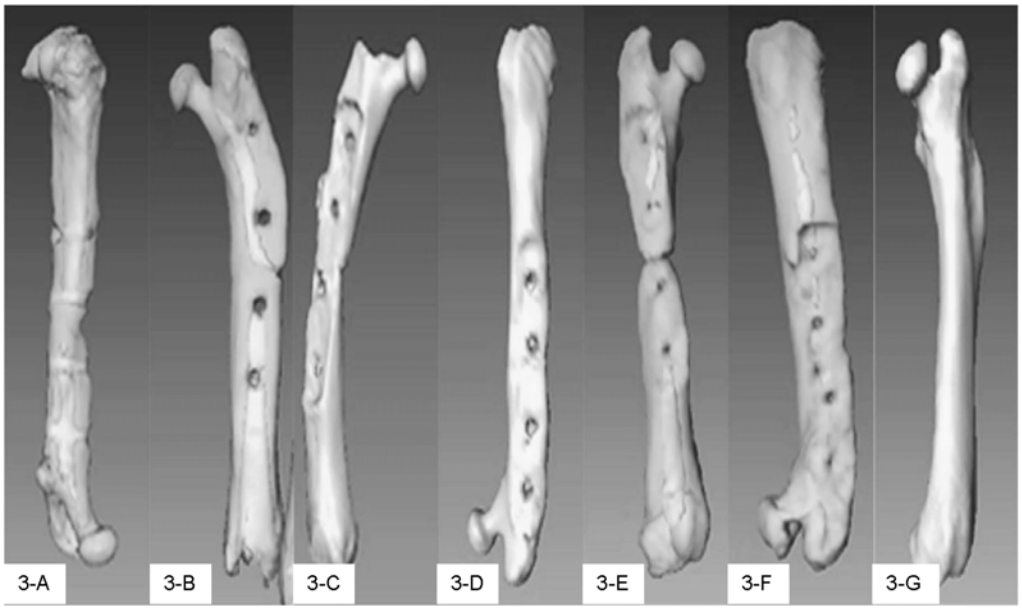


Figure 3. Microcomputed Tomography 3-D Images. (Figure 3-A is an operated control specimen after 3 months. An incomplete intermedullary canal as well as the presence of a surface defect can be visualized. Figure 3-B is an experimental sample after 3 months. Increase in the intermedullary canal and a decrease in the surface defect (versus the operated control sample from the same time point) can be seen. Figure 3-C is the operated control sample after 6 months. An apparent surface defect is present compared to the experimental sample (from the same time point) in figure 3-D. Figure 3-E and 3-F are the operated control and experimental samples, respectively, after 9 months. A complete defect, with boney ingrowth, can be seen in the operated control sample (Figure 3-E). Lack of significant surface defect can be visualized in the experimental sample (Figure 3-F) along with complete defect repair. Figure 3-G is an unoperated control.)

nine months (**Figures 3-D** & **3-F** respectively) the defect-free bone is seen in the experimental group. However, the repair was incomplete in operated controls 3-A, 3-C and 3-E. This confirmed bone growth was stimulated in the experimental group.

4. DISCUSSION

4.1. Undecalcified Histological Sectioning and Staining

During each time interval (three, six and nine months) repair occurred faster in experimental groups compared to than in the operated control groups. The repaired area in the experimental group lacked fibrous tissue and, histologically, appeared to consist primarily of cortical bone. In comparison, operated control samples demonstrated repair with a relative high concentration of voids and fibrous tissues. Defects were present, in the operated control femurs, after even nine months of repair time.

The color variation in the sections was due to specimen thickness, whereby more tissue was stained. Using the Exakt system, exact specimen thickness is difficult to control. General colors were similar with slight shade variation. Exact duplication in slide thickness was attempted but not always achieved. Thus histological analysis demonstrated the presence of bone.

At the three month time interval, bone infiltrated into the defect site although infiltration was not complete, the defect could still be visualized. In operated control samples, incomplete healing began at the defect site. Bone infiltration was sporadic, filled with large voids, and the defect was still present (**Figures 1-A** & **1-B**). In experimental samples, the defect was completely filled with mature bone and the intermedullary canal was beginning to reform (**Figure 1-C**). In the operated control histological sections, repaired bone could be visualized, although the rate of repair was far inferior to repair seen in the experimental group.

Six month undecalcified histological operated control sections displayed significant fibrous growth at the defect site. Tissue juxtaposed to the defect site was highly infiltrated with fibrous tissue (**Figure 1-D**). Analysis of later sections revealed the area of fibrous tissue decreased in the defect site. Six month experimental histological sections displayed superior growth compared to the operated control samples. Repaired bone tissue displayed continuous growth although large voids, white spaces, were noticed (**Figure 1-E**). Six month experimental histological sections showed improved repair compared to operated control samples. While voids were present in experimental samples, operated control samples had complete defects present.

Nine month undecalcified operated control sections displayed a high relative percentage of fibrous tissue in the defect site. Unlike the six month group, fibrous tissue was juxtaposed to cortical bone (**Figure 1-F**). In the nine month undecalcified experimental sections, defects were only slightly present with no signs of fibrous tissue or voids (**Figure 1-I**). All defects, in experimental femurs, were repaired with, histologically, cortical bone. After surgery, rats gained weight due to lack of significant movement and therefore had less space for movement. This lack of significant movement reduced the mechanostimulation on all loaded skeletal structures, including the defected bone. Since repair processes strive under loaded conditions and these lethargic animals presented reduced loaded conditions, bone growth decreased from the three month group to the nine month group. During all time intervals, animals were confined to Institutional Animal Care and Use Committee (IACUC) approved cages 10.5" W x 19" L x 8" H (Allentown Caging Equipment, Inc., Allentown, N.J.), physical activity was negligible and feed intake was constant (free choice). Reduced physical activity led to decreased mechanostimulation which led to weakened repaired bone [21,22]. Nevertheless the nine month experimental group demonstrated a relative higher percentage of bone compared to operated control femurs from the same time period.

4.2. Non-Destructive Testing by Microcomputed Tomography (µCT)

MicroCT (µCT) is a series of x-ray images compiled together into a three dimensional image. Individual slices were available for viewing as well as three dimensional images. Microcomputed tomography images of the bone graft site displayed progressive repair in experimental samples compared to operated control femurs, which displayed a complete defect throughout all time periods. Operated control femurs, across all time intervals, displayed similar characteristics. All operated control samples had incomplete intermedullary canals, defect presence (appropriate for a critical sized defect) and dense tissue growth over bone ends at defect site. Likewise, experimental samples displayed similar characteristics across all time intervals. All experimental samples had a slight surface defect present (although size decreased as time progressed) and intermedullary canal was continuous throughout entire femoral length. Thus µCT provided a non-destructive technique to view internal composition without compromising sample integrity. Non- destructive microcomputed tomography testing allowed femurs to undergo additional testing, thus maximizing the quantity of data per sample.

In the operated control group, repair was present along the ventral femoral surface. Repair was incomplete (discontinuous intermedullary canal with defect on dorsal femoral surface) and resultant area of repair visually appeared as dense cortical bone in a single slice image. In **Figure 3-A**, the surface defect can be visualized along with an incomplete intermedullary canal. In the experimental group, significant bone repair was observed. In **Figure 3-B**, the defect was almost completely repaired with the exception of a small surface disruption and the intermedullary canal has begun to reform.

In the six month operated control group, the defect was still visible and bone was somewhat deformed (**Figure 3-C**). In the experimental six month group, repair was significantly improved compared to the operated control group during the same time interval. The intermedullary canal was continuous throughout the femur. A slight remnant of the defect in cortical bone on ventral surface remained. **Figure 3-D** shows the dorsal plane of the defect completely healed and indistinguishable from surrounding bone.

In the nine month operated control group, cortical bone defined edges of the defect (**Figure 3-E**) and repair of the intermedullary canal was discontinuous. In the nine month experimental group, repair was essentially complete with only small amounts of defect present; the three dimensional rendering showed the absence of a visually apparent defect (**Figure 3-F**). **Figure 3-G** showed the unoperated control femur.

5. CONCLUSIONS

The specific conclusions were:

1. The animal evaluations of the composite paste for 3, 6 & 9 months examined by undecalcified histology and microCT (μCT) demonstrated a high degree of bone ingrowth for the experimental group, compared to samples in the operated control group (defect repaired without paste).

2. The use of μCT as a non-destructive tool to follow bone ingrowth was found to be very valuable. Although we used this technique mainly for qualitative data, the data could be digitized to make more quantitative predictions.

3. Purified crawfish chitosan compounded with calcium sulfate in paste form can be successfully used to deliver osteoblasts that will allow bone growth to occur in defects.

ACKNOWLEDGMENTS

The authors appreciate the help of Errin Robinson and Alan L. Ogden for surgical assistance, Dollie Smith for growing the osteoblasts and Shelia Rogers for guidance regarding the sectioning and staining of slides. The plates were supplied by Synthes (Paoli, PA).

REFERENCES

[1] Q. Li, E. T. Dunn, E. W. Grandmasion, M. F. A. Goosen. (1992) Applications and Properties of Chitosan. Applications of Chitin and Chitosan, ed. M. F. A. Goosen. Technomic Publishing Company, Inc.: Lancaster. 3−29.

[2] H. No, S. Meyers, (1995) Preparation and characterization of chitin and chitosan-A review. Journal of Aquatic Food Product Technology. 4(2): 27−52.

[3] H. K. No, S. P. Meyers, (1997) Preparation of chitin and chitosan, in Chitin Handbook, R. A. A. Muzzarelli, Peter, M. G., Editor. 475−489.

[4] S. B. Rao, C. P. Sharma, (1997) Use of chitosan as a biomaterial: Studies on its safety and haemostatic potential. Journal of Biomaterials Research. 34: 21−28.

[5] T. Jiang, C. Pilane, C. (2005) Laurencin, Fabrication of Novel Porous Chitosan Matrices as Scaffolds for Bone Tissue Engineering. Materials Research Society Symposium. 845: 187−192.

[6] Y. Chen, Y. Chung, L. Wang, K. Chen, S. Li, (2002) Antibacterial Properties of Chitosan in Waterborne Pathogens. Journal of Environmental Science and Health. A37 (7): 1379−1390.

[7] Z. Shi, K. Neoh, E. Kang, W. Wang, (2006) Antibacterial and mechanical properties of bone cement impregnated with chitosan nanoparticles. Biomaterials. 27: 2440−2449.

[8] T. Don, C. Chuang, W. Chiu, (2002) Studies on the Degradation Behavior of Chitosan-g-Poly (acrylic acid) Copolymers. Tamkang Journal of Science and Engineering. 5(4): 235−240.

[9] P. VandeVord, H. M., S. DeSilva, L. Mayton, B. Wu, P. Wooley, (2002) Evaluation of the biocompatibility of a chitosan scaffold in mice. Journal of Biomedical Materials Research. 59: 585−590.

[10] Y. Okamoto, R. Yano, K. Miyatake, I. Tomohiro, Y. Shigemasa, S. Minami, (2003) Effects of chitin and chitosan on blood coagulation. Carbohydrate Polymers. 53: 337−342.

[11] M. Jumaa, F. H. Furkert, B. W. Müller, (2002) A new lipid emulsion formulation with high antimicrobial efficacy using chitosan. European Journal of Pharmaceutics and Biopharmaceutics. 53: 115−123.

[12] P. A. Felse, T. Panda, (1999) Studies on applications of chitin and its derivatives. Bioprocess and Biosystems Engineering, 20 (6): 505−512.

[13] T. Chandy, C. P. Sharma, (1990) Chitosan as a Biomaterial. Biomaterials, Artificial Cells and Artificial Organs. 18(1): 1−24.

[14] P. K. Dutta, J. Dutta, M. C. Chattopadhyaya, V. S. Tripathi, (2004) Chitin and Chitosan: Novel Biomaterials Waiting for Future Development. Journal of Polymer Materials. 21(3): 321−333.

[15] S. Mallapragada, B. Narasimhan, (2006) Handbook of Biodegradable Polymeric Materials and their Applications. Vol. 1 & 2.

[16] A. K. Singla, M. Chawla, (2001) Chitosan: some pharmaceutical and biological aspects-an update. Journal of Pharmacy and Pharmacology. 53(8): 1047−1067.

[17] D. Mukherjee, (2001) Method for producing chitin or chitosan, US patent 6310, 188B1, Oct. 30.

[18] T. A. Einhorn, J. M. Lane, A. H. Burstein, C. B. Kopman and V. J. Vigorita, (1984) The healing of segmental defects induced by demineralized bone matrix, J. of Bone and Joint Surgery, 66−A: 274−279.

[19] D. W. Wolff, V. M. Goldberg and S. Stevenson (1994) Histomorphometric analysis of the repair of a segmental diaphyseal defect with ceramic and titanium fiber metal implants: effect of bone marrow. J. of Orthopaedic Research. 12(3): 439−446.

[20] D. P. Mukherjee, A. S. Tunkle, R. A. Roberts, A. Clavenna, S. Rogers, D. Smith, (2003) An Animal Evaluation of a Paste of Chitosan Glutamate and Hydroxyapatite as a Synthetic Bone Graft Material. Journal of Biomedical Material Research, Part B. 67B: 603−609.

[21] T. L. N. Jarvinen, P. Kannus, H. Sievanen, P. Jolma, A. Heinonen, M. Jarvinen, (1998) Randomized controlled study of effects of sudden impact loading on rat femur. Journal of Bone and Mineral Research. 13(9): 1475−1482.

[22] Y. Umemura, T. Ishiko, T. Yamauchi, M. Kurono, S. Mashiko, (1997) Five jumps per day increase bone mass and breaking force in rats. Journal of Bone and Mineral Research. 12(9): 1480−1485.

J. Biomedical Science and Engineering, 2009, 2, 30-35

Investigating connectional characteristics of motor cortex network*

Dong-Mei Hao[1], Ming-Ai Li[2]

[1]School of Life Science and Bioengineering. [2]School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China. Correspondence should be addressed to Dong-Mei Hao (haodongmei@bjut.edu.cn)

Received September 10th, 2008; revised November 12th, 2008; accepted December 11th, 2008

ABSTRACT

To understand the connectivity of cerebral cortex, especially the spatial and temporal pattern of movement, functional magnetic resonance imaging (fMRI) during subjects performing finger key presses was used to extract functional networks and then investigated their characteristics. Motor cortex networks were constructed with activation areas obtained with statistical analysis as vertexes and correlation coefficients of fMRI time series as linking strength. The equivalent non-motor cortex networks were constructed with certain distance rules. The graphic and dynamical measures of motor cortex networks and non-motor cortex networks were calculated, which shows the motor cortex networks are more compact, having higher statistical independence and integration than the non-motor cortex networks. It indicates the motor cortex networks are more appropriate for information diffusion.

Keywords: Motor Cortex Network, Connectivity, Correlation Coefficient, Functional Magnetic Resonance Imaging (fMRI), Activation Area

1. INTRODUCTION

Neuroanatomical studies have revealed a large number of connections linking different brain structures. There is a wealth of information about the patterning and functional impact of connection pathways linking segregated areas of the cerebral cortex. The brain consists of networks of highly interconnected regions, coordinating major aspects of behaviour. The brain is inherently a dynamic system, in which the traffic between regions, during behavior or even at rest, creates and reshapes continuously complex functional networks of correlated dynamics [1]. The cerebral motor cortex which is closely associated with movement is innervated by a number of anatomical and functional connections.

*This work was supported by the National Natural Science Foundation of China (30670543).

Over recent years, neurophysiological and neuroimaging experiments as well as detailed computer simulations of neuronal networks have contributed to our understanding of the neural mechanisms generating functional connectivity. Functional magnetic resonance imaging (fMRI) is a non-invasive and widely available technique for mapping brain functions. It is based upon the blood oxygenation level-dependent (BOLD) effect. As concerns the motor system, the available functional imaging studies indicate a mass activation effect within the hand representation area during finger-tapping or finger-to-thumb opposition tasks in terms of either a stepwise or a linear function between movement rate and hemodynamic response [2].

Some computational approaches such as covariance structural equation modelling (SEM) aim at inferring causal relations between brain areas from their pattern of covariance, by extracting networks of effective connectivity [3]. Dynamic causal modelling (DCM) characterizes the dynamics of interactions among states of brain regions with bilinear approximations of intrinsic coupling among neuronal states and the influence of external inputs [4]. Granger causality mapping (GCM) extends the vector autoregressive (VAR) technique to capture interactions among brain regions, assuming a causal and dynamic system of linear interactions, driven by stochastic innovations [5]. A graphical approach linking the notions of graphical models and Granger causality has been applied to describe dynamic dependencies in neural systems [6, 7]. Several principled approaches such as non-metric multidimensional scaling [8], hierarchical analysis [9] and cluster analysis [10] have been used to derive numerical descriptions of the organization of the network from neuroanatomical connection data. Using basic and general concepts from information theory, entropy and mutual information, O. Sporns *et al.* have developed a theoretical measure that captures the interplay of functional segregation and integration within a given system [11]. A close look at the anatomical and functional organization of the cerebral cortex provides important clues for formulating a potential general mechanism for neural integration.

In this paper we present our current research investigating the connectivity of the motor system, the network underlying the generation of movement. We propose a

method to extract motor networks and non-motor networks, as revealed by fMRI when humans perform finger tasks, and then analyze them in the context of the current understanding of complex networks.

2. MATERIALS AND METHODS

Considering the multiple processes taking place at different brain regions and interacting with one another in executing a specific task, extracting brain connectivity from fMRI data facilitates our understanding of brain function. The states of activated brain regions are fully observed as intensity variations of fMRI time-series.

2.1. Data Acquisition

The fMRI data center [12] afforded the fMRI data submitted by Kathleen Y. Haaland [13]. Fourteen healthy right-handed volunteers between the ages of 20 and 40 participated in this study. Subjects performed finger key presses in response to numeric sequences presented visually on the screen. The index ("1"), middle ("2"), and ring ("3") fingers of the right or left hand were placed on response keys. Two sequence conditions were used. The simple condition required repetition of one of three sequences (i.e., 11111, 22222, or 33333) and the complex condition consisted of heterogeneous sequences (i.e., 12131, 23231, or 32321). Each 3-sec trial began with the appearance of a five-digit number sequence presented vertically on the screen for 2.5 sec, cueing subjects to immediately perform the sequence as quickly and accurately as possible. The task paradigm consisted of ten 12-s epochs alternating between rest and activation.

Functional MRI was obtained on a 1.5-T General Electric Signa scanner. Echo-planar (EP) images were collected using a single-shot, blipped, gradient echo EP pulse sequence: echo time (TE)=40 msec, data acquisition time =40 msec, field of view (FOV)=24 cm, resolution=64×64. Twenty-two contiguous sagittal 6-mm thick slices provided coverage of the entire brain (voxel size: 3.75×3.75×7mm). Prior to functional imaging, high-resolution 3-D spoiled gradient-recalled at steady-state anatomic images were collected: TE =5 msec, repetition time (TR)=24 msec, 40° flip angle, number of excitations =1, slice thickness=1.2 mm, FOV=24 cm, resolution=256×128. Refer to [13] for detail.

2.2. Motor Cortex Network, MCN

fMRI data were analyzed using SPM5 (Welcome Department of Cognitive Neurology, London, UK) and MATLAB 7 (The Mathworks Inc.) for all subjects. The first six images of each time series were discarded to eliminate signal intensity variations arising from progressive saturation. Echo-planar images were realigned to the first functional image of each time series to remove residual head movement. The functional images of each subject were coregistered with the mean functional image from realignment, normalized to MNI (Montreal Neurological Institute) standard space and spatially smoothed using a Gaussian filter of 6 mm FWHM. First level analysis of each individual was conducted; one-tailed Student t-tests were used to identify brain regions most responsive for finger key presses. The maximum intensity projection of the statistical map and coordinates in MNI space for each maximum were obtained.

We define the maximum intensity voxels (activation areas) as vertexes of a motor cortex network (MCN). The activity of voxel x at time t is denoted as V(x, t) after fMRI pre-processing. We calculate the linear correlation coefficient between any pair of voxels, x_1 and x_2 as formula (1), where $\sigma^2(V(x)) = \langle V(x,t)^2 \rangle - \langle V(x,t) \rangle^2$, $\langle \bullet \rangle$ represents temporal averages. The correlation coefficient is used as the connection strength or weight between these two nodes. Therefore MCN is constructed with motor association cortices as vertexes and their correlation coefficients as linking weights.

$$r(x_1, x_2) = \frac{\langle V(x_1,t)V(x_2,t) \rangle - \langle V(x_1,t) \rangle \langle V(x_2,t) \rangle}{\sigma(V(x_1))\sigma(V(x_2))} \quad (1)$$

2.3. Non-motor Cortex Network, Non-MCN

The equivalent non-motor cortex network composes of k vertexes which were generated randomly in the spatial coordinates range of cerebral cortex and were far away from (greater than threshold) any vertex in MCN. Suppose (x_n, y_n, z_n) is a voxel coordinates in non-motor cortex and (x_{mi}, y_{mi}, z_{mi}) is the i voxel coordinates in motor cortex. For all i (i=1…k), it satisfies the following condition :

$$\sqrt{(x_n - x_{mi})^2 + (y_n - y_{mi})^2 + (z_n - z_{mi})^2} > thred \quad (2)$$

Therefore non-MCN was constructed with k voxels as nodes and the correlation coefficients among them according to formula (1) as connection strength.

2.4. Measures of the Brain Functional Network

Structural aspects are captured using concepts and measures provided by graph theory. All structural analyses are performed on the network's connection matrix, which provides a complete description of all connections and pathways between the network's individual units. Functional connectivity which is the temporal correlation between remote neurophysiological events can be reflected by information theory.

- Density [14]

Density is defined as the sum of the ties divided by the number of possible ties (i.e. the ratio of all tie strength that is actually present to the number of possible ties). The density of a network may give us insights into such phenomena as the speed at which information diffuses among the nodes, and the extent to which nodes have high levels of communicating capital and /or communicating constraint.

- Characteristic path length: lpath

lpath is the global mean of the lengths of the shortest path linking any pairs of nodes, and can be used to describe the connectivity of a network.

- System entropy [11]

The cerebral cortex networks are implemented as dynamical system. The neural activities can be described as a Gaussian multidimensional stationary stochastic process. The network units interact with each other and deviate from statistical independence through connections. The entropy H(X) of a system measures its overall degree of independence. Assuming stationarity, the entropy of a system X composed of n units is computed as formula (3) with $|\cdot|$ indicating the matrix determinant. COV is the covariance matrix of the system and can be obtained analytically from the connection matrix

$$H(X) = 0.5\ln((2\pi e)^n |COV(X)|) \quad (3)$$

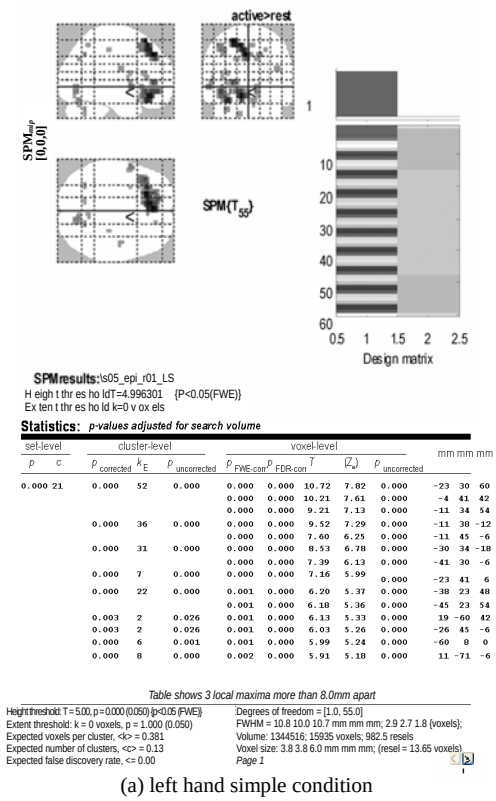
- Integration I(X) [11]

The integration I(X) measures the overall degree to which a system deviates from statistical independence. This measure is derived as the difference between entropies of the individual components of X, considered independently, and the entropy H(X) of the entire system:

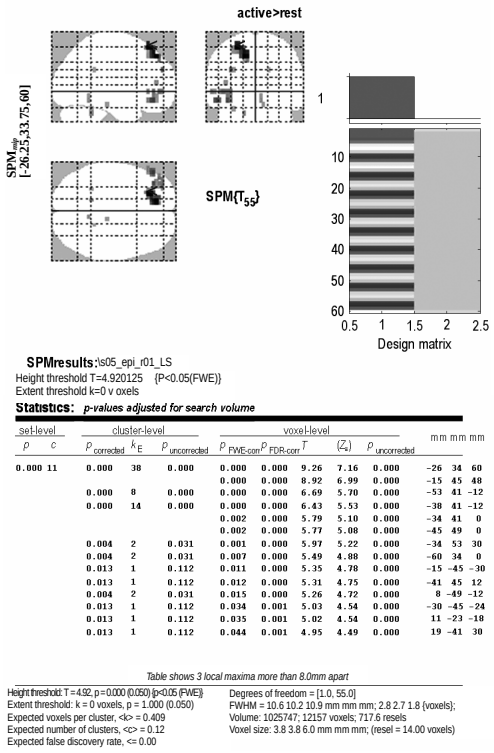
$$I(X) = \sum_{i=1}^{n} H(x_i) - H(X) \quad (4)$$

3. RESULTS

We analyzed the fMRI data of 14 subjects with finger key presses task. 4 subjects had only one activation area in the simple condition, 3 subjects had none or only one activation area in the complex condition. Therefore we only studied the other 10 subjects in the simple condition and 11 subjects in the complex condition, and constructed a MCN and 50 non-MCNs for each of them.Figure 1 is an example of SPM showing bilateral activation of motor cortex of subject 5, in which (a) is the left hand simple condition and (b) is the left hand complex condition. It lists all clusters above the chosen level of significance with details of significance thresholds. Figure 2 shows the motor cortex networks (MCNs) of the same subject in left hand simple and complex conditions. A vertex indicates an activation area, a line indicates a bidirectional connection, and the thicker a line is, the stronger a connection is. Figure 3 shows the non-motor cortex networks (non-MCNs) of the left hand simple and complex conditions for subject 5. A vertex representing a non-activation area is at least 15mm apart from the activation area according to experience [15], and a thicker line also indicates a stronger connection. In Figure 2 (a) vertex 1, 3, 9, 10 locate in BA10 (Brodmann Area), vertex 2 in BA6, vertex 4 in BA10, vertex 5, 8 in BA32, vertex 6, 7 in BA47, vertex 11 in BA7, vertex 12 in BA11, vertex 13 in BA22, vertex 14 in BA18; (b) vertex 1, 2 locate in BA8, vertex 3, 8 in BA47, vertex 4 in BA11, vertex 6 in BA10, vertex 7 in BA9, vertex 10 in BA46, vertex14 in BA31 using Talairach Client 2.4 (Research Imaging Center, University of Texas Health Science Center at San Antonio) [16]. We calculated the average characteristic parameters of 50 non-MCNs and compared them with the corresponding MCNs for each subject in the simple and complex conditions, see Figure 4 and Figure 5, in which "*" indicates a MCN and "o" indicates a non-MCN. We noticed the density (p=0.000),

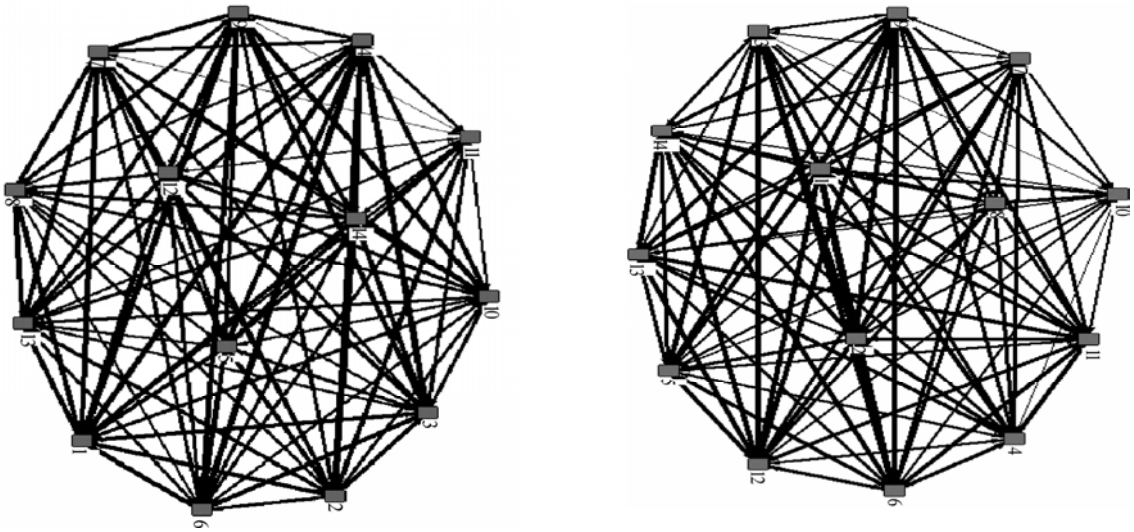


(a) left hand simple condition



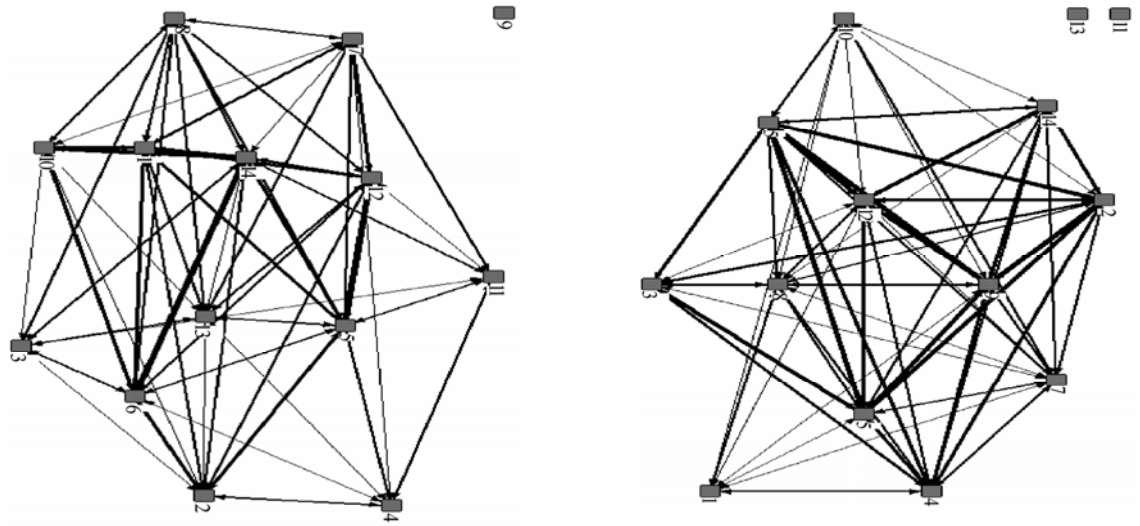
(b) left hand complex condition

Figure 1. SPM showing activation of motor cortex



(a) left hand simple condition (b) left hand complex condition

Figure 2. Cerebral motor cortex networks



(a) left hand simple condition (b) left hand complex condition

Figure 3. Cerebral non-motor cortex networks

characteristic path length (p=0.000), system entropy (p=0.000) and integration (p< MCNs are all significant different from those of non-MCNs regardless of the simple or complex condition. In **Figure 6**, we compared some features of MCNs in the simple and complex condition. "*" indicates a MCN in the simple condition and "o" indicates a MCN in the complex condition. The density (p=0.668), characteristic path length (p=0.728), system entropy (p=0.411) and integration (p=0.243) in the simple condition have no significant difference from those in the complex condition.

3. DISCUSSION

As shown in **Figure 2** and **Figure 3**, MCNs have denser and stronger connections than non-MCNs. In **Figure 4** and **Figure 5**, the density of MCNs are much larger than that of non-MCNs and the characteristic path length of MCNs are much shorter than that of non-MCNs. It indicates the information such as motor strategy, spatial-temporal arrangement and sensorimotor message can diffuse among motor regions with high speed and high level. The entropy and integration of MCNs are larger than those of non-MCNs, which indicate MCNs have higher overall degree of statistical independence than non-MCNs and at the same time there are more statistical dependencies among the motor regions. We can deduce that in our experiment, the simple task and the complex task are at the same cognitive level and therefore have similar functional connectivity patterns for healthy subjects. It is different from other studies [17] in which complex movement increases activity in regions and involvement of areas. Therefore how to define complexity in an experiment context is to be considered.

Movement is an essential part of our daily life activities and the movement handicapped cannot lead productive independent lives due to inability to control their activities of daily living. We attempt to understand the

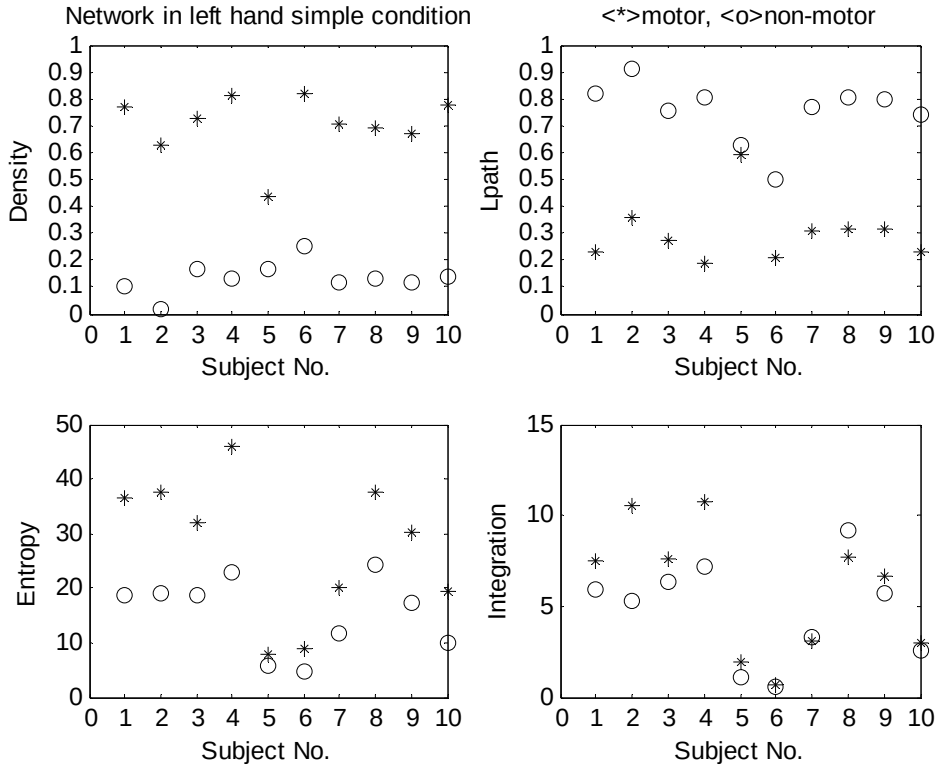


Figure 4. Characteristics comparison of MCNs and non-MCNs in hand simple condition

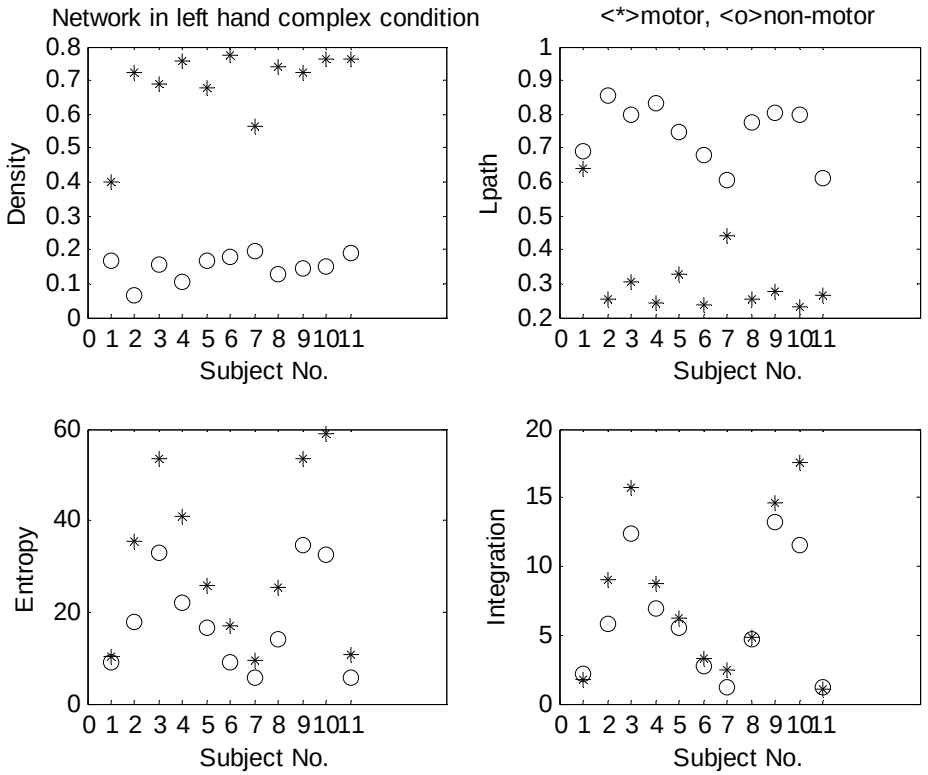


Figure 5. Characteristics comparison of MCNs and non-MCNs in left hand complex condition

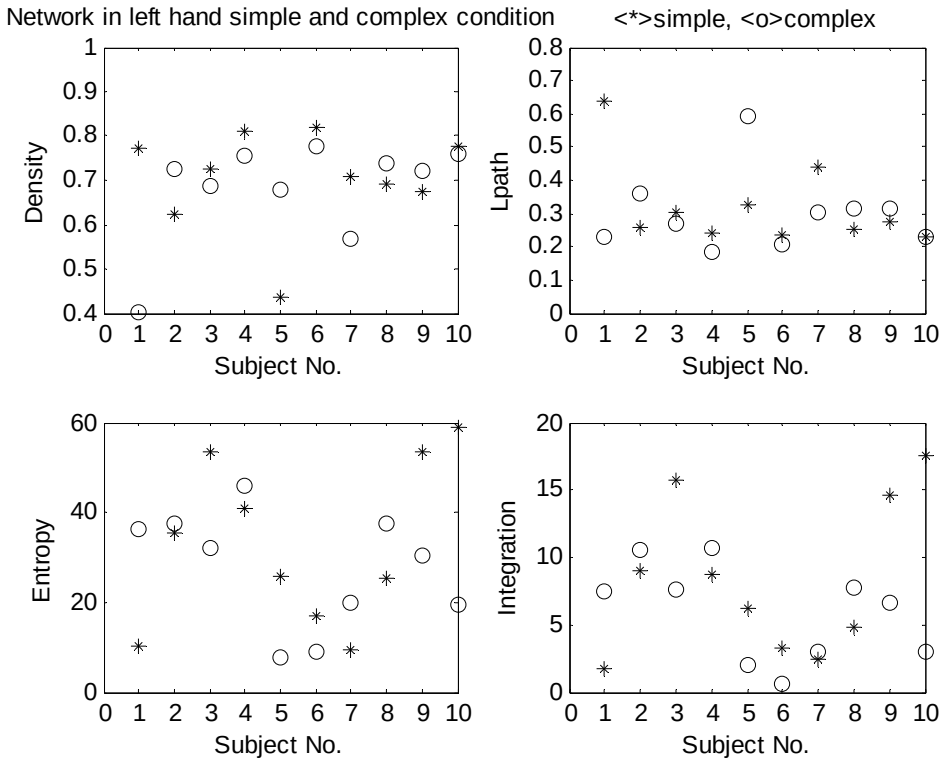


Figure 6. Characteristics comparison of MCNs in left hand simple and complex conditions

features of motor cortex networks with some statistical measures, which may be associated with movement generation and information transfer and help to study the motor skills disruption and rehabilitation.

ACKNOWLEDGEMENTS

The authors' work was supported by grants from the National Natural Science Foundation of China (30670543). The data sets were supported by fMRIDC (The fMRI Data Center, Dartmouth College, http://www. fmridc.org, accession number: 2-2003-114E5).

REFERENCES

[1] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi *et al*. (2005) Scale-Free Brain Functional Networks. Physical Review Letters, 14 January: 018102−1~018102−4.

[2] A. Riecker, D. Wildgruber, K. Mathiak *et al*. (2003) Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fMRI study. NeuroImage, 18: 731−739.

[3] J. C. Zhuang, S. LaConte, S. Peltier, *et al*. (2005) Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination. NeuroImage, 25 (2): 462−470.

[4] K. J. Friston, (2003) Dynamic causal modeling. NeuroImage, 19: 1273−1302.

[5] L. Harrison, W. D. Penny, K. J. Friston, (2003) Multivariateautoregressive modelling of fMRI time series. NeuroImage, 19: 1477–1491.

[6] M. Eichler, (2005) A graphical approach for evaluating effective connectivity in neural systems. Philos. Trans. R. Soc. B, 360: 953−967.

[7] S. Yang, D. Knoke. (2001) Optimal connections: strength and distance in valued graphs. Social Networks, 23: 285−295.

[8] K. E. Stephan, C. C. Hilgetag, G. A. P. C. Burns *et al*. (2000) Computational analysis of functional connectivity between areas of primate cerebral cortex. Phil. Trans. R. Soc. Lond. B, 355, 111−126.

[9] D. J. Felleman & D. C. Van Essen, (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex, 1: 1−47.

[10] S. Dodel, J. M. Herrmann, T. Geisel. (2002) Functional connectivity by cross-correlation clustering. Neurocomputing, 44−46:1065−1070.

[11] S. G. Tononi and G. M. Edelman. (2000) Theoretcal Neuroanatomy: Relating Anatomical and Functional Connectivity in Graphs and Cortical Connection Matrices, Cerebral Cortex, 10: 127−141.

[12] http://www.fmridc.org.

[13] K. Y. Haaland, C. e L. Elsinger *et al*. (2004) Motor Sequence Complexity and Performing Hand Produce Differential Patterns of Hemispheric Lateralization. Journal of Cognitive Neuroscience, 16(4): 621−636.

[14] http://www. analytictech.com.

[15] A. McNamara, M. Tegenthoff, H. Dinse, *et al*. Increased functional connectivity is crucial for learning novel muscle synergies. Neuroimage, 2007, 35: 1211−1218 (p1213, ROI= 6mm, 7voxels).

[16] http://www.talairach.org.

[17] D. J. Serrien, R. B. Ivry, S. P. Swinnen. (2007) The missing link between action and cognition. Progress in Neurobiology, 82: 95−107.

J. Biomedical Science and Engineering, 2009, 2, 36-40

Biodegradable and bioactive porous polyurethanes scaffolds for bone tissue engineering

Mei-Na Huang[1], Yuan-Liang Wang[1*], Yan-Feng Luo[1]

National 985 Research Center of Bioinspired Material Science and Engineering, Bioengineering College,Chongqing University, Chongqing 400030, People's Republic of China, Correspondence to Yuan-Liang Wang (wyl@cqu.edu.cn). Tel. /fax: +86-23-65102509.

Received September 22nd, 2008; revised November 12th, 2008; accepted November 19th, 2008

ABSTRACT

Biodegradable porous polyurethanes scaffold have themselves opportunities in service, including controlled degradation rate, no-toxic degradation products. However, polyurethanes are lack of bioactive groups, which limits their application. This review gives the common modification methods, surface functionalization and blending modification. In finally, the review puts forward to the bulk modification as a new method to enhance the bioactivity of polyurethanes.

Keywords: Polyurethanes, Bioactivity, Biodegradation, Bone Repair

1. INTRODUCTION

Currently, tissue engineering involving synthetic materials offers a practical approach for bone repair and regeneration. In this approach, a 3-D porous biodegradable scaffold is beneficial to guide cell attachment, proliferation and tissue regeneration [1,2]. Therefore, a number of researchers are interested in developing biodegradable polymeric scaffolds for bone engineering repair [3,4,5,6]. Polyurethane, which concludes the polyurethane urea elastomer, is regarded as a kind of bone repair materials for its nice mechanical property and their special shape memory function.

Biodegradable polyurethanes, made from degradable polyester/polyether with hydrophilic group of ether bond, aliphatic diisocyanate, having the hydrophobic group of alkly and chain extenders [7,8]. Due to these special group, polyurethanes have controlled degradation rate, in general, the degradation time can reach to some months with changing of the ratio polyester/polyether to diisocyanate [7,9], which fits to the growth rate of osteoblast. Moreover, the degradation give rise to non-toxic products, which will not produce side effect for body. Besides polyester/polyether and diisocyanate, chain extender is also a key factor. In order to regulate the pH of degradation products, and avoid the acid auto-catalytic effect in the degradation process, and then further controlling the easily control of degradation rate, some researches choose diamines [10]. *Guan et al* [4] synthesized (poly (etherurethane urea), PEUU) with PCL and 1, 4-diisocyanatobutane (BDI) and putrescine. And then, PEUU was made into highly porous, biodegradable polyurethane scaffold for tissue engineering. In this study, BDI was used, since it could release putrescine, a polyamine that is essential for cell growth and proliferation. *Zhang et al* [11] synthesized polyurethane by reacting of highly pure lysine diisocyanate with glucose, which resulted in major degradation products lysine and glucose (LDI-glucose), and then completely degradate and enter into human circulation system.

The degradation mechanisms of polymers are important and need to be investigated further. Non-toxic degradation products are necessary and, moreover, mechanical properties are also influenced by degradation mechanisms. LDI-glucose [11] polymer, for example, is degraded by hydrolysis of urethane bonds to liberate lysine, glucose, ethanol, and CO_2. Ethanol could inhibit cell-cell adhesion, but a study reported that concentrations less than 30mM are harmless to the cell. Moreover, in contrast to PLA and PLGA degradation mechanisms, the study showed that the degradation of polyurethane with diamine no significant increase in pH of the solution. PEUU degradation products were also shown to be non-toxic to endothelial cells. The polymer showed a linear degradation with no signs of autocatalytic effects when compared to PLA or PLGA degradation behaviour. In addition, regulating ratio of polyester/polyether to diisocyanate can change the molecular weight of polyurethane, and then control their degradation rate. The two regulation methods make it be balance with growth of cell/tissue and realize the real tissue engineering repair.

However, polyurethanes as a potential, biodegradable materials are lack of bioactive groups, which limits their applications. Therefore, how to ensure biodegradation and bioactive of polyurethane are two key factors for it's application in bone repair [12,13]. A further requirement for scaffold, particularly used for bone engineering, is controllable interconnected porosity for cells to grow into the desired physical form and to compete vascularization of the ingrown tissue [12]. Other highly desirable

features concerning the scaffold processing are near-net -sHAe fabrication and scalability for cost-effective industrial production [12,14].

In the paper, we only discuss how to enhance the bioactivity of porous polyurethane scaffold. In general, bioactive functionalization methods of polyurethanes can be concluded to three major design strategies [15,16,17, 18,19]. One approach is blending the polyurethanes with tricalcium phosphate/ hydroxyapatite or other inorganic ceramic [16,17,18,19]. Various bioactive factors further enhance the cellular compatibility. The inorganic ceramic have another advantage, the function of bone induction and the conduction [16,17,18,19]. The other approach involves endowing the biomaterials with bioactivity by incorporating soluble bioactive molecules, such as growth factors and plasmid DNA, into biomaterial carriers so that the bioactive molecules can be released from the materials and trigger or modulate new tissue formation [20,21,22]. The last one is incorporation of cell-binding peptides into biomaterials via chemical or physical modification. The cell-binding peptides include a native long chain of extracellular matrix (ECM) proteins as well as short peptide sequences derived from intact ECM proteins that can incur specific interactions with cell receptors [15,23,24,25] This paper reviews above methods and focuses on their opportunities as a kind of bone repair materials, and puts forward a new method to improve the bioactivity of biodegradable polyurethanes.

2. BIOACILITY OF POLYURETHANE

Tissue engineering applies methods from materials engineering and life sciences to artificial construction new tissue. Two common approaches are transplanting the biomaterials with cell [26] or the biomaterials with some bioactive factor/bioactive substance for the cell homing to realize restoration. Facing the complex biological and sensitive human body, requirements of biomaterials are extremely challenging. The First and most, compared to other bioactive materials, polyurethanes are lack of bioactive factors and cytocompatibility [27], which can be well solved by introduction of bioactive substances, including the inorganic phosphate, growth factors and extracellular matrix.

2.1. Introduction of Inorganic Phosphate Into Polyurethanes

Hydroxyapatite, glasses, glass-ceramics or calcium phosphates having similar components with natural bone [14,20], are important categories of bioactive materials. Coating and blending are the most common methods to modify polymer with inorganic phosphate. Biomimetic method is a chemical modification with inorganic phosphate [28,29,30].

Hydroxyapatite (HA), the most important inorganic phosphate, has been extensively investigated over the past few decades as a biomedical material. It can be designed as a bioactive material, besides it is similar composition with natural bone, osteoconducive, osteoinductivity, biodegra-dability, high mechanical strength and their medical products such as screws, plates and rods have been commercial forms a strong bond to natural bone in vivo [31,32,33]. Moreover, the introduction of HA can regulate the pH of biomaterials. Above properties of hydroxyapatite and other inorganic phosphate can induct the growth of bone and prevent the inflammatory reaction [31,32,33,34].

Rezwan, K. et al [13] reviewed the function of bioactive glasses, glass-ceramics and the calcium phosphates or HA in the enhancement of the bioactivity of polyurethanes. It has been found that reactions on bioactive glass surfaces can release critical concentrations of soluble Si, Ca, P and Na ions, depending on the processing route and particle size. The released ions induce intracellular and extracellular responses. One key reason that makes bioactive glassed-correlation material is the possibility of controlling a range of chemical properties and thus the rate of bioresorption. *Park, Y.S. et al* [35] investigated the fabrication method of a three-dimensional reticulated scaffold with interconnected pores of several hundred micrometers using calcium phosphate glass in the system of CaO-CaF_2-P_2O_5-MgO-ZnO and a polyurethane sponge as a template. It is thought that this kind of biodegradable glass scaffold combined with osteogenic cells has potential to be studied further as a tissue engineered bone substitute. The structure and chemistry of glasses, in particular sol-gel derived glasses, can be tailored at a molecular level by varying either composition, or thermal or environmental processing history.

Above inorganic phosphate is important bioactive modification material, however, current technology is difficult to solve the compatibility between inorganic phosphate and polyurethanes. It is difficult to make a uniform matrix, particularly, the current coating/blending methods，which result in that it is difficult to form a uniform matrix, particularly, the content of inorganic ceramic is high [36,37]. Some researches found that some of HA/PLA composites lost their strengths rapidly in physiological environment and failures occur mainly at the interface of HA and the polymer matrix. Two main reasons may take responsibility for these interfacial failures: one is lack of effective adhesion between ceramic phase and polymer matrix; the other is self-catalytic degradation of hydroxyl groups on HA surfaces to polymer main chains. The structure of polyurethanes/HA is similar to HA/PLA, which may result in the same interface separation. For solving the problem, Xian, Y.M adopted chemical reaction to produce HA crystal on the polymer surface, the chemical reaction to make inorganic phosphate in the surface polymer can solve the interface separation, however, another problem appeared [30]. The reaction of making HA/polymers crystal is similar to the biomimetic calcification, which lasted for more than one week, and then make negative effect on the polymers. Moreover, the products can not ensure the crystal structure.

Introducing inorganic phosphate can not wholly solve the bioactivity problem. Some researchers use bioactive factor, which can react with polymers to enhance their bioactivity [38,39,40], such as RGD, moreover, the bioactive factors is important for cell homing.

2.2. Surface Modification of Porous Polyurethanes Scaffold with Bioactive Factors

2.2.1. Arg-Gly-Asp(RGD) Modified Biomimetic Polyurethanes

In an effort to improve the adhesion and retention of cells to polymer scaffolds, researches typically coated with various extracellular matrix proteins [40,41,42]. These studies highlight that extracellular proteins played an important role in attachment and spreading of cells to surface, where specific domains on cell membrane bind directly with extracellular matrix moleculaes *via* integrins [43,44]. A number of specific cell-recognition sequences have been identified, the most extensively studied sequence being the arginine-glycine-aspartic acid (RGD) motif present in matrix molecules such as vitronectin, fibronectin, laminin and collagen, fibrillin [40,45,46,47].

RGD peptide is one of the major bioactive factor to design biomimetic polyurethanes and has been widely researched in recent years [38,39,40]. In order to provide a stable linking, RGD peptides should be covalently attached to polymer via functional groups like hydroxyl-, amino-, or carboxyl-groups. Some polyurethanes are amino-terminated [4,38], which can react with the carboxyl-groups of RGD, with 1,3-Dicyclohexylcarbodiimide (DCC) as catalyst. Other polyurethanes are hydroxyl-terminated, the hydroxyl-also can react with the carboxyl-group of RGD [48].

Moreover, in order to enhance the surface functionalization, polymeric materials, such as polyurethane must be functionalized before bioactive peptides or proteins are immobilized on their surfaces [44]. In general, the functionalization can be realized by a variety of means, either by introduced the multi-functional groups monomer or polymer[39], or by subsequent surface modification by plasma treatment [45] ozone oxidation [46] surface graft polymerization [40] or site-specific reactions [47]. Here, we put emphasis on two examples to demonstrate the successful application of linking group in surface modification. One example [39], the difunctional spacer molecule-diisocyanate is introduced as the linking group of polyurethane film and RGD, realizing the surface functionalization of polyurethane. Another example, *Jozwiak, A.B* [40] used two steps to enhance the introduction rate. First, the multi-amino group-polyethyleneimine (PEI) is introduced, a medium sized molecular weight branched form of PEI was used here in order to provide a large number of reactive primary amine groups and enhance its entrapment within the polyurethane surface. Second, introducing the dextran, which is functional spacer molecule and can link the RGD easily.

2.2.2. Growth Factors Modified Biomimetic Polyurethanes

Chemotaxis, proliferation, differentiation and matrix synthesis are essential in natural tissue/organ development and wound healing [45]. Owing to the rapid advances in recombinant technology and the availability of large scale manufacturing of cytokines and growth factors, many recent tissue engineering strategies have turned to specific growth factors to stimulate cellular activity in vitro and to improve functional neotissue formation in vivo [47,48]. Characteristic of these bioactive factors is that they can effective release at specific site and realize the function of improving cell proliferation and recruitment [46,49]. Incorporation of angiogenic growth factors such as basic fibroblast growth factor (bFGF) and vascular endothelial growth factor (VEGF), among others, into scaffolds for controlled release has been shown to promote lacal angiogenesis [50]. Platelet-derived growth factor (PDGF) has been demonstrated to stimulate proliferation and recruitment of both periodontal ligament and bone cells in vitro. In vivo study also showed that PDGF-BB enhances the ability of healing [46].

There are many methods to incorporate growth factors into synthetic scaffolds, such as absorbing growth factor to scaffold, and blending growth factor containing microspheres into the scaffold [46], or directly mixing growth factor containing protein powder into the scaffold during processing [50]. However, absorbing growth factors onto the scaffold has the drawback of low loading efficiency and rapid releasing, which may be associated with in bioactivity due to harsh solvents such as hexane [46] or methylene chloride [51]. Incorporating growth factor directly into the scaffold can potentially avoid these shortcomings.

Whether or not has bioactivity of the released bioactive factors is an essential problem. Bioactivity of the factors can be assessed in two methods [47,50,52]. First, bioactivity of the released factor can be determined through the direct method-human gingival fibroblase DNA synthesis as measured by specific composition [46]. Second, the bioactivity is assessed in terms of its ability to stimulate the growth of cells [50,52].

3. CONCLUTION AND PRESPECT

3.1. Possibility and Challenge of Bulk Modification for Polyurethanes

Besides above methods, how can we improve bioactivity of polymers? Now, a great wealth of knowledge about the biology of integrin mediated cell adhesion has proved that the modification of polyurethanes with RGD peptides or other bioactive factors are useful tool to design bioactive porous scaffolds that can provide biological cues elicit specific cellular responses and direct new tissue formation. However, the surface modification has some limitations. Since surface modification has been performed on well-defined model surfaces and the

evaluation of cell behavior on material has been conducted under serum free media, the results may not properly indicate complicated events associated with in vivo environments. Even though some model surfaces may be useful to provide fundamental knowledge to understand cell behavior through specific binding, they may not be directly used as tissue engineering scaffolds.

If we use bulk designing of polyurethanes, incorporated RGD or collagen may result in recognition sites is present not only on the surfaces but also in the bulk of the materials. *Niu, X.F. et al* [53] review the bulk modification, which describe the bulk modification of biomaterials is beneficial to tissue engineering applications where injectable biomimetic materialsare required to match the complex HA of native tissue at defect sites. *Cook et al* [54] and *Barrera et a*l [55] conducted a lot of investigations in understanding the effects of bulk modification via RGD peptides. They synthesized RGD bulk modified poly (lactic acid-co-lysine) and successfully blended it with PLA to fabricate a thin film. When this film was exposed to endothelial cell suspended media for 4 h, the specific function of RGD was maintained to facilitate cell spreading.

Polyurethanes are the biomaterials with hydroxyl-terminated and amine-terminated. The RGD or other peptide can react with the terminal group of polyurethanes, which may result in bioactive polyurethanes. My laboratory chose the bulk modification to introduce the bioactive factor, such as RGD/MGF, and then emulsion/freeze drying mean was adopted to make porous polyurethane scaffold with bioactivity polyurethane.

3.2. The Possibility of Introduction of Inorganic Phosphate by Chemical Reaction

Inorganic phosphate is important component of natural bone, however, current technology is difficult to solve the compatibility between the inorganic phosphate and polyurethanes.

How to introduce the inorganic phosphate, and at the same time avoid above disadvantage is a key problem for enhancing the stability of polyurethane/inorganic phosphate composition. In order to overcome these limitations of composite, covalently attached the inorganic phosphate to polyurethanes by linking group may be a feasible method. Linking group should easily react with the hydroxyl-from inorganic phosphate and the carboxyl- or amino-group from polyurethanes. Silane derivatives are used as modification molecular to link hydroxyl groups (-OH) in HA surface to polymer main chain, which is carried out via direct reactions of –OR groups on HA surfaces. At the same time, other functional groups (-NH_2) of silane derivatives may further react towards the terminal groups carboxylic group or hydroxyl group. Moreover, glutaraldehyde [43] may be the important cross-linking agent. In addition, in order to ensure the homogeneity of composite, the effective connection of emulsion blending-chemical crosslinking may be an efficient method [44].

For realizing biodegradation, bioactivity and mechanical property of the bone repair materials, the paper puts forward two methods to make the biodegradable materials, which are equipped with the uniform structure and bioactive components.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from the "Eleven-Five" National Science and Technology Support Program of China (2006BA103B04).

REFERENCES

[1] J. R. Hench, (2003) Regeneration of trabecular bone using porous ceramics. Current Opinion in Solid State and Materials Science, 7, 301−307.

[2] Z. H. Zhou, J. M. Ruan, (2008) Preparation and bioactivity of sol-gel macroporous bioactive glass. Journal of University of Science and Technology, 15, 290−298.

[3] J. J. Guan, W. R. Wagner, (2005) Synthesis, characterization and cytocompatibility of polyurethane urea elastomers with designed elastase sensitivity. Biomacromolecules, 6, 2833−2842.

[4] J. J. Guan, K. L. Fujimoto, (2005) Preparation and characterization of highly porous, biodegradable polyurethane scaffolds for soft tissue applications. Biomaterials, 26, 3961−3971.

[5] Y. Wang, G. B. Ameer, (2002) A tough biodegradable elastomer. Nature Biotechnology, 20, 602−606.

[6] K. E. Healy, A. Rezania, (1999) Designing biomaterials to direct biological responses. Annals of the New York Academy of Sciences, 875, 24−25.

[7] S. A. Guelcher, K. Gallagher, (2005) Synthesis of biocompatible segmented polyurethanes from aliphatic diisocyanates and diurea diol chain extenders. Acta Biomaterialia, 1, 471−484.

[8] K. D. Kavlock, T. W. Pechar, (2007) Synthesis and characterization of segmented poly(esterurethane urea) elastomers for bone tissue engineering. Acta Biomaterialia, 3, 475−484.

[9] S. A. Guelcher, (2008) Synthesis, mechanical properties, biocompatibility, and biodegradation of polyurethane networks from lysine polyisocyanates. Biomaterials, 29(12), 1762−1775.

[10] Y. F. Luo, Y. L. Wang, (2008) Evaluation of the cytocompatibility of butanediamine and RGDS-grafted poly (d, l-lactic acid). European Polymer Journal, 44, 1390−1402.

[11] J. Y. Zhang, E. J. Beckman, (2002) Synthesis, biodegradability, and biocompatibility of lysine diisocyanate-glucose polymers. Tissue Engineering, 8(5), 771−785.

[12] X. Miao, Y. Hu, (2004) Porous calcium phosphate ceramics prepared by coating polyurethane foams with calcium phosphate cements. Materials Letters, 58, 397−402.

[13] K. Rezwan, Q. Z. Chen, (2006) Biodegradable and bioactive porous polymer /inorganic composite scaffolds for bone tissue engineering. Biomaterials, 27(18), 3413−3431.

[14] M. Gelinsky, P. B. Welzel, (2008) Porous three-dimensional scaffolds made of mineralised collagen: Preparation and properties of a biomimetic nanocomposite material for tissue engineering of bone. Chemical Engineering Journal, 137, 84−96.

[15] H. Shin, S. Jo, (2003) Biomimetic materials for tissue engineering. Biomaterials, 24, 4353−4364.

[16] M. Bil, J. Ryszkowska, (2007) Bioactivity of polyurethane-based scaffolds coated with Bioglass. Biomedical Materials, 2(2), 93−101.

[17] A. Chetty, T. Steynberg, (2008) Hydroxyapatite-coated polyurethane for auricular cartilage replacement: an in vitro study. Journal of Biomedical Materials Research A, 84 (2), 475−482.

[18] X. Huang, X. Miao, (2007) Novel Porous Hydroxyapatite Prepared by Combining H_2O_2 Foaming with PU Sponge and Modified with PLGA and Bioactive Glass. Journal Biomaterials Applications, 21(4), 351−374.

[19] C. Vitale-Brovarone, E. Verne, (2007) Development of glass-ceramic scaffolds for bone tissue engineering: Characterisation, proliferation of human osteoblasts and nodule formation. Acta Biomaterialia, 3, 199–208.
[20] M. J. Whitaker, R. A. Quirk, (2001) Growth factor release from tissue engineering scaffolds. Journal of Pharmacy and Pharmacology, 53, 1427–1437.
[21] T. P. Richardson, W. L. Murphy, (2001) Polymeric delivery of proteins and plasmid DNA for tissue engineering and gene therapy. Gene Expression, 11, 47–58.
[22] J. E. Babensee, L. V. McIntire, (2000) Growth factor delivery for tissue engineering. Pharmaceutical Research, 17, 497–504.
[23] H. Shin, S. Jo, (2002) Modulation of marrow stromal osteoblast adhesion on biomimetic oligo[poly-thylene glycol) fumarate] hydrogels modified with Arg-Gly-Asp peptides and a poly (ethyleneglycol) spacer. Journal of Biomedical Materials Research, 61, 169–179.
[24] Y. Suzuki, M. Tanihara, (2000) Alginate hydrogel linked with synthetic oligopeptide derived from BMP-2 allows ectopic osteoinduction in vivo. Journal of Biomedical Materials Research, 50, 405–409.
[25] F. Buket Basmanav, G. T. Kose, (2008) Sequential growth factor delivery from complexed microspheres for bone tissue engineering. Biomaterials, 29, 4195–4204.
[26] C. M. Hill, Y. H. An, (2007) Osteogenesis of Osteoblast Seeded Polyurethane-Hydroxyapatite Scaffolds in Nude Mice. Macromolecular Symposium, 253, 94–97.
[27] G. Ryan, A. Pandit, (2006) Fabrication methods of porous metals for use in orthopaedic applications. Biomaterials, 27, 2651–2670.
[28] T. Sakura, C. Tanaka, M. Yang, (2004) Production and characterization of a silk-like hybrid protein, based on the polyalanine region of Samia cynthia ricini silk fibroin and a cell adhesive region derived from fibronectin. Biomaterials, 25(4), 617–624.
[29] A. Rainer, S. Maria, (2008) Fabrication of bioactive glass-ceramic foams mimicking human bone portions for regenerative medicine" Acta Biomaterialia, 4, 362–369.
[30] Y. M. Xiao, D. X. Li, (2007) Preparation of nano-HA/PLA composite by modified-PLA for controlling the growth of HA crystals. Materials Letters, 61, 59–62.
[31] L. L. Hench, (1997) Sol-gel materials for bioceramic applications. Current Opinion in Solid State and Materials Science, 2, 604–610.
[32] Z. K. Hong, P. B. Zhang, (2005) Nano-composite of poly (L-lactide) and surface grafted ydroxyapatite:Mechanical properties and biocompatibility. Biomaterials, 26, 6296–6304.
[33] Qiu, X. Y., Hong, Z. K. (2005) Hydroxyapatite Surface Modified by L-Lactic Acid and Its Subsequent Grafting Polymerization of L-Lactide. Biomacromolecules, 6, 1193–1199.
[34] P. L. Lin, H. W. Fang, (2007) Effects of hydroxyapatite dosage on mechanical and biological behaviors of polylactic acid composite materials. Materials Letters, 61, 3009–3013.
[35] H. J. Moon, K. N. Kim, (2006) Effect of calcium phosphate glass on bone formation in calvarial defects of Sprague-Dawley rats. Journal of Materials Science Materials in Medicine, 17 (9), 807–813.
[36] J. Russias, E. Saiz, (2006) Fabrication and mechanical properties of PLA/HA composites: A study of in vitro degradation. Materials Science and Engineering C, 26, 1289–1295.
[37] S. M. Zhang, J. Liu, (2005) Interfacial fabrication and property of hydroxyapatite/polylactide resorbable bone fixation composites. Current Applied Physics, 5, 516–518.
[38] H. J. Salacinski, G. Hamilton, (2003) Surface functionalization and grafting of heparin and/or RGD by an aqueous-based process to a poly(carbonate-urea)urethane cardiovascular graft for cellular engineering applications. Journal of Biomedical Materials Research Part A, 3(66A), 688–697.
[39] J. J. Guan, S. Michael, (2004) Biodegradable poly (ether ester urethane) urea elastomers based on poly (ether ester) triblock copolymers and putrescine: synthesis, characterization and cytocompatibility. Biomaterials, 25, 85–96.
[40] A. B. Jozwiak, C. M. Kielty, (2008) Surface functionalization of polyurethane for the immobilization of bioactive moieties on tissue scaffolds. Journal of Materials Chemistry, 18, 2240–2248.
[41] C. M. Li, C. Vepari, (2006) Electrospun silk-BMP-2 scaffolds for bone tissue engineering. Biomaterials, 27, 3115–3124.
[42] M. Nsksmura, M. Mie, (2008) Construction of multi-functional extracellular matrix proteins that promote tube formation of endothelial cells. Biomaterials, 29, 2977–2986.
[43] Y. M. Yue, K. Xu, (2008) Preparation and Characterization of Interpenetration Polymer Network Films Based on Poly (vinyl alcohol) and Poly (acrylic acid) for Drug Delivery. Journal of Applied Polymer Science, 6(108), 3836–3842.
[44] X. L. Xu, X. S. Chen, (2007) Electrospun poly (L-lactide)-grafted hydroxyapatite/poly (L-lactide) nanocomposite fibers. European Polymer Journal, 43, 3187–3196.
[45] G. B. Wei, Q. M. Jin, (2006) Nano-fibrous scaffold for controlled delivery of recombinant human PDGF-BB. Journal Controlled Release, 112, 103–110.
[46] J. Y. Lee, S. H. Nam, (2002) Enhanced bone formation by controlled growth factor delivery from chitosan-based biomaterials. Journal Controlled Release, 78 (1-3), 187–197.
[47] J. A. Jansen, J. W. Vehof, (2005) Growth factor-loaded scaffolds for bone engineering. Journal Controlled Release, 101 (1–3), 127–136.
[48] S. A. Guelcher, A. Srinivasan, (2008) Synthesis, mechanical properties, biocompatibility, and biodegradation of polyurethane networks from lysine polyisocyanates. Biomaterials, 25, 1762– 1775.
[49] J. J. Guan, J. J. Stankus, (2007) Biodegradable elastomeric scaffolds with basic fibroblast growth factor release. Journal Controlled Release, 120, 70–78.
[50] J. Ziegler, U. Mayr-Wohlfart, (2002) Adsorption and release properties of growth factors from biodegradable implants. Journal of Biomedial Materials Research, 59, 422–428.
[51] T. W. King, C. W. Patrick, (2000) Development and in vitro characterization of vascular endothelial growth factor (VEGF)-loaded poly (DL-lactic-co-glycolic acid)/poly(ethylene glycol) microspheres using a solid encapsulation/single emulsion/ solvent extraction technique. Journal of Biomedial Materials Research, 51, 383–390.
[52] F. Bono, P. Rigon, (1997) Heparin inhibits the binding of basic fibroblast growth factor to cultured human aortic smooth-muscle cells. Biochemical Journal, 326, 661–668.
[53] X. F. Niu, Y. L. Wang, (2005) Arg-Gly-Asp (RGD) modified biomimetic polymeric materials. Journal of Materials Science and Technology, 21(4), 571–576.
[54] A. D. Cook, J. S. Hrkach, (1997) Characterization and development of RGD-peptide-modified poly (lactic acid-co-lysine) as an interactive, resorbable biomaterial. Journal of Biomedial Materials Research, 4, 513–523.
[55] D. A. Barrera, E. Zylstra, (1993) Synthesis and RGD Peptide Modification of a New Biodegradable Copolymer: Poly (lactic acid-celysine). Journal of the American Chemical Society, 115, 11010–11011.

J. Biomedical Science and Engineering, 2009, 2, 41-50

EEA algorithm model in estimating spread and evaluating countermeasures on high performance computing

Si-Yuan Liu[1,3], Chao Liu[2], Yu Liu[4], Yi-Ming Luo[3], Gao-Jin Wen[1], Jian-Ping Fan[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. [3]Hong Kong University of Science and Technology, Hong Kong, China. [4]Royal Institute of Technology, Stockholm, Sweden. Correspondence should be addressed to Si-Yuan Liu (liusiyuan@ict.ac.cn).

Received March 12th, 2008; revised December 7th, 2008; accepted December 29th, 2008

ABSTRACT

This work started out with the in-depth feasibility study and limitation analysis on the current disease spread estimating and countermeasures evaluating models, then we identify that the population variability is a crucial impact which has been always ignored or less emphasized. Taking HIV/AIDS as the application and validation background, we propose a novel algorithm model system, EEA model system, a new way to estimate the spread situation, evaluate different countermeasures and analyze the development of ARV-resistant disease strains. The model is a series of solvable ordinary differential equation (ODE) models to estimate the spread of HIV/AIDS infections, which not only require only one year's data to deduce the situation in any year, but also apply the piecewise constant method to employ multi-year information at the same time. We simulate the effects of therapy and vaccine, then evaluate the difference between them, and offer the smallest proportion of the vaccination in the population to defeat HIV/AIDS, especially the advantage of using the vaccination while the deficiency of using therapy separately. Then we analyze the development of ARV-resistant disease strains by the piecewise constant method. Last but not least, high performance computing (HPC) platform is applied to simulate the situation with variable large scale areas divided by grids, and especially the acceleration rate will come to around 4 to 5.5.

Keywords: EEA, ODE, HIV/AIDS, Spread Estimating, Countermeasure Evaluation, High Performance Computing

1. INTRODUCTION

There is an ancient Arabic saying, *those who predict the future, lie, even if they think they are telling the truth.* This saying succinctly sums up the great uncertainty in projecting the future, especially for a complex problem such as HIV/AIDS spread estimation issues. As the HIV/AIDS pandemic enters its 27th year, both the number of infections and number of deaths due to the HIV/AIDS continue to leap. Even though an enormous amount of effort, our global society remains uncertain on how to most effectively estimate the spread of this disease, evaluate different countermeasures to it and allocate resources to fight this epidemic.

Nevertheless, attempts to predict future trends and prevalence of HIV/AIDS have been carried out with a wide range of errors, using the following methods for estimating HIV/AIDS prevalence, such as back-calculation method, "ratio" method, multiplying the estimated annual HIV/AIDS cases by 20, using the results of serological surveys and extrapolating these data to the total 15–49 years old population and some recently developed methods, the workbook method, and the special computer models [1,2].

Based on the differential equation theory，we propose a novel algorithm model system, EEA model system, a new way to estimate the spread of HIV/AIDS, evaluate different countermeasures to HIV/AIDS and analyze the development of ARV-resistant disease strains. It is a series of solvable ordinary differential equation (ODE) models to estimate the spread of HIV/AIDS infections, which not only require only one year's data to deduce the situation in any year, but also apply the piecewise constant method to employ multi-year information at the same time, overcoming the limitation of the classic infection model (SI model) which ignores the change of the population, and the scarcity and error of data.

We simulate the effects of therapy and vaccine, then evaluate the difference between them, and offer the smallest proportion of the vaccination in the population to defeat HIV/AIDS, especially the advantage of using the vaccination while the deficiency of using therapy separately. At last, we analyze the development of ARV-resistant disease strains by the piecewise constant method.

According to our models, we can firstly outline the spread period of HIV/AIDS without any control in a country can be reasonably divided into three main periods (**Figure 1.**) [3,4]: free spread period, burst spread period and stable spread period.

In recent years, the demand on modeling capability has increased rapidly in the areas of disease analysis, drug design study, environmental assessment, etc. Most modeling approaches are still based on the traditional single-CPU reservoir simulators and have reached their limits with regard to what can be accomplished with them. During the same period, high performance computing (HPC) technology has increasingly been recognized as an attractive alternative modeling approach to resolving large-scale or multi-million-cell simulation problems [5]. As a result, parallel computing techniques have received more attention in this modeling community.

2. AN IDEAL MODEL TO ESTIMATE THE SPREAD OF HIV/AIDS

Build a solvable model to estimate the proportion of the number of HIV/AIDS infections in the population for any country in any year, in the absence of any additional interventions. **Model 1**

2.1. Assumptions and Definitions

Ⅰ. All the derivatives referred in the equations exist.

Ⅱ. t is time (unit: year).

N (t) is the population of the country at t.

N (t) obeys Logistic Population Model

Ⅲ. The crowd is divided into the susceptible crowd and the infective crowd.

i (t) is the proportion of the infected population in the total population at t.

The rest of the N (t) is the susceptible crowd.

Ⅳ. Two main transmission ways:

Cross infections: **T-1**

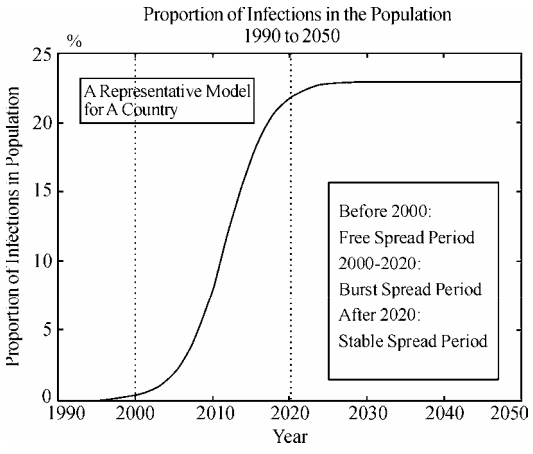


Figure 1. Proportion of infections in the Population 1990 to 2050

Sharing drug injecting equipment and transfusion of blood or blood-derived products. Vaginal intercourse without a condom (man to woman and woman to man) and anal sex without a condom (both partners are at risk). We assume as the average number of persons infected by an HIV patient per year. Single-chain infections: **T-2**

An infected mother to her baby during pregnancy, at childbirth, or by breast feeding. And is the birth rate of infected infants in infected crowd per year.

Ⅴ. K_3 is the rate of patient death.

2.2. Design of the Model

Through the first transmission way, every patient can infect $K_1(1-i(t))$ healthy people, so there are $K_1(1-i(t))i(t)N(t)$ people infected every year.

Through the second transmission way, the number of infected people is $K_2N(t)i(t)$.

We can get

$$\frac{d}{dt}\left(N(t)i(t)\right) = K_1N(t)(1-i(t))i(t) + K_2N(t) - K_3N(t)i(t) \tag{1}$$

After deduce, we can get

$$i'(t) = (K_1 + K_2 - K_3)\mathrm{i}(t) - \frac{N'(t)}{\mathrm{N}(t)}i(t) - K_1i(t)^2,$$

So that we can get

$$i'(t) = (K_1 + K_2 - K_3 - \frac{d}{dt}\ln N(t))i(t) - K_1i(t)^2$$

The following question is how to solve those equations. Based on the classic Logistic population model, N (t) should satisfy

$$\frac{dN}{dt} = -aN^2 + bN,$$

Its solution is

$N(t) = \frac{b}{a + ce^{-bt}}$, where α and β are constants. c is determined by initial condition.

After fitting, we find out, for the countries in the chart, α is smaller than β by several orders of magnitude averagely. So we let $\alpha = 0$ in computation.

The population data of China in this diagram come from[13].Then,

$$N(t) \approx c^{-1}\beta e^{\beta t} \tag{2}$$

So $\frac{d}{dt}\ln \mathrm{N}(t) = \beta$ is a constant value for the same country.

Let a= K_1,

$$b = K_1 + K_2 - K_3 - \frac{d}{d(t)}\ln N(t) \tag{3}$$

and add the initial condition I (t_0) = c, meaning c is the proportion of the infected people at t_0, then we can get

$$\begin{cases} i'(t) = -ai(t)^2 + bi(t) \\ i(t_0) = c \end{cases} \quad (4)$$

Finally we get

$$i(t) = \frac{b}{a - (a - \frac{b}{c})e^{-b(t-t_0)}} \quad (5)$$

where a, b are undetermined coefficients.

When it comes to how to get a and b, we believe the best way is to get the statistical value of K_1, K_2, K_3 and $\frac{d}{dt}\ln(N(t))$, then get a and b by (3). Of course this will burden the workload of the public health department, but that worth it greatly, because these two constants will tell people the destination and rapidity of the HIV/AIDS infections.

It's easy that

$$\lim_{t\to\infty} i(t) = \frac{b}{a} \quad (6)$$

That means, without any additional interventions, the proportion of infected population will steadily come to $\frac{b}{a}$. Meanwhile b indicates the speed of i (t) which tends to limit. So the larger the value of b of a country is, the faster the rate of change in the number of HIV/AIDS infections for this country, and the more attention should be paid to this country. So based on abundant data, it's significant to estimate the value of a and b. Then we can realize the potential destroy of HIV/AIDS to a country.

It is a pity that we can not get the believable conclusion about K_1, K_2, K_3, so we have to apply the fitting method as a makeshift.

The detailed method is as the followings.

Because the solution of (4) exists and only exists, we can get a series of equations about a and b by referring to the data between 1995 to 1997, and get a series of approximate solutions, while the rest of the data are applied to test the model.

The reason why we take the data in early times is that in those days the infection situation we take as without too many additional interventions which is very necessary and reasonable for our model.

Also we can forecast the trend of HIV/AIDS infections in a country, especially the coming of the peak and the appropriate time to take appropriate measures to control the situation.

Table 1. Data of china

Year	1997	1999	2001	2003
Number of HIV/AIDS infections	300,000 [8]	500,000 [12]	660,000 [10]	840,000 [10]

This model ignores some additional interventions, but is easy to be modified to satisfy the requirements in the following models, focusing on two interventions: provision of antiretroviral (ARV) drug therapies, and provision of a hypothetical HIV/AIDS preventative vaccine.

The data in **Figure 2** refer to **Table I**.

The situation of the spread of HIV/AIDS in China accords with our three main periods (as **Figure 3** shows). The spread of HIV/AIDS in China is in the free spread period.

3. MODELS TO EVALUATE COUNTERMEASURES TO HIV/AIDS

Based on the estimating model, we take the following scenarios into consideration to build three models to solve different problems and come to the significant results.

1. Antiretroviral (ARV) drug therapy. **Model 2.1**
2. A preventative HIV/AIDS vaccine. **Model 2.2**
3. Both ARV provision and a preventative HIV/AIDS vaccine. **Model 2.3**

Assume in these scenarios that there is no risk of emergence of drug-resistant strains of HIV which we will examine this issue later.

3.1. Model in Scenario 1: Antiretroviral (ARV) Drug Therapy

3.1.1. Assumptions and Definitions

First, based on Ⅰ, Ⅱ, Ⅳ and Ⅴ in **Model 1**.

And,

Ⅲ.

i(t) is the percentage of the infection in the population (N (t)) at t, which is infective.

r(t) is the percentage of patients accepting ARV in the population at t ($r(t) \le i(t)$).

These people take the lower death rate p and are stopped to take the infection actions, while all the non-infected are susceptible.

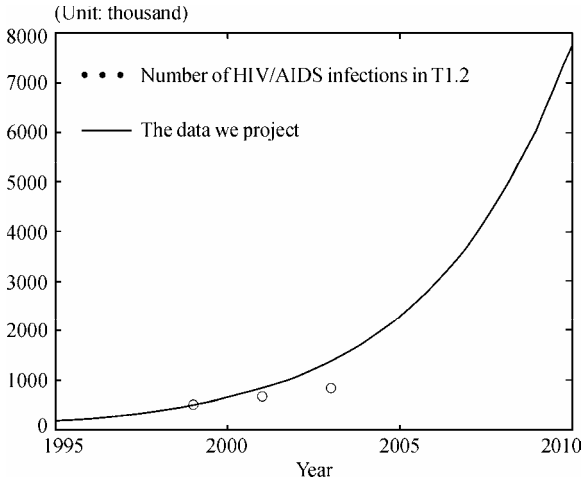


Figure 2. Number of infections in China 1995 to 2010

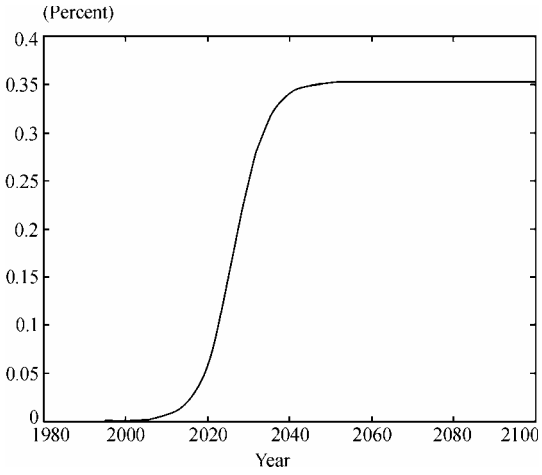


Figure 3. Proportion of infections in population China, 2000 to 2100

Ⅵ.

$\forall$ n, $\exists$ r_n s.t. $\max_{n-1<t\le n} |r(t)-r_n|$ is very small, so, $r(t)\approx r_n$, at $n-1<t\leqslant n$

During solve the equation, we consider r (t) is a constant in the same year, that is, when $n\leqslant t<n+1$, $r(t)=r_n$, where n is the number of years.

We can assert the expression of i(t), which is not a approximate solution calculated by computer simulation. Meanwhile, from the subsequent data, we can find out there is no obvious error. Actually, our calculated result is very accurate.

And because we can take $\frac{d}{d(t)}\ln N(t)$ as a constant value (see (1)), the deformation course of the equation will be

$$i\frac{dN}{dt}+N\frac{di}{dt}=(-K_1i^2+(K_1(1-r_n)+K_2-K_3)-c)N,$$

$$\frac{di}{dt}=-ai^2+(b+r_na)i-c,$$

where $i\frac{dN}{dt}+N\frac{di}{dt}=(-K_1i^2+(K_1(1-r_n)+K_2-K_3)-c)N$, and $c=(a-b-p)r_n$.

These are all undetermined coefficients, while r_n is determined by outside conditions.

With an initial condition, $i(n)=i_n,(0\le i_n\le 1)$

$$\begin{cases} i'=-ai^2+(b+r_na)i-c \\ i(n)=i_n \quad (n\le t\le n+1) \end{cases}$$

Let $i_{n+1}=i(n+1)$, we can backwards calculate all the values in n years.

Similarly we can get

$$\begin{cases} i'=-ai^2+(b+r_na)i-c \\ i(n+1)=i_{n+1} \qquad (n\le t\le n+1) \end{cases}$$

Let $i_n=i(n)$, so we can calculate forwards. Then we finish the solving course of **Model 2.1**.

Now we consider how to apply it into the numeric calculation.

First, based on assumptions IV, a and b here are the same with those are in the **Model 1**.

So we can continue to use the precious fitted data in **Model 1**, while r (n) is determined by outside aid and the population in this country which is given before. So based on the above solving course of the two problems, we just have to get the proportion of the infections in the population in any year after the appearance of HIV/AIDS, and put the data into them inductively to forecast or backdate the value of t in this country in any year.

Then the calculation of the rate of the change of the infections this year is as follows, $\frac{d}{dt}(Ni)=aN(i-r)(1-i)+(b+\beta)Ni$, where N can be looked up in the data[13], while $b=\frac{d}{dt}\ln N$ can be fitted by the least-square estimation based on N, which is different for different countries.(**Figure 1**)

3.1.2. Model Analysis and Application

The following results base on the credible data from [10] [12]. From the diagram (**Figure 4**), we can find out the effect

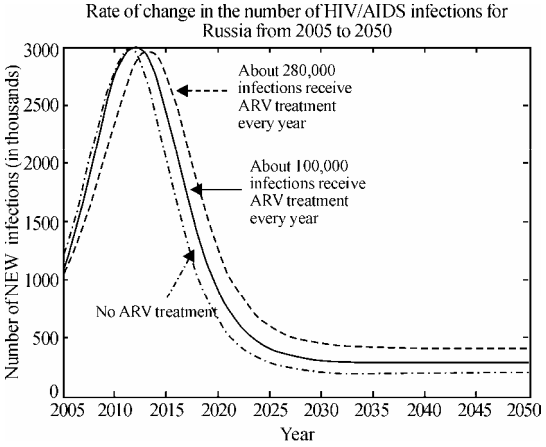


Figure 4. Rate of change in the number of HIV/AIDS infections for Russia from 2005 to 2050

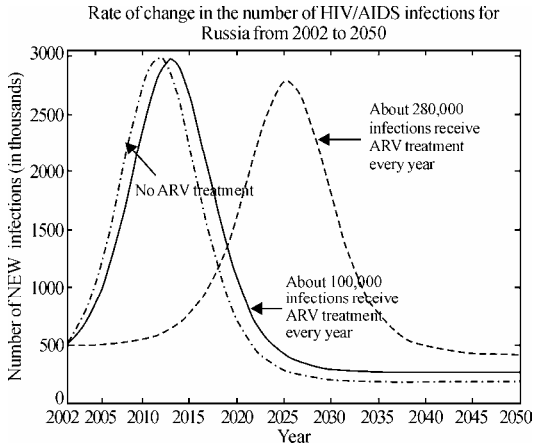


Figure 5. Rate of change in the number of HIV/AIDS infections for Russia from 2002 to 2050

of the Antiretroviral (ARV) drug therapy is tiny, as Russia is in Burst Spread Period, despite of the number of infections who received the treatment is very large.

However, if the therapy can be offered earlier, the effect is delectable. As we can see in the diagram (**Figure 5**) below, about 280,000 infections have received ARV treatment every year since 2002. At that time, Russia is in the end of Free Spread Period. So the ARV drug therapy should be offered as early as possible (better before the Burst Spread Period).

So what is the effect of the ARV drug therapy? Based on our model, we can find out it will delay the coming of the infection peak. As what the following diagram shows, the peak will come late at least 5 years, which means we have 5 more precious years to solve it.

3.2. Model in Scenario 2: Preventative HIV/AIDS Vaccine

3.2.1. Assumptions and Definitions

First, based on Ⅰ, Ⅱ, Ⅳ and Ⅴ. And,

Ⅲ".

We assume vaccine is fully effective without any inefficacy and drug-resistant strains, and it will give the patient whole life immunity.

i(t) is the percentage of the infections in the population (N (t)) at t, which is infective.

s(t) is the percentage of vaccine injections in the population at t, which is not susceptible.

The rest of the N (t) is the susceptible crowd.

Ⅶ.

$\forall$ n, $\exists$ s_n s. t. $\max_{n-1<t\le n} |s(t)-s_n|$ is very small, so, $s(t)\approx s_n$, when n-1$<t\leqslant n$.

During solving the equation, we consider r(t) is a constant in the same year, that is, when $n\le t\le n+1$, $s(t)=s_n$, where n is the number of years.

So we can directly deduce the expression of i(t), which is not a approximate solution calculated by computer simulation. Meanwhile, from the subsequent data, we can find out there is no obvious error. Actually, our calculated result is very accurate.

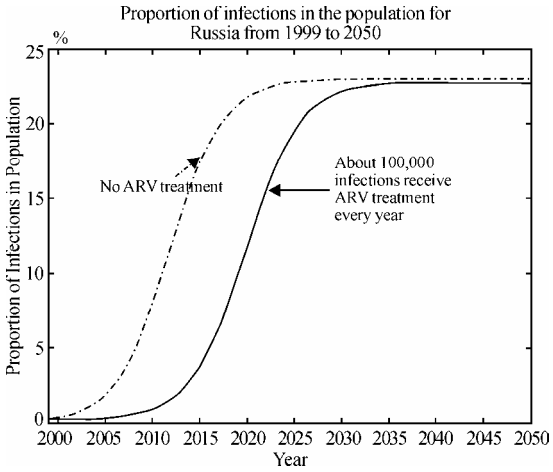


Figure 6. Proportion of infections in the population for Russia from 1999 to 2050

And because we can take $\frac{d}{d(t)}\ln N(t)$ as a constant value (**Model 1**), the deformation course of the equation will be

$$i\frac{dN}{dt}+N\frac{di}{dt}=(-K_1i^2+(K_1(1-s_n)+K_2-K_3)i)N,$$

$$\frac{di}{dt}=-ai^2+(b-s_na)i, \tag{7}$$

where $a=K_{1,}$ $b=K_1+K_2-K_3-\frac{d}{dt}(\ln N)$, and these are all undetermined coefficients, while s_n is determined by outside conditions.

With an initial condition, $i(n)=i_n,(0\le i_n\le 1)$

$$\begin{cases} i'=-ai^2+(b-s_na)i \\ i(n)=i_n \qquad (n\le t\le n+1) \end{cases} \tag{8}$$

So we can get

$$i(t)=(b-as_n)/(a-\frac{a(i_n+s_n)-b}{i_n}e^{-(b-as_n)(t-n)})$$

$$(n\le t<n+1) \tag{9}$$

Similarly we can get

$$\begin{cases} i'=-ai^2+(b-s_na)i \\ i(n+1)=i_{n+1} \qquad (n\le t\le n+1) \end{cases}$$

E.2-2-3'

$$i(t)=(b-as_n)/(a-\frac{a(i_{n+1}+s_n)-b}{i_{n+1}}e^{-(b-as_n)(t-n)})$$

$$(n\le t<n+1) \tag{10}$$

Then we finish the solving course of **Model 2.1**.

From (11), we can get $\lim_{t\to\infty} i(t)=\frac{b}{a}-s$, so when $\frac{b}{a}=s$, i(t)=0. That is, when $s\ge\frac{b}{a}$, maybe we will clear up HIV/AIDS. That means, when the percentage of vaccine injections in the population is lower than b/a, the percentage of the infections in the population will finally tend to b/a-s, so only when it equals to or is higher than b/a, we can finally defeat HIV/AIDS. That really provides a great meaning suggestion to the government layout and deployment.

Now we consider how to apply it into the numeric calculation.

First, it's not hard to find out, a, b here have the same definitions with those are in the **Model 1**. So we can continue to use the precious fitted data in **Model 1**, while s(n) is determined by outside aid and the population in this country which is given before. So based on

the above solving course of the two problems, we just have to get the proportion of the infections in the population in any year after the appearance of HIV/AIDS, put the data into the inductive formula, $i_n = i(n) = \lim_{x \to n-0} i(x)$, so we can forecast or backdate the value of t of this country in any year.

3.2.2. Model Analysis and Application

Take the situation in Brazil as an example to explain our model and result.

The reason why we take Brazil as an example is that the percentage of ARV treatment there has come to 100 %. [12] So we believe it's reasonable and representative to select Brazil to show the effect of the vaccination.

So in Brazil we let the proportion of vaccination is 0.081, which approximates to b/a (refer to **Model 1**). When we use the vaccine measurement in 2010, the rate of change in the number of HIV/AIDS infections will slow down obviously. So if the proportion of vaccination is over 0.08 every year (see **Figure 7**), and the trend of defeating HIV/AIDS maybe will appear.

So we come to the conclusion: Compared with ARV treatment, vaccination is a better way to clear up the HIV/AIDS, which has the following advantages:

Less cost

Convenient application

Better efficiency in short period

Clearing up HIV/AIDS finally

But, it's a pity that the efficient vaccine is not available now.

4. MODEL IN SCENARIO 3: BOTH ARV PROVISION AND PREVENTATIVE HIV/AIDS VACCINE

4.1. Assumptions and Definitions

First, based on Ⅰ, Ⅱ, Ⅳ, Ⅴ, Ⅵ, and Ⅶ. And, Ⅲ'''

i(t) is the percentage of the infection in the population (N (t)) at t, which is infective.

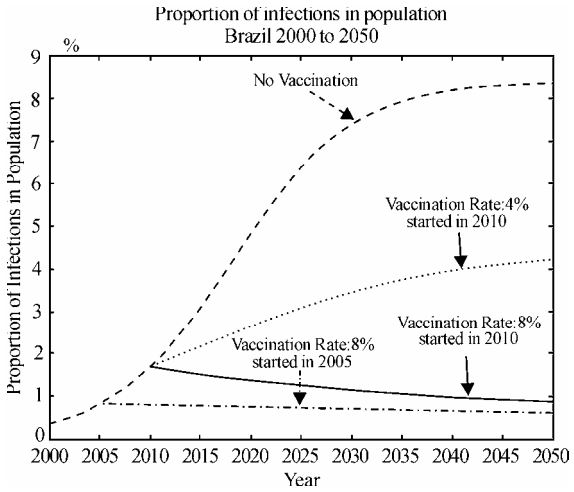


Figure 7. Proportion of infections in population Brazil from 2000 to 2050

s(t) is the percentage of patients accepting ARV in the population at $t(s(t) \le i(t))$. These people take the lower death rate p and are stopped to take the infection actions, while all the non-infected are susceptible.

r(t) is the percentage of vaccine injection in the population at t, which is not susceptible.

4.2. Design of the Model

During solving the equation, we consider r(t) and r(t) are constants in the same year, that is,

when $n \le t < n+1$, $\begin{cases} s(t) = s_n \\ r(t) = r_n \end{cases}$, where n is the number of years.

So we can directly deduce the expression of i(t), which is not a approximate solution calculated by computer simulation. Meanwhile, from the subsequent data, we can find out there is no obvious error. Actually, our calculated result is very accurate.

And because we can take $\frac{d}{d(t)} \ln N(t)$ as a constant value (**Model 1**), the deformation course of the equation will be

$$i\frac{dN}{dt} + N\frac{di}{dt} = (-K_1 i^2 + (K_1(1 - s_n + r_n) + K_2 - K_3)i - c)N,$$

$$\frac{di}{dt} = -ai^2 + (b - s_n a + r_n a)i - c,$$

where $a = K_1$, $b = K_1 + K_2 - K_3 - \frac{d}{dt}(\ln N)$, and $c = a(1 - s_n)r_n$, and these are all undetermined coefficients, while s_n is determined by outside conditions.

With an initial condition, $i(n) = i_n, (0 \le i_n \le 1)$

$$\begin{cases} i' = -ai^2 + (b + (r_n - s_n)a)i - \text{r(-as+b+}\frac{\text{d}}{\text{dt}}\text{lnN-p)} \\ i(n+1) = i_n \qquad\qquad (n \le t \le n+1) \end{cases} \tag{11}$$

$$d = b + (r_n - s_n)a,$$

$$c = r(-as + b - p + \frac{d}{dt}\ln(N)), \quad (n \le t \le n+1).$$

Similarly we can get

$$\begin{cases} i' = -ai^2 + (b + (r_n - s_n)a)i - \text{r(-as+b+}\frac{\text{d}}{\text{dt}}\text{lnN-p)} \\ i(n+1) = i_{n+1} \qquad\qquad (n \le t \le n+1) \end{cases} \tag{12}$$

$$i(t) = \frac{1}{2a}(arctg(-\frac{1}{2}(t - n - 1)\sqrt{4ca - d^2} + \frac{arctg(-d + 2i_{n+1}a)}{\sqrt{4ca - d^2}})\sqrt{4ca - d^2} + d) \tag{13}$$

where

$d = b + (r_n - s_n)a$, $c = r(-as + b - p + \frac{d}{dt}\ln(N))$, $(n \le t \le n+1)$.

Then we finish the solving course of **Model 2.1**

First, it's not hard to find out, a, b here have the same definitions with those are in the **Model 1**. So we can continue to use the precious fitted data in Model 1, while s(n) is determined by outside aid and the population in this country which is given before. So based on the above solving course of the two problems, we just have to get the proportion of the infections in the population in any year after the appearance of HIV/AIDS, and put the data into the above formulas inductively, to forecast or backdate the value of t of this country in any year. Then the calculation of the rate of the change of the infections this year is as follows,

$$\frac{d}{dt}(Ni) = aN(i - r_n)(1 - i - s_n) + (b + \beta)N(i - r_n) + pNr_n,$$

where N can be looked up in the data that ICM provides, while $\beta = \frac{d}{dt}\ln N$ can be fitted by the least-square estimation based on N, which is different for different countries.

4.3. Model Analysis and Application

In this model, we can estimate the expected rate of change in the number of HIV/AIDS infections for a country under realistic assumptions for two scenarios: Antiretroviral (ARV) drug therapy and a preventative HIV/AIDS vaccine.

Take the situation in India as an example to explain our model and result [12].

The reason why we take India as an example is that India, a developing country, will have the largest population in the world in the not far future.

As the above two conclusions we come to, the effi ciency of vaccination is better than treatment. So in the less fund situation, vaccination is the first choice.

To India, a developing country with a very large population, should apply vaccination, under the condition, certain amount of treatment.

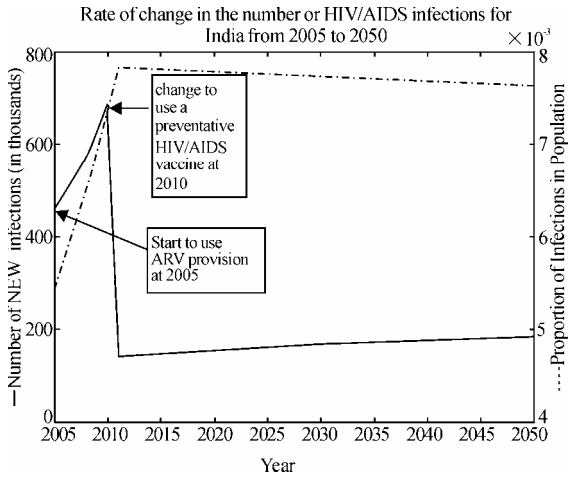


Figure 8. Rate of change in the number of HIV/ADIS infections for India from 2005 to 2050

5. Models to Analyze the Development of ARV-resistant Disease Strains

We will re-formulate the three models developed in last section, taking into consideration the following assumptions about the development of ARV-resistant disease strains.

A person receiving ARV treatment with adherence below 90 percent has a 5 percent chance of producing a strain of HIV/AIDS which is resistant to standard first-line drug treatments. Second-and third-line ARV drug therapies are available, but assume that these drugs are prohibitively expensive to implement in countries outside of Europe, Japan, and the United States. **Model 3**

5.1. Assumptions and Definitions

First, based on Ⅰ, Ⅱ, Ⅳ, Ⅴ, Ⅵ, and Ⅶ.

And,

Ⅲ'''

i(t) is the percentage of the infection in the population (N (t)) at t, which is infective.

r(t) is the percentage of patients accepting ARV in the population at t ($s(t) \le i(t)$).

These people take the lower death rate p and are stopped to take the infection actions, while all the non-infected are susceptible.

s(t) is the percentage of vaccine injection in the population at t, which is not susceptible.

Ⅷ

p in Ⅲ''' is the function of t , and

$\forall$ n, $\exists$ p_n s.t. $\max_{n-1<t\le n} |p(t) - p_n|$ is very small, so, $p(t) \approx p_n$, at $n - 1 < t \le n$.

5.2. Design of the Model

At first, we have to do some necessary modifications.

The development of ARV-resistant disease strains doesn't have effect on the model without ARV treatment, **Model 2.2**. So we just have to focus on **Model 2.1** and **Model 2.3**. Because of the similarity between these two models, we just take **Model 2.1** as an example.

The development of ARV-resistant disease strains will cause the rising of the proportion of the patients with ARV treatment, that is, p is bigger. To get the more general result, we assume p is the function of t, p(t).

Remain the rest of the assumptions, the model will be modified into

$$\begin{cases} \frac{d}{dt}(N(t)i(t)) = K_1 N(t)(1 - i(t))(i(t) - r(t)) \\ +(K_2 - K_3)N(t)(i(t) - r(t)) - p(t)Nr(t) \\ N(n)i(n) = N_n i_n \end{cases}$$

Just as how we deal with r(t) and s(t), we assume $p(t) = p_n (n \le t < n+1)$.

Then deduce functions, we get

$$\begin{cases} \dfrac{di}{dt} = -ai^2 + (b+ar)i - c \\ i(n) = i_n \qquad (n \le t \le n-1) \end{cases}$$

and

$$\begin{cases} \dfrac{di}{dt} = -ai^2 + (b+ar)i - c \\ i(n+1) = i_{n+1} \qquad (n \le t \le n-1) \end{cases}$$

where $a = K_1$, $b = K_1 + K_2 - \dfrac{d}{dt}\ln N$, $c = (a+b-p_n)r_n$, $(n \in Z)$.

The solutions of the above questions is uniquely exists, and have the expressions, which we will not give unnecessary details to.

Then with $i_m = i(m) = \lim_{x \to m} i(x)$, we can get a satisfied curve of the rate of the change of infections starting at one point.

5.3. Model Analysis and Application

The result of the model is as the following diagram shows (**Figure 9**).

We take India as an example to explain the result. (HIV/AIDS infections data from [10])

The development of ARV-resistant disease strains will cause the rising of the proportion of the patients with ARV treatment. And India has a long way to solve the HIV/AIDS problem.

We can satisfy all the requirements in the last section, under the conditions in this section, using our models with the similar methods, so we will not give unnecessary details to the realization of other models in this section.

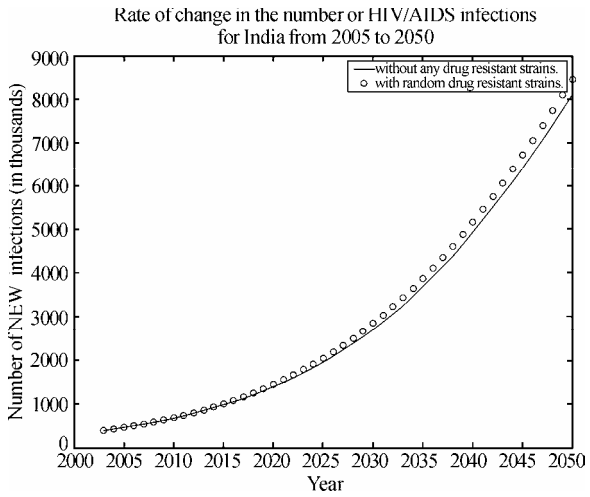


Figure 9. Rate of change in the number of HIV/ADIS infections for India from 2005 to 2050

6. HIGH PERFORMANCE COMPUTING PLATFORM

Dawning 4000L, IDC data processing machine, as **Figure 10**, shows, located in Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, has been designed to provide more than 3 teraflops computing capability with 644 processors, 644GB physical memory and 100 terabytes storage. The machine can be upgraded online to 6.75 teraflops computing capability with 1300 processors, 4000GB physical memory and 600 terabytes storage.

The computing capability of the common PC will be enough for our model, while in simulating a large scale and establish the relation between time and variation ratio in real time, especially when amount of data become very large, the performance of database management system will drop sharply , ability of data of organization and management weaken greatly , can't realize the rapid searching for the a great deal of data .And what is more, it can even cause the breakdown of system when amount of data is getting larger, so high performance computing platform will come to be necessary.

Each CPU will compute the variable situation in a certain area which divided by grids ($P(n): M(n) \to n$, P is the processor, M is the certain area, and n is the number. As **Figure 11** shows, with the communication in the results, finally, we can get the whole situation in a large scale area, such as a country, a continent and the global world, and the acceleration rate will come to around 4 to 5.5.

7. CONCLUSION AND FUTURE WORK

We identify the scenarios and problems in spreading and countermeasures evaluating, and propose a novel algorithm model system, the EEA model system, with three distinctive main conclusions. First, three main periods spread of HIV/AIDS, and finally comes to the stable spread period. Secondly, the limitations and applications of the antiretroviral (ARV) drug therapies. Thirdly the minimum proportion of vaccination in a country to eliminate HIV/AIDS.

Based on the design, analysis, and application of our model system which based on high performance computing platform, we can safely draw the conclusion that the EA model system exploits a new way to estimate the spread of HIV/AIDS, evaluate different countermeasures to HIV/AIDS and analyze the development of ARV-resistant disease strains and will behave a great positive effect on defeating HIV/ AIDS.

The future work we should focus on is the im provement of algorithm for this issue on the HPC platform to arise the accuracy of estimation, the stability of countermeasures evaluation and identification of more significant issues, such as this model maybe will be applicable for the terrorist infection spreading and counter measures evaluation.

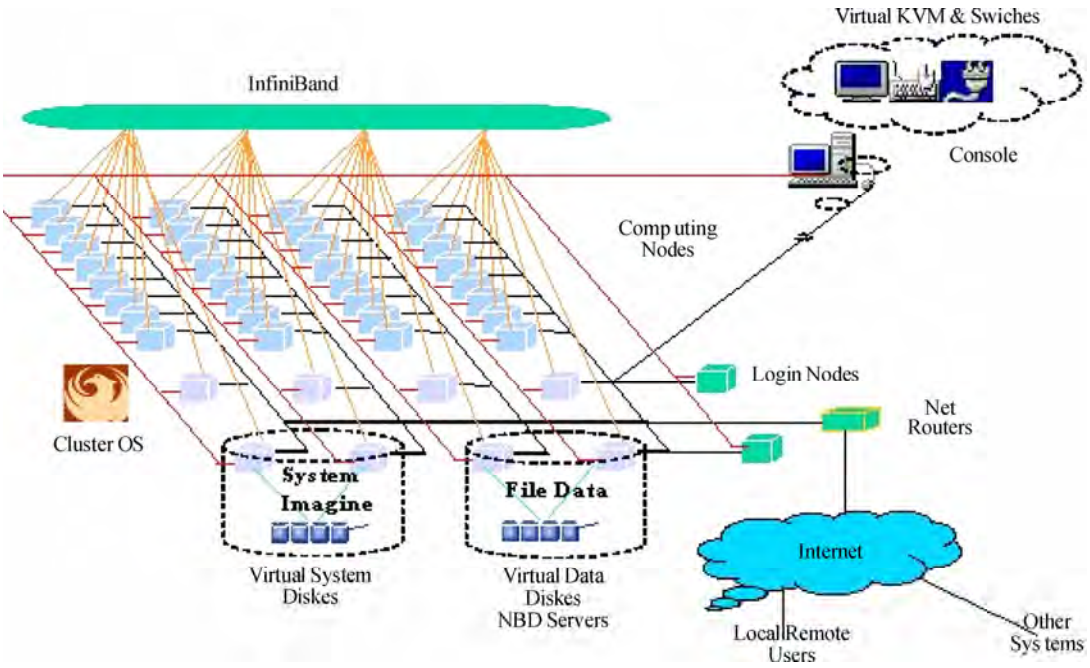


Figure 10. High-performance Computing Platform

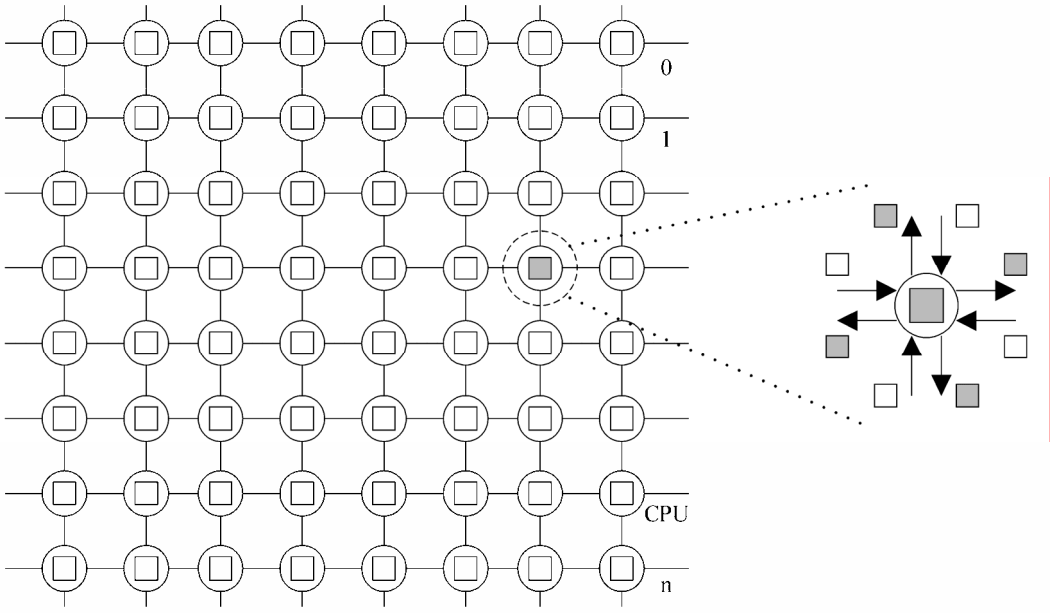


Figure 11. Computation and communication in the CPUs

REFERENCES

[1] B. H. Reinhard; K. S. Bernd; G. Peter; K. Bernhard; S. Shlomo; H. Eilke B 1. (1997) 'Changing incidence of AIDS-defining illnesses in the era of antiretroviral combination therapy', AIDS. 11(14): 1731−1738, November 15, 1997.

[2] B. Rapatski, P. Klepac, S. Dueck, M. X. Liu and L. I. Weiss, (2001) 'Mathematical epidemiology of HIV/AIDS in Cuba during the period 1986−2000', Mathematical Biosciences and Engineering, Vol. 3, No. 3, pp. 545−556.

[3] C. Willard JR. MD, MPH; The American Social Health Association Panel (1999) 'Estimates of the Incidence and Prevalence of Sexually Transmitted Diseases in the United States', Sexually Transmitted Diseases. Vol .26, No.4, Supplement: S2−S7, April.

[4] Lewis. (2001) 'Three stage AIDS incubation period: a worst case scenario using addict-needle interaction assumptions'. Mathematical Biosciences, Vol. 169, No.01, pp. 53−87.

[5] F. Berman, G. Fox, T. Hey. (2003) Grid Computing: Making the Global Infrastructure a Reality. John Wiley and Sons. P747−771.

[6] G. J. Dore, P. K. Correll, Y. Li, J. M. Kaldor, D. A. Cooper, B. J. Brew. (1999) 'Changes to AIDS dementia complex in the era of highly active antiretroviral therapy'. PubMed, Jul 9, 13(10), 1249−53.

[7] D. J. Gregory, C. K. Patricia, Y. M. Li,; J. M. Kaldor,.; C. A. David, B. J. Bruce (1999) 'Epidemiology and Social', AIDS, Vol.13, No.10, pp.1249−1253, July 9.

[8] H. L. Wu. (2000) 'Modeling the HIV epidemic: a state-space approach'. Mathematical and Computer Modeling, Vol. 32, No. 1−2, pp.197−215.

[9] Ministry of Health of the People's Republic of China (2003) 'Report on the Chinese AIDS Epidemic (Chinese version, December 2003)', December 1.

[10] UNAIDS/WHO. (2003) 'Overview of making estimates of HIV/AIDS and its impact in countries with low-level or concentrated epidemics: The Workbook Method'. UNAIDS/WHO, June.

[11] UNAIDS. (2004) 'Report on the global AIDS epidemic (English original, June 2004)'.

[12] UNGASS. (2003) 'Program monitoring data from UNGASS (United Nations General Assembly Special Session) country reports 2003'.

[13] UNAIDS. (2000) 'UNAIDS Epidemiological Fact Sheets by Country', June.

[14] United Nations. (2005) 'Population Division, Department of Economic and Social Affairs (2005) World Population Prospects: The 2004 Revision'.

[15] I. Foster, C. Kesselman. (1999) The Grid Blueprint for a New Computing Infrastructure [M]. Morgan Kaufmann Publishers, Inc., 6−8.

[16] I Foster, C Kesselman, S Tueeke. (2001) The anatomy of the grid: Enabling scalable virtual organizations. International Journal of Supercomputer Applications, 15(3): 200−222.

[17] Z. W. Xu, W. Li, L. Zha, *et al*. (2004) Vega: A Computer Systems Approach to Grid Computing. Journal of Grid Computing. Vol. 2(2): 109−120.

[18] Y. L. Gong, F. P. Dong, W. Li, Z. W. Xu. (2003) VEGA Infrastructure for Resource Discovery in Grids. Journal of Computer Science & Technology. Vol. 18, No.4, 413−422.

[19] W Allcock, J Bresnahan, J Bester. *et a1*. (2002) Grid FTP Protocol Specification [Z]. GGF Grid FTP Working Group Document.

[20] J Bester, I Foster, C Kessdman *et a1*. (1999) GASS：A data movement and access service for wide a computing systems [A]. Sixth Workshop on I/O in Parallel and Distributed System[C].

[21] S. Y. Liu, C. Liu, C. Z. Zhao and Yu Liu. (2007) Mathematical Models and Optimization Discussions on EA System on AIDS/HIV Spread Estimating and Countermeasures Evaluating. IEEE 7th International Symposium on Bioinformatics and Bioengineering.

[22] D. Patterson, A. Brown *el a1*. (2002) Recover oriented computing (ROC): Motivation, definition, techniques, and ease studies. UC Berkeley, Tech Rep: UCB/CSD-02-1175.

[23] L. Chen, C.L. Wang, F. C. M. Lau, and R. K. K. Ma, (2002) "A Grid Middleware for Distributed Java Computing with MPI Binding and Process Migration Supports," International Workshop on Grid and Cooperative Computing (GCC-2002), December 26−28, Hainan, China, pp. 640−652.

[24] M. J. M. Ma, C. L. Wang, F. C. M. Lau. (2000) "JESSICA: Java-Enabled Single-System-Image Computing Architecture," Journal of Parallel and Distributed Computing, Vol. 60, No. 10, pp. 1194−1222.

[25] H. Stockinger, F. Donno, E. Laure, S. Muzaffar, P. Kunszt, G. Andronico, P. Millar. (2003) Grid Data Management in Action: Experience in Running and Supporting Data Management Services in the EU Data Grid Project. Computer Science.

[26] K. Gor, D. Ra, S. Ali, L. Alves, N. Arurkar, I. Gupta, A. Chakrabarti, A. Sharma, S. Sengupta. (2005) Scalable enterprise level workflow and infrastructure management in a grid computing environment. Cluster Computing and the Grid.

[27] K. Yang, A. Galis, C. Todd, (2002) Policy-based active Grid management architecture. Networks. ICON 2002.

[28] A. Boukerche, A. Roy. (2002) Dynamic Grid-Based Approach to Data Distribution Management. Journal of Parallel and Distributed Computing. Volume 62, Issue 3, Pages 366−392.

[29] A. -M. Vandamme, K. Van Laethem, and E. De Clercq, (1999) "Managing Resistance to Anti-HIV Drugs: An Important Consideration for Effective Disease Management," Drugs, Vol. 57, >No. 3, 337−361.

[30] H. Walter *et al*., (1999) "Rapid, Phenotypic HIV-1 Drug Sensitivity Assay for Protease and Reverse Transcriptase Inhibitors," J. Clinical Virology, Vol. 13, Nos. 1−2, 71−80.

[31] D. Wang, S. Bloor, and B. A. Larder, (2000) "The Application of Neural Networks in Predicting Phenotypic Resistance from Genotypes for HIV-1 Protease Inhibitors," Antiviral Therapy, Vol. 5, supplement 3, 51−52.

[32] J. R. Quinlan, (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco.

[33] C. J. C. Burges, (1998) "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, Vol. 2, No. 2, 121−167.

[34] T. Joachims, (1999) "Making Large-Scale Support Vector Machine Learning Practical," Advances in Kernel Methods: Support Vector Learning, B. Schölkopf, C. Burges, and A. Smola, eds., MIT Press, Cambridge, Mass., 169−184.

[35] J. Selbig, T. Mevissen, and T. Lengauer, (1999) "Decision Tree-Based Formation of Consensus Protein Secondary Structure Prediction," Bioinformatics, Vol. 15, No. 12, 1039−1046.

[36] X. Shen, W. K. Liao, A. Choudhary, G. Memik, M. Kandemir, (2003) "A high-performance application data environment for large-scale scientific computations," IEEE Transactions on Parallel and Distributed Systems, Volume: 14 Issue: 12 Date: Page(s): 1262−1274.

J. Biomedical Science and Engineering, 2009, 2, 51-56

Scientific Research Publishing

Using position specific scoring matrix and auto covariance to predict protein subnuclear localization

Rong-Quan Xiao[1], Yan-Zhi Guo[2], Yu-Hong Zeng[2], Hai-Feng Tan[1], Xue-Mei Pu[2], Meng-Long Li[1,2*]

[1]College of Life Sciences, Sichuan University, Chengdu 610064. [2]College of Chemistry, Sichuan University, Chengdu 610064, P.R. China. Correspondence should be addressed to Meng-Long Li(liml@scu.edu.cn). Tel: +86 28 89005151; Fax: +86 28 85412356.

Received September 8th, 2008; revised November 13th, 2008; accepted November 20th, 2008

ABSTRACT

The knowledge of subnuclear localization in eukaryotic cells is indispensable for understanding the biological function of nucleus, genome regulation and drug discovery. In this study, a new feature representation was proposed by combining position specific scoring matrix (PSSM) and auto covariance (AC). The AC variables describe the neighboring effect between two amino acids, so that they incorporate the sequence-order information; PSSM describes the information of biological evolution of proteins. Based on this new descriptor, a support vector machine (SVM) classifier was built to predict subnuclear localization. To evaluate the power of our predictor, the benchmark dataset that contains 714 proteins localized in nine subnuclear compartments was utilized. The total jackknife cross validation accuracy of our method is 76.5%, that is higher than those of the Nuc-PLoc (67.4%), the OET-KNN (55.6%), AAC based SVM (48.9%) and ProtLoc (36.6%). The prediction software used in this article and the details of the SVM parameters are freely available at http://chemlab.scu.edu.cn/predict_SubNL/index.htm and the dataset used in our study is from Shen and Chou's work by downloading at http://chou.med.harvard.edu/bioinf/Nuc-PLoc/Data.htm.

Keywords: Position Specific Scoring Matrix; Auto Covariance; Support Vector Machine; Protein Subnuclear Localization Prediction

1. INTRODUCTION

The cell nucleus is complex, important subcellular organelle in eukaryotes cell. It organizes the comprehensive assembly of our genes and their corresponding regulatory factors [1]. Meanwhile, it also reflects various intricate biological activities, and controls various kinds of biologic processes [2]. Many proteins, from outside a nuclear, trend to be localized into specific subnuclear locations of the nucleus [3]. If proteins can not be correctly localized into its specific subnuclear locations in human, it will lead to genetic disease [4], cancer [5] or virally infected cells [6]. Thus, it's desirable to get the knowledge of protein subnuclear localization for in-depth understanding cell biological processes and genomic regulation. However, it is costly and time-consuming to assay the subnuclear localization of proteins by biology experiments [7]. The number of protein sequences is increasing more rapidly than that of identified proteins [7]. So it is of great practical significance to develop computational approaches for identifying the protein subnuclear localizations in cell nucleus. At the same time, many lines of evidences have indicated that computational approaches, such as structural bioinformatics [8], molecular docking [9], pharmacophore modelling [10], QSAR [11,12,13], protein subcellular location prediction [7,14], identification of membrane proteins and their types [15], identification of enzymes and their functional classes [16], identification of proteases and their types [17], protein cleavage site prediction [18,19], and signal peptide prediction [20,21] can provide very useful information for both basic research and drug discovery in a timely manner. The present study is devoted to develop a new method for predicting protein subnuclear localization in hope to stimulate the development of the relevant areas.

Recently, many algorithms have already been developed for predicting protein subcellular localizations [22, 23,24,25,26,27,28,29,30,31,32,33], as reviewed by Chou [7]. Even several web severs have been constructed for predicting subcellular localization of various organisms [14,34,35,36,37]. However, there are only a few computational methods for predicting protein subnuclear localization [38,39,40,41], such as OET-KNN [42], ProLoc [43], Nuc-PLoc [44], and AdaBoost classifiers [45].

Compared to the conventional amino acid composition (AAC), pseudo amino acid (PseAA) composition [46], originally introduced by Chou [47,48], can include the sequence-order information of sequences. Similarly, the PsePSSM was also proposed by Shen and Chou in order to incorporate the evolution information of proteins [44]. They built a new web server called Nuc-PLoc for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM with a promising prediction result. In this study, we developed a new method by fus-

ing position specific scoring matrix (PSSM) and auto covariance (AC), so that this method can incorporate sequence-order information by AC and the evolutionary information by PSSM. A classifier based on SVM was constructed to predict protein subnuclear localization using jackknife test. The result indicates that our method has successfully enhanced accuracies of the existing methods for predicting protein subnuclear localization.

2. MATERIALS AND METHODS

2.1. Data Sets

In this paper, our dataset is obtained from article by Shen and Chou [44]. And anyone can freely download it at this page (http://chou.med.harvard.edu/bioinf/Nuc-PLoc/Data.htm). This dataset consists of nine classes and 714 proteins in total. Details of this benchmark dataset are shown in **Table 1.** S_i (i=1, 2… 9) is used to represent each of nine subsets and S represents the total dataset.

2.2. Feature Representations

2.2.1. Auto Covariance (AC)

We selected three common physicochemical properties, hydrophobicity [49], volumes of side chains of amino acids [50], and polarity [51], to represent the structure and function [52], the stereospecific blockade [53] and the electronic property [54] of residues in a protein respectively. These original values were taken from Guo *et al*. [55] and were first normalized to zero mean value and unit standard deviation (SD) by Equation (1):

$$P'_{i,j} = \frac{P_{i,j} - \overline{P}_j}{S_j} \quad (1)$$

(i=1, 2, 3; j=1, 2, 3…, 20.)

Where $P_{i,j}$ is the *i-th* descriptor value for *j-th* amino acid, $\overline{P}_j$ is the mean of the *j-th* descriptor of the 20 amino acids and S_j is the value of SD. So each protein sequence was translated into three vectors with each amino acid represented by the normalized values.

There are many approaches to convert the protein sequences into numerical order sequences, including autocorrelations and auto covariance (AC). Autocorrelations, quite similar to AC, has been used in the prediction of secondary structure content [56,57,58] and structural class [59,60,61,62]; however, AC as a statistical tool for analyzing sequences of vectors has also been successfully adopted by our research group for protein classifications [55,63] from primary sequence. So in our study, AC was selected to transform these numerical vectors into uniform matrices in order to take the neighboring effect of the sequences into account. Here, *lag* is the distance between one residue and its neighbour, a certain number of residues away. The AC variables are calculated by the Equation (2) [55].

$$AC_{lag,j} = \frac{1}{L-lag}\sum_{i=1}^{L-lag}(P_{i,j} - \frac{1}{L}\sum_{i=1}^{L}P_{i,j}) \times (P_{(i+lag),j} - \frac{1}{L}\sum_{i=1}^{L}P_{i,j}) \quad (2)$$

Where i is the position in the sequence P, j is one descriptor, L is the length of the sequence P and *lag* is the value of the lag.

In this way, the number of AC variables, D, can be calculated according to Equation (3) [55].

$$D = lg \times p \quad (3)$$

Where *lg* is the maximum *lag* (*lag*=1, 2, 3…, *lg*) and p represents the number of descriptors.

2.2.2. Position Specific Scoring Matrix (PSSM)

A PSSM is a Position Specific Scoring Matrix and is a commonly used representation of motifs (patterns) in biological sequences [64]. So far, this method has been used for predicting protein subcellular localization [65] and subnuclear localization [40,44].

For a protein sequence P with L amino acid residues, PSSM is obtained according to the following Equation [44].

$$P_{PSSM} = \begin{bmatrix} P_{1\to1} & P_{1\to2} & \cdots & P_{1\to j} & \cdots & P_{1\to20} \\ P_{2\to1} & P_{2\to2} & \cdots & P_{2\to j} & \cdots & P_{2\to20} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{i\to1} & P_{i\to2} & \cdots & P_{i\to j} & \cdots & P_{i\to20} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{L\to1} & P_{L\to2} & \cdots & P_{L\to j} & \cdots & P_{L\to20} \end{bmatrix} \quad (4)$$

In Equation (4), where $i \to j$ describes *i-th* amino acid residue of the protein sequence P being mutated to amino acid type j in the biology evolution process, $P_{i\to j}$ is the score of this mutation and L is the length of the sequence P. Here we used the numerical codes 1, 2, 3… 20 to represent the single character of ordered 20 native amino acid types in Equation (4). To get the $L \times 20$ scores of the P_{PSSM} in the Equation (4), we used three iterations of PSI-BLAST [66] with default threshold (the default E-value is 0.001) to search the Swiss-Prot database (version 54.4, released on 25 Oct. 2007) for multiple sequence alignment against the protein P. Then, the value of $P_{i\to j}$ is standardized by Equation (5), as given below.

$$P_{i\to j} = \frac{P^o_{i\to j} - \frac{1}{20}\sum_{j=1}^{20}P^o_{i\to j}}{\max(P^o_{i\to j}) - \min(P^o_{i\to j})} \quad (5)$$

Table 1. The benchmark dataset consists of 714 nuclear proteins classified into nine subnuclear localizations

Subnuclear localization	Subset	No. of proteins
Chromatin	S_1	99
Heterochromatin	S_2	22
Nuclear envelope	S_3	61
Nuclear matrix	S_4	29
Nuclear pore complex	S_5	79
Nuclear speckle	S_6	67
Nucleolus	S_7	307
Nucleoplasm	S_8	37
Nuclear PML body	S_9	13
Total	S	714

(i= 1, 2, 3… L; j= 1, 2, 3…20)

Where $\boldsymbol{P}^{o}_{i\to j}$ is the original scores generated by PSI-BLAST, $\boldsymbol{P}_{i\to j}$ is a zero mean value over the 20 native amino acids and the value is between -1 and 1. However, because of proteins with different lengths L, the matrices of the PSSM descriptor in Equation (4) have different numbers of rows. To gain the uniform matrix for protein sequences of different lengths, we converted the PSSM of protein P to a uniform vector through the Equation (6) [44].

$$\overline{P}_{PSSM} = \left[\overline{P}_1 \quad \overline{P}_2 \quad \cdots \quad \overline{P}_j \quad \cdots \quad \overline{P}_{20}\right]^{\mathrm{T}} (j=1,2,\cdots,20) \quad (6)$$

Where T is the transpose operator, $\overline{P}_j$ is the average score over *j-th* column in Equation (4).

Finally, the $\overline{P}_{PSSM}$ describes the evolutionary information of a protein sample, and AC variables contain the interaction information between two amino acid residues of a sequence. So each protein sequence was converted into a numerical vector by concatenating PSSM and AC. Here, each AC variable was appended a weight factor of 0.05.

2.2.3. Accuracy and Matthew's Correlation Coefficient (MCC)

To evaluate the performance of this method, two parameters, accuracy and Matthew's correlation coefficient (MCC), were selected in this article. They are calculated by Equation (7) and Equation (8), respectively.

$$Accuracy = \frac{TP}{TP+FN} \quad (7)$$

$$MCC = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \quad (8)$$

Where TP represents the true positive; TN, the true negative; FP, the false positive and FN, the false negative.

3. RESULTS AND DISCUSSION

In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, subsampling test, and jackknife test [67]. However, as elucidated in [14] and demonstrated by Eq.50 of [7], among the three cross-validation methods, the jackknife test is deemed the most objective that can always yield a unique result for a given benchmark dataset, and hence has been increasingly used by tigators to examine the accuracy of various predictors (see, e.g., [7,33,68,69,70,71,72,73,74,75,76,77,78,79,80, 81,82]). So in this paper, the jackknife test was chosen to validate the current algorithm. Because the benchmark dataset used has nine subsets, the one-to-one multiclass classification system led to 9*(9-1)/2=36 SVM models for one single encoding methods. Meanwhile, for AC variables, the value of *lg* was optimized as 13 through a series of control experiments, and the value of p is 3. So, the number of AC variables, D, is 39 ($D = lg \times p = 13\times 3 = 39$) according to Equation (3).

Amino acid composition (AAC) has been widely used for predicting subcellular localizations [7,14,22,23,24, 25,26,27,28,30,31,32,34,35,36,37,83,84,85], so it was also used as a substitution model in our study. And thus, three SVM models based on AAC, AC and PSSM, were respectively constructed.

The results according to jackknife test are listed in **Table 2**. As can be seen from **Table 2**, the prediction accuracy of PSSM based model is nearly equal to that of AAC based model. However, AC based model gives the lower accuracy of 64.13%. Then we constructed models by fusing the three substitution models, so four fused classifier were built. **Table 2** shows that the accuracies of the four fused models are higher than those of the three anterior models. Among those four fused models, the accuracy of the model combining PSSM, AAC and AC is lower than that of PSSM and AC based model that obtains the best performance with an accuracy of 76.45%. So the final SVM model was built based on PSSM and AC. The kernel function of SVM is radio basis function (rbf), and the parameters of C and γ are listed in the table by downloading at http://chemlab.scu.edu.cn/predict_SubNL/index.htm.

In order to further examine the prediction power of the current classifier, the performance of this method was also compared with those of the existing methods on the same training dataset. The results obtained by several algorithms with different substitution models were summarized in **Table 3**. From **Table 3**, we can see that the accuracy obtained by Nuc-PLoc [44] is much higher than those of ProtLoc [43], AAC based SVM and OET-KNN [42]. When compared to Nuc-PLoc, our method obtains a better performance with the accuracy of 76.5%. It means our method is successful in predicting protein subnuclear localization only using primary sequences of proteins

Table 2. Overall accuracies by jackknife tests with different substitution models on the benchmark dataset of Table 1

Substitution Model	AAC[a]	AC[b]	PSSM[c]	AAC+AC[d]	PSSM+AAC	PSSM+AC[d]	PSSM+AAC+AC[d]
Accuracy	73.82%	64.13%	73.85%	74.05%	75.97%	76.45%	75.99%

a: Amino acid composition
b: Auto covariance
c: Position specific scoring matrix
d: While fused models were constructed, a weight factor added on AC is 0.05.

Table 3. Overall accuracy by jackknife tests with different algorithms on the benchmark dataset of Table 1

Algorithm	Protein sample descriptor	Overall accuracy
ProtLoc[a,d]	Amino acid composition	261/714=36.6%
SVM[d]	Amino acid composition	349/714=48.9%
OET-KNN[b,d]	PseAA Composition	397/714=55.6%
Nuc-PLoc[c,d]	Fusion of PsePSSM and PseAA Composition	481/714=67.4%
Our method	Combination of PSSM and AC	546/714=76.5%

a: See Cedano *et al*. (1997)[86]
b: See Shen and Chou (2005)[42]
c: See Shen and Chou (2007)[44]
d: The results were from Shen and Chou (2007)[44], and the original data could been seen in that article.

Table 4. The MCC values obtained by the jackknife tests with Nuc-PLoc and our method on the benchmark dataset of Table 1

Subnuclear localization	Matthew's correlation coefficient	
	Nuc-PLoc[a]	Our method[b]
Chromatin S_1	0.60	0.55
Heterochromatin S_2	0.52	0.58
Nuclear envelope S_3	0.53	0.65
Nuclear matrix S_4	0.52	0.61
Nuclear pore complex S_5	0.70	0.72
Nuclear speckle S_6	0.43	0.57
Nucleolus S_7	0.57	0.57
Nucleoplasm S_8	0.31	0.54
Nuclear PML body S_9	0.32	0.51

a: The results were from Shen and Chou (2007)[44], and the original data could be seen in that article.
b: The classifier fused PSSM and AC.

In addition, to evaluate the stability of our method, the values of the MCC for the nine subsets were compared based on Nuc-PLoc and our current predictor, respectively, as seen in **Table 4**. For nine subsets, our method yields a higher MCC than Nuc-PLoc, except the subset S_1. So, compared to the existing methods, our classifier combined with PSSM and AC has further improved the prediction accuracy of protein subnuclear localization.

4. CONCLUSION

In this paper, a new classifier was developed by fusing PSSM and AC for predicting protein subnuclear localization only using the primary sequences of nuclear proteins. The SVM predictor was constructed based on PSSM and AC. AC variables represent the interactions between amino acids in protein sequences; PSSM describes the evolutionary information. So the method incorporated not only the evolution information, but also the sequence-order information. Compared with the current methods, this method successfully raises the prediction accuracy. Hence, it may be a good supplementary tool for protein function studies.

ACKNOWLEDGEMENT

The authors gratefully thank Shen and Chou for sharing the benchmark dataset. The work was funded by the National Natural Science Foundation of China (No. 20775052).

REFERENCES

[1] G. S. Stein, J. B. Lian, W. A. van, J. L. Stein, A. Javed, G. Barnes, L. Gerstenfeld, D. Vradii, S. K. Zaidi, and J. Pratap, *et al*. (2006) Organization of transcriptional regulatory machinery in nuclear microenvironments: Implications for biological control and cancer. Cancer Treatment Reviews, 32, 13−13.

[2] A. I. Lamond, and W. C. Earnshaw, (1998) Structure and function in the nucleus. Science, 280, 547−553.

[3] J. M. Bridger, and W. A. Bickmore, (1998) Putting the genome on the map. Trends in Genetics, 14, 403−409.

[4] H. G. Sutherland, G. K. Mumford, K. Newton, L. V. Ford, R. Farrall, G. Dellaire, J. F. Caceres, and W. A. Bickmore, (2001) Large-scale identification of mammalian proteins localized to nuclear sub-compartments. Human Molecular Genetics, 10, 1995−2011.

[5] S. K. Zaidi, D. W. Young, A. Javed, J. Pratap, M. Montecino, W. A. van, J. B. Lian, J. L. Stein, and G. S. Stein, (2007) Nuclear microenvironments in biological control and cancer. Nature Reviews Cancer, 7, 454−463.

[6] R. D. Phair, and T. Misteli, (2000) High mobility of proteins in the mammalian cell nucleus. Nature, 404, 604−609.

[7] K. C. Chou, and H. B. Shen, (2007) Recent progress in protein subcellular location prediction. Analytical Biochemistry, 370, 1−16.

[8] K. C. Chou, (2004) Structural bioinformatics and its impact to biomedical science. Current Medicinal Chemistry, 11, 2105−2134.

[9] K. C. Chou, D. Q. Wei, and W. Z. Zhong, (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. Biochemical and Biophysical Research Communications, 308, 148−151.

[10] Sirois, S., Wei, D. Q., Du, Q. S. and Chou, K. C. (2004) Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore pointst. Journal of Chemical Information and Computer Sciences, 44, 1111−1122.

[11] Du, Q. S., Mezey, P. G. and Chou, K. C. (2005) Heuristic molecular lipophilicity potential (HMLP): A 2D-QSAR study to LADH of molecular family pyrazole and derivatives. Journal of Computational Chemistry, 26, 461−470.

[12] Du, Q. S., Huang, R. B., Wei, Y. T., Du, L. Q. and Chou, K. C. (2008) Multiple field three dimensional quantitative structure-activity relationship (MF-3D-QSAR). Journal of Computational Chemistry, 29, 211−219.

[13] Prado-Prado, F. J., Gonzalez-Diaz, H., de, V. O., Ubeira, F. M. and Chou, K. C. (2008) Unified QSAR approach to antimicrobials. Part 3: First multi-tasking QSAR model for Input-Coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. Bioorganic and Medicinal Chemistry, 16, 5871−5880.

[14] Chou, K. C. and Shen, H. B. (2008) Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. Nature Protocols, 3, 153−162.

[15] Chou, K. C. and Shen, H. B. (2007) MemType-2L: a web server

for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochemical and Biophysical Research Communications, 360, 339−345.
[16] Shen, H. B. and Chou, K. C. (2007) EzyPred: a top-down approach for predicting enzyme functional classes and subclasses. Biochemical and Biophysical Research Communications, 364, 53−59.
[17] Chou, K. C. and Shen, H. B. (2008) ProtIdent: A web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochemical and Biophysical Research Communications, 376, 321−325.
[18] Chou, K. C. (1996) Prediction of human immunodeficiency virus protease cleavage sites in proteins. Analytical Biochemistry, 233, 1−14.
[19] Shen, H. B. and Chou, K. C. (2008) HIVcleave: a web-server for predicting human immunodeficiency virus protease cleavage sites in proteins. Analytical Biochemistry, 375, 388−390.
[20] Chou, K. C. and Shen, H. B. (2007) Signal-CF: A subsite-coupled and window-fusing approach for predicting signal peptides. Biochemical and Biophysical Research Communications, 357, 633−640.
[21] Shen, H. B. and Chou, K. C. (2007) Signal-3L: A 3-layer approach for predicting signal peptides. Biochemical and Biophysical Research Communications, 363, 297−303.
[22] Chou, K. C. (2000) Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. Biochemical and Biophysical Research Communications, 278, 477−483.
[23] Chou, K. C. and Cai, Y. D. (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. Journal of Biological Chemistry, 277, 45765−45769.
[24] Cai, Y. D., Liu, X. J., Xu, X. B. and Chou, K. C. (2002) Support vector machines for prediction of protein subcellular location by incorporating quasi-sequence-order effect. Journal of Cellular Biochemistry, 84, 343−348.
[25] Cai, Y. D., Liu, X. J. and Chou, K. C. (2002) Artificial neural network model for predicting protein subcellular location. Computers and Chemistry, 26, 179−182.
[26] Chou, K. C. and Cai, Y. D. (2004) Prediction and classification of protein subcellular location-sequence-order effect and pseudo amino acid composition. Journal of Cellular Biochemistry, 90, 1250−1260.
[27] Chou, K. C. and Cai, Y. D. (2004) Prediction of protein subcellular locations by GO-FunD-PseAA predictor. Biochemical and Biophysical Research Communications, 320, 1236−1239.
[28] Gao, Q. B., Wang, Z. Z., Yan, C. and Du, Y. H. (2005) Prediction of protein subcellular location using a combined feature of sequence. Febs Letters, 579, 3444−3448.
[29] Zhang, T. L., Ding, Y. S. and Chou, K. C. (2006) Prediction of protein subcellular location using hydrophobic patterns of amino acid sequence. Computational Biology and Chemistry, 30, 367−371.
[30] Chou, K. C. and Shen, H. B. (2006) Predicting protein subcellular location by fusing multiple classifiers. Journal of Cellular Biochemistry, 99, 517−527.
[31] Chou, K. C. and Shen, H. B. (2006) Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-Nearest Neighbor classifiers. Journal of Proteome Research, 5, 1888−1897.
[32] Zhou, X. B., Chen, C., Li, Z. C. and Zou, X. Y. (2008) Improved prediction of subcellular location for apoptosis proteins by the dual-layer support vector machine. Amino Acids, 35, 383−388.
[33] Shi, F., Chen, Q. J. and Li, N. N. (2008) Hilbert Huang transform for predicting proteins subcellular location. Journal of Biomedical Science and Engineering, 1, 59−63.
[34] Chou, K. C. and Shen, H. B. (2006) Hum-PLoc: A novel ensemble classifier for predicting human protein subcellular localization. Biochemical and Biophysical Research Communications, 347, 150−157.
[35] Shen, H. B., Yang, J. and Chou, K. C. (2007) Euk-PLoc: an ensemble classifier for large-scale eukaryotic protein subcellular location prediction. Amino Acids, 33, 57−67.
[36] Shen, H. B. and Chou, K. C. (2007) Hum-mPLoc: An ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites. Biochemical and Biophysical Research Communications, 355, 1006−1011.
[37] Chou, K. C. and Shen, H. B. (2007) Euk-mPLoc: A fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. Journal of Proteome Research, 6, 1728−1734.
[38] Lei, Z. D. and Dai, Y. (2005) An SVM-based system for predicting protein subnuclear localizations. BMC Bioinformatics, 6, 291−298.
[39] Lei, Z. D. and Dai, Y. (2006) Assessing protein similarity with Gene Ontology and its use in subnuclear localization prediction. BMC Bioinformatics, 7, 491−500.
[40] Mundra, P., Kumar, M., Kumar, K. K., Jayaraman, V. K. and Kulkarni, B. D. (2007) Using pseudo amino acid composition to predict protein subnuclear localization: Approached with PSSM. Pattern Recognition Letters, 28, 1610−1615.
[41] Li, F. M. and Li, Q. Z. (2008) Using pseudo amino acid composition to predict protein subnuclear location with improved hybrid approach. Amino Acids, 34, 119−125.
[42] Shen, H. B. and Chou, K. C. (2005) Predicting protein subnuclear location with optimized evidence-theoretic K-nearest classifier and pseudo amino acid composition. Biochemical and Biophysical Research Communications, 337, 752−756.
[43] Huang, W. L., Tung, C. W., Huang, H. L., Hwang, S. F. and Ho, S. Y. (2007) ProLoc: Prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features. Biosystems, 90, 573−581.
[44] Shen, H. B. and Chou, K. C. (2007) Nuc-PLoc: a new web-server for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM. Protein Engineering Design and Selection, 20, 561−567.
[45] Jiang, X. Y., Wei, R., Zhao, Y. J. and Zhang, T. L. (2008) Using Chou's pseudo amino acid composition based on approximate entropy and an ensemble of AdaBoost classifiers to predict protein subnuclear location. Amino Acids, 34, 669−675.
[46] Shen, H. B. and Chou, K. C. (2008) PseAAC: A flexible web server for generating various kinds of protein pseudo amino acid composition. Analytical Biochemistry, 373, 386−388.
[47] Chou, K. C. (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins-Structure Function and Genetics, 43, 246−255.
[48] Chou, K. C. (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics, 21, 10−19.
[49] Tanford, C. (1962) Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. Journal of the American Chemical Society, 84, 4240−4247.
[50] Krigbaum, W. R. and Komoriya, A. (1979) Local interactions as a structure determinant for protein molecules: II. Biochimica et Biophysica Acta, 576, 204−248.
[51] Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. Science, 185, 862−864.
[52] Guo, Y. Z., Li, M., Lu, M., Wen, Z., Wang, K., Li, G. and Wu, J. (2006) Classifying G protein-coupled receptors and nuclear receptors on the basis of protein power spectrum from fast Fourier transform. AMINO ACIDS, 30, 397−402.
[53] Hackel, M., Hinz, H. J. and Hedwig, G. R. (1999) Partial molar volumes of proteins: amino acid side-chain contributions derived from the partial molar volumes of some tripeptides over the temperature range 10−90 degrees C. BIOPHYSICAL CHEMISTRY, 82, 35−50.
[54] Guo, Y. Z., Li, M. L., Wang, K. L., Wen, Z. N., Lu, M. C., Liu, L. X. and Jiang, L. (2006) Fast fourier transform-based support vector machine for prediction of G-protein coupled receptor subfamilies. (Vol 37, pg 759, 2005). ACTA BIOCHIMICA ET BIOPHYSICA SINICA, 38, 456−456.
[55] Guo, Y.Z., Yu, L.Z., Wen, Z.N. and Li, M.L. (2008) Using sup-

port vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. Nucleic Acids Research, 36, 3025−3030.

[56] Lin, Z. and Pan, X. M. (2001) Accurate prediction of protein secondary structural content. Journal Of Protein Chemistry, 20, 217−220.

[57] Zhang, C. T., Lin, Z. S., Zhang, Z. D. and Yan, M. (1998) Prediction of the helix/strand content of globular proteins based on their primary sequences. Protein Engineering, 11, 971−979.

[58] Zhang, Z. D., Sun, Z. R. and Zhang, C. T. (2001) A new approach to predict the helix/strand content of globular proteins. Journal Of Theoretical Biology, 208, 65−78.

[59] Kedarisetti, K. D., Kurgan, L. and Dick, S. (2006) Classifier ensembles for protein structural class prediction with varying homology. Biochemical And Biophysical Research Communications, 348, 981−988.

[60] Kurgan, L. A. and Homaeian, L. (2006) Prediction of structural classes for protein sequences and domains-Impact of prediction algorithms, sequence representation and homology, and test procedures on accuracy. Pattern Recognition, 39, 2323−2343.

[61] Li, X. and Pan, X. M. (2001) New method for accurate prediction of solvent accessibility from protein sequence. Proteins-Structure Function And Genetics, 42, 1−5.

[62] Kurgan, L. and Chen, K. (2007) Prediction of protein structural class for the twilight zone sequences. Biochemical And Biophysical Research Communications, 357, 453−460.

[63] Guo, Y. Z., Li, M. L., Lu, M. C., Wen, Z.N. and Huang, Z. T. (2006) Predicting G-protein coupled receptors-G-protein coupling specificity based on autocross-covariance transform. Proteins, 65, 55−60.

[64] Ben-Gal, I., Shani, A., Gohr, A., Grau, J., S, A., Shmilovici, A., Posch, S. and Grosse, I. (2005) Identification of transcription factor binding sites with variable-order Bayesian networks. Bioinformatics, 21, 2657−2666.

[65] Xie, D., Li, A., Wang, M. H., Fan, Z. W. and Feng, H. Q. (2005) LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. Nucleic Acids Research, 33, 105−110.

[66] Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V. and Altschul, S. F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Research, 29, 2994−3005.

[67] Chou, K. C. and Zhang, C. T. (1995) PREDICTION OF PROTEIN STRUCTURAL CLASSES. Critical Reviews in Biochemistry and Molecular Biology, 30, 275−349.

[68] Chen, Y. L. and Li, Q. Z. (2007) Prediction of the subcellular location of apoptosis proteins. Journal of Theoretical Biology, 245, 775−783.

[69] Chen, Y. L. and Li, Q. Z. (2007) Prediction of apoptosis protein subcellular location using improved hybrid approach and pseudo-amino acid composition. Journal of Theoretical Biology, 248, 377−381.

[70] Zhou, X. B., Chen, C., Li, Z. C. and Zou, X. Y. (2007) Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. Journal of Theoretical Biology, 248, 546−551.

[71] Chen, C., Chen, L. X., Zou, X. Y. and Cai, P. X. (2008) Predicting protein structural class based on multi-features fusion. Journal of Theoretical Biology, 253, 388−392.

[72] Chen, K., Kurgan, L. A. and Ruan, J. S. (2008) Prediction of protein structural class using novel evolutionary collocation-based sequence representation. Journal of Computational Chemistry, 29, 1596−1604.

[73] Du, P. F. and Li, Y. D. (2008) Prediction of C-to-U RNA editing sites in plant mitochondria using both biochemical and evolutionary information. Journal of Theoretical Biology, 253, 579−586.

[74] Jiang, X. Y., Wei, R., Zhang, T. L. and Gu, Q. (2008) Using the concept of Chou's Pseudo Amino Acid composition to predict apoptosis proteins subcellular location: An approach by approximate entropy. Protein and Peptide Letters, 15, 392−396.

[75] Jin, Y. H., Niu, B., Feng, K. Y., Lu, W. C., Cai, Y. D. and Li, G. Z. (2008) Predicting subcellular localization with AdaBoost Learner. Protein and Peptide Letters, 15, 286−289.

[76] Li, F. M. and Li, Q. Z. (2008) Protein Subcellular Location Using Chou's Pseudo Amino Acid Composition and Improved Hybrid Approach. Protein and Peptide Letters, 15, 612−616.

[77] Lin, H. (2008) The modified Mahalanobis Discriminant for predicting outer membrane proteins by using Chou's pseudo amino acid composition. Journal of Theoretical Biology, 252, 350−356.

[78] Lin, H., Ding, H., Guo, F. B., Zhang, A. Y. and Huang, J. (2008) Predicting Subcellular Localization of Mycobacterial Proteins by Using Chou's Pseudo Amino Acid Composition. Protein and Peptide Letters, 15, 739−744.

[79] Niu, B., Jin, Y. H., Feng, K. Y., Liu, L., Lu, W. C., Cai, Y. D. and Li, G. Z. (2008) Predicting membrane protein types with bragging learner. Protein and Peptide Letters, 1**5**, 590−594.

[80] Wang, T., Yang, J., Shen, H. B. and Chou, K. C. (2008) Predicting membrane protein types by the LLDA algorithm. Protein and Peptide Letters, 15, 915−921.

[81] Wu, G. and Yan, S. M. (2008) Prediction of mutations in H3N2 hemagglutinins of influenza A virus from North America based on different datasets. Protein and Peptide Letters, 15, 144−152.

[82] Zhang, G. Y. and Fang, B. S. (2008) Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo-amino acid composition. Journal of Theoretical Biology, 253, 310−315.

[83] Cai, Y. D. and Chou, K. C. (2003) Nearest neighbour algorithm for predicting protein subcellular location by combining functional domain composition and pseudo-amino acid composition. Biochemical and Biophysical Research Communications, 305, 407−411.

[84] Park, K. J. and Kanehisa, M. (2003) Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. Bioinformatics, 19, 1656−1663.

[85] Chou, K. C. and Cai, Y. D. (2004) Predicting subcellular localization of proteins by hybridizing functional domain composition and pseudo-amino acid composition. Journal of Cellular Biochemistry, 91, 1197−1203.

[86] Cedano, J., Aloy, P., PerezPons, J. A. and Querol, E. (1997) Relation between amino acid composition and cellular location of proteins. Journal of Molecular Biology, 266, 594−600.

J. Biomedical Science and Engineering, 2009, 2, 57-62

Down regulation of surviving gene and up regulation of p53 gene expression by siRNA induces apoptosis in human hepatocellular carcinoma cell line HepG2

Yun-Hua Lu [1,2,3], Cong Tang [1,2] , Wei Wang [1,2], Tao Xi [1,2*]

[1]Research Centre of Biotechnology, China Pharmaceutical University, NanJing, 210009, China. [2]JiangSu Key Laboratory of Carcinogenesis and intervention, China Pharmaceutical University. [3]School of Chemistry & Biotechnology, YiChun University, YiChun, JiangXi, 33600, China. *Correspondence should be addressed to Yun-Hua Lu (lyh655@yahoo.com.cn)

Received April 16th, 2008; revised October 31st, 2008; accepted October 31st, 2008

ABSTRACT

Survivin gene may be a good target for cancer gene therapy because it is over expressed in a variety of human tumors including human hepatocellular carcinoma but not in differentiated adult tissues. To explore the effects of the siRNA of survivin gene inducing apoptosis in human hepatocellular cancer cells, three siRNAs cpusiRNA1, cpusiRNA2 and cpusiRNA3 were designed and transferred into human hepatocellular carcinoma cell line HepG2 (HepG2) by lipofection. MTT test showed that the growth of HepG2 decreased when it was transfected with 25nM, 50nM, 100nM, 150nM, 200nM, 400nM siRNA respectively after 48 hours. And the change of mRNA and protein of survivin gene and p53 gene had been detected by RT-PCR and Western blot. Cells presented an increase in apoptosis index was assayed by flow cytometry. Small interfering RNA can exert a knockdown of survivin gene expression and up regulation of p53 gene to induce apoptosis and to inhibit the growth of HepG2.

Keywords: RNAi, Survivin gene, p53 gene, Apoptosis, HepG2

1. INTRODUCTION

Hepatocellular carcinoma (HCC) was the most pernicious cancer with a high mortality rate. Despite the improvement of surgical techniques and other conditions, the prognosis of HCC remained poor. Therefore, there is a need for development of new therapy way to improve the survival rate for potential use in HCC. Many studies have shown that the survivin gene was a new member of inhibitors of the apoptosis protein (IAP) family, which had been implicated in both control of cell division and inhibition of apoptosis. It was selectively overexpressed in the majority of human tumor types including liver, lung, breast, colon, hepatocellular, oesophageal, pancreatic, bladder, uterine and ovarian cancers, largecell non-Hodgkin's lymphomas, leukaemias, neuroblastoma, brain tumors, pheochromocytoma, soft tissue sarcomas, melanomas and other skin cancer, but not in adult differentiated tissues with the exception of thymus and genital gland, and was associated with the aggressiveness of diseases and unfavorable outcomes[1,2,3,4,5]. Ling X et al reported that induction of survivin by taxol in MCF-7 cells was an early event and was independent of taxol-mediated G2/M arrest, thus suggesting a role for survivin in taxol resistance not only during mitosis but outside of the mitotic checkpoint as well[6]. On the basis of these findings survivin had been proposed as an attractive target for new anticancer interventions. Several preclinical studies had demonstrated that downregulation of survivin expression, accomplished through the use of small interfering RNA to increase the apoptotic rate and to reduce tumor growth potential. It could specifically and efficiently degrade mRNA, resulting in post uctranscriptional gene silencing (PTGS) [7,8,9,10,11,12,13,14, 15], which was a natural mechanism in organisms underlying the resistance to virus invasion and inhibition of transposon mobility. Its blocking action on gene expression had been successfully observed in rat and human cells cultured in vitro, and the knockdown of genes in cells had been achieved [11,15].A study[16] had shown that 21−25 nt small interference RNA (siRNA) could mediate specific gene silencing in mammal cells. Being effective and highly specific, RNAi probably becomed a new technique in knocking gene down and plays an important role in gene therapy of diseases. Three small interfering RNAs were designed according to the sequence of survivin gene and they were transferred into human hepatocellular carcinoma cell line HepG2 by lipofection to observe survivin and p53 gene expression changes and their effects on cell apoptosis and growth,

which laid a foundation for further studies on the functions of survivin gene and genetic therapy involved in HCC.

2. MATERIALS AND METHODS

2.1. Main Reagents

Trizol reagent and M-MLV were purchased from Gibco BRL (Carlsbad, CA). Taq DNA RNA-Mate kit was bought from Genepharma (Shanghai, China). Polyclonal rabbit anti-human polymerase, dNTPs and DNA Marker were obtained from Takara (Dalian City, China). survivin, p53 and β-Actin antibody were purchased from WuHan Boster Co. (WuHan City, China).

2.2. Small Interfering RNA Design

CpusiRNA1 sense: 5'-ACUGCAGAGAAAGAGCC dTdT-3', cpusiRNA1 anti-sense: 5'-GGCUCUUUCUCU GUCCAGUTT-3'; cpusiRNA2 sense: 5'-GAAUUUGAG GAAACUGCGAdTdT-3', cpusiRNA2 anti-sense: 5'-UC GCAGUUUCCUAACUUACTT-3'; cpusiRNA3 sense: 5'-AGCAUUCGUCCGGUUGCGCdTdT-3', cpusiRNA3 anti-sense: 5'-GCGCAACCGGACGAAUGCUTT-3'; Negative control cpusiRNA4 sense: 5'-UUCUCCGAAC GUGUCACGUdTdT-3', Negative control cpusiRNA4 anti-sense: 5'-ACGUGACACGUUCGGAGAATT-3'. Each sense chain of siRNA oligonucleotides was added two thymidine residues (dTdT) at the 3' ends. After these siRNAs having been designed, they were compared with sequence in the human EST (expressed sequence tag) database to confirm that no other genes were targeted, and then they were sent to Shanghai Genepharma Co., Ltd (Shanghai city in China) to be synthesized.

2.3. Primers of Survivin Gene Design

After primers of survivin gene (forward primer P1: 5'CGACGTTGCCCCCTGCCTG3', reverse primer P2: 5'AAGGAAAGCGCAACCGGACGA3'), p53 gene (forward primer P3: 5'AGCGATGGTCTGGCCCCTCC3', reverse primer P4: 5' GCGCCGGTCTCTCCAGGA3') and GAPDH gene (forward primer P5: 5'-ATTCAACGG CA-CAGTCAAGG-3', forward primer P6: 5'-GCAGAA GGGGCGGAGATGA-3') having been designed, they were sent to Shanghai Sangon Biological Co. (China) to be synthesized.

2.4. Cell Culture

Human hepatocellular carcinoma cell line HepG2 was maintained in our laboratory. And it was grown in RPMI-1640 supplemented with 100 mL/L fetal bovine serum (FBS) and incubated in a humidified incubator containing 50mL/L CO_2 at 37℃.

2.5. Transfection

Twenty-four hours before transfection, cells were trpsinized, diluted in fresh media without antibiotics and transferred to 96-well and 6-well plates, These cells grown to a confluency of 50%–60% were transfected with 0, 25 nmol/L, 50 nmol/L, 100 nmol/L, 200 nmol/L and 400 nmol/L of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 per well using RNAi-Mate media according to manufacturer's recommendations.

2.6. MTT Assay

HepG2 cell (5×10^3) was placed into every well of 96-well plate with 180μL culture medium RPMI-1640 containing 10% NBS. After the cell having been incubated at 37℃ for 24 hours, 20μL siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 with increasing concentrations 0, 25 nmol/L, 50 nmol/L, 100 nmol/L, 200 nmol/L and 400 nM was added to every well respectively. When the cell had been incubated at 37℃ for 44h, cell proliferation was assessed by MTT, After the cell having been incubated for 4 h, the reaction was stopped by the addition of 150 μL DMSO, After 30 min incubation and shaking, the absorbency of the samples was determined at 490 nm(A_{490}).

2.7. RT-PCR

Total RNA was extracted from HepG2 cell using Trizol reagent (Gibco BRL) according to the manufacturer's instructions. Complementary DNA (cDNA) was generated from total RNA using M-MLV. PCR of cDNA was performed in a final volume of 50 μL containing 4μL of 4×dNTPs, 2 units of Taq DNA polymerase, and 20 mmol/L of each primer. The samples were amplified 35 cycles at 94 for 30 s, at 58 for 30 s, and at 72 for 50 s, and finally at 72 for 10 min. Amplification of human GAPDH served as a control for a sample loading and integrity. the length of amplified fragments of survivin gene P53 and GAPDH gene were 250 bp, 306 bp and 213 bp.

2.8. Western Blot

The cells were washed twice with cold PBS and were lysed in radioimmunoprecitation assay buffer [50 mmol/L of Tris (pH 7.4), 150 mmol/L NaCl, 1% Triton X-100, 1% deoxycholic phenylmethyl-sulfonyl fluoride, 1μg/ml of aprotinin, and 1 mmol/L of DTT] for 10min and scraped. The extracts were centrifuged at 4000 g at 4℃ for 15min. Protein (100 μg) were resolved by sodium dodecyl sulfatepolyacry-lamide gel electrophoresis (SDS-PAGE) and transferred to nitrocellulose membranes (Sigma Co., Ltd in USA). After blocking with 5% skim milk in Tris-HCl (pH 7.5) at room temperature for 2h, the nitrocellulose membranes were reacted for 2h with specific antibodies in the same blocking solution. After extensive washing with Tris-HCl containing 0.05% Tween 20, the membrances were reacted with rabbit anti-human polyclonal antibody for Survivin, p53 and β-Actin protein detection.

2.9. Flow Cytometric Analysis

Cell cycle distributions were determined by measuring the cellular DNA content using flow cytometry (BD Biosciences Clontech, Palo Alto, CA). Cells were washed with PBS twice, fixed with 700 mL/L ethanol for 20 min and stored at 4℃ overnight, then washed with PBS twice, and stained with 100 μL of 50 mg/L PI

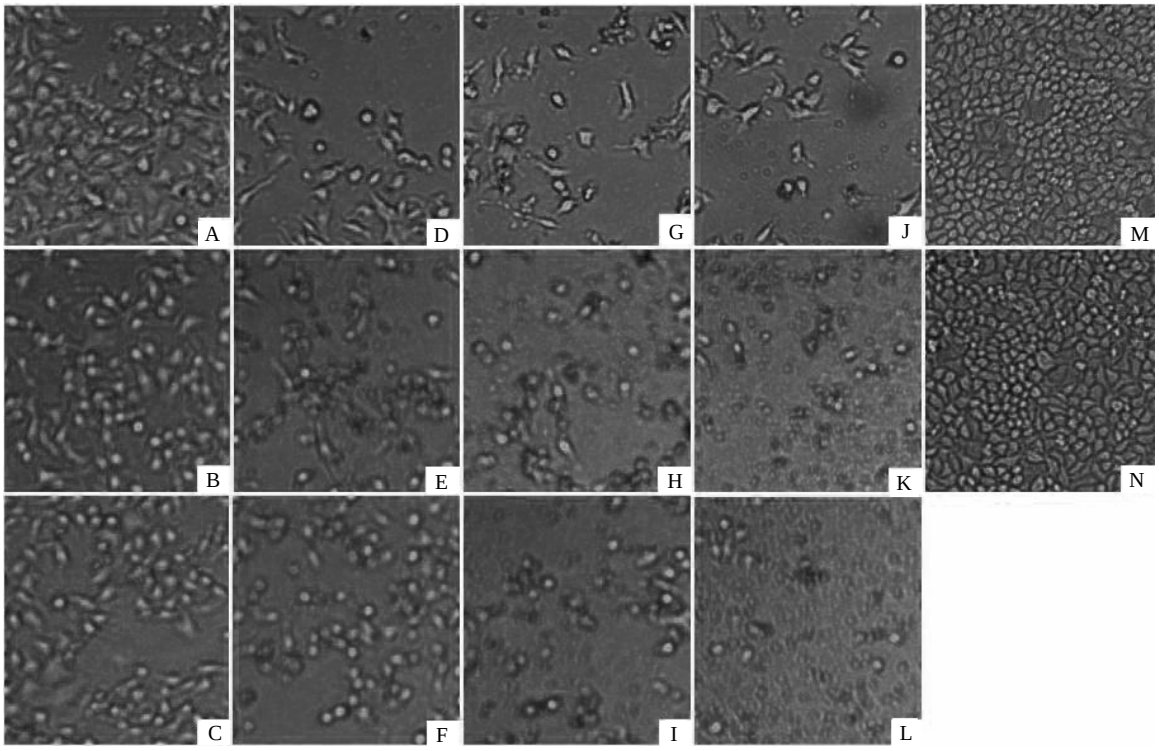


Figure 1. HepG2 cells transfected with siRNA. **A-C** HepG2 cells transfected with 50 nM nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 after 48 hours. **D-F** HepG2 cells transfected with 100 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 after 48 hours. **G-I** HepG2 cells transfected with 200 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 after 48 hours. **J-L** HepG2 cells transfected with 400 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 after 48 hours. **M** HepG2 cells transfected with 200 nM of siRNA cpusiRNA4 after 48 hours. **N** Untransfected HepG2 cells.

at 4℃ for 30min. Apoptotic cells were assayed using the Elite ESP flow cytometry at 488nm, and data were analyzed with the WinMID2.9 software.

2.8. Statistical Analysis

Data were expressed as mean ± SD. Statistical significance was determined by the Students' t-test. $P<0.05$ was considered statistically significant.

3. RESULTS

3.1. Change of Cellular Morphology

After the cells were transfected with 0, 50nM, 100nM, 200nM and 400nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 respectively for 48 h, The cellular morphology of HepG2 cells had changed greatly, such as the volume of cell changing small, the cellular morphology becoming irregularity, cell shrinking, and nucleus pycnosis. But the cellular morphology of control cells was normal (**Figure 1**).

3.2. Cellular Proliferation

The number of the cells became small when the cells were tranfected with 0, 25nM, 50nM, 100nM, 200nM and 400nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 respectively after 48 hours, and the inhibition effect to the HepG2 fade up in evidence. But the negative control cpusiRNA4 have no evident effect on the growth of HepG2 cells. The inhibition rate was determined by MTT assay as **Table 1**. And tumor curative Cisplatin was used as positive control to inhibit the growth of HepG2 cells 10μM, 20μM, 40μM, 80μM and 160μM. The inhibition rate was determined by MTT assay as **Table 1** and **2**. The IC_{50} of the cpusiRNA2 and cpusiRNA3 to the HepG2 cells is 88.1nM, 89.2nm, while the IC_{50} of Cisplatin is 32.1μM which is about 365 times higher than that of cpusiRNA2 and cpusiRNA3.

Table 1. Effects of siRNA and Cisplatin on the growth of HepG2 cells determined by MTT assay (n =6), (mean ± SD)

siRNA	0	25nM	50nM	100nM	200nM
cpusiRNA1(%)	0	17.2±2.6	36.6±4.2	50.2±3.7	66.3±3.8
cpusiRNA2(%)	0	16.8±2.5	33.9±4.2	54.5±3.7	75.3±5.2
cpusiRNA3(%)	0	10.5±3.4	20.3±3.9	41.7±3.6	61.5±4.2
Cisplatin	0	10μM	20μM	40μM	80μM
Cisplatin(%)	0	16.1±3.8	42.4±4.3	83.1±4.1	92.7±3.9

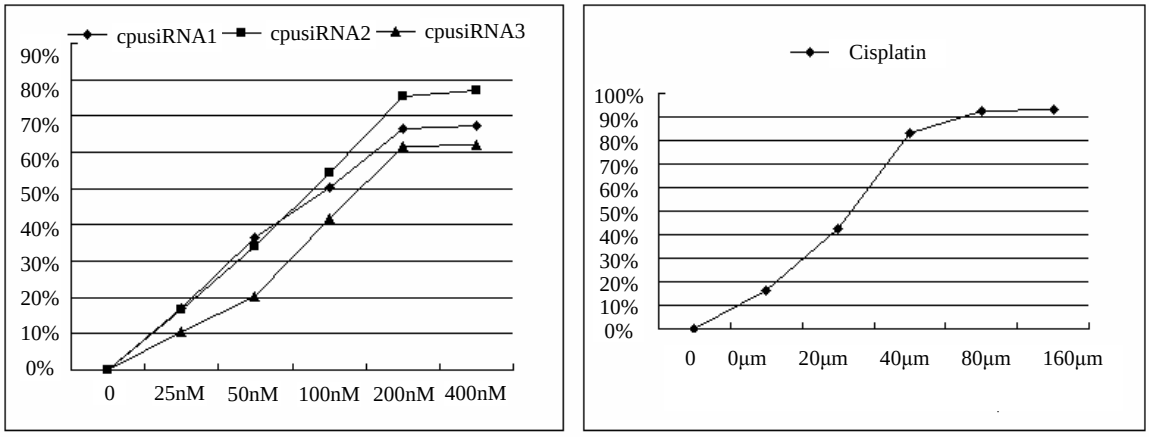


Table 2. The curve of the effects of siRNA and Cisplatin on the growth of HepG2 cells

3.3. Change of the Survivin and p53 mRNA

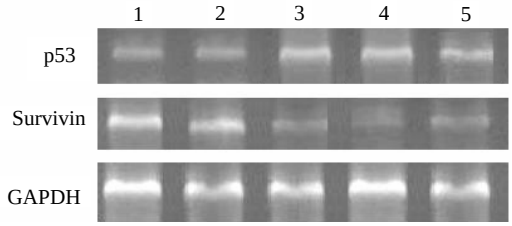


Figure 2. RT-PCR show the change of survivin and p53 mRMA. 1 HepG2 cells transfected with 200 nM of siRNA cpusiRNA4 after 48 hours; 2 Untransfected HepG2 cells; 3. 4. 5. HepG2 cells transfected with 100 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 after 48 hours

Four-eight hours after the transfection of survivin siRNA, the semi-quantitative reverse transcription polymerase chain reaction amplified 250 bp fragment of survivin gene, 306 bp fragment of p53 gene and 213 bp fragment of GAPDH gene as **Figure 2**. The result suggested not only that the mRNA of survivin gene declined, but also that the mRNA of p53 gene increased when the cells were transfected with 100 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 at 48 h in HepG2 cells (**Figure 2**).

3.4. Change of the Survivin and p53 Protein

After four-eight hours treatment of siRNA, the Western blot showed that the amount of survivin protein reduced with suppression of the survivin mRNA, and that the expression of p53 gene up-regulated because of increasing of p53 mRNA when HepG2 cells were transfected with 100 nM of siRNA after 48 hours as **Figure 3**.

3.5. Flow Cytometry Analysis

After the HepG2 cells were transfected with 100nM of siRNA cpusiRNA4, cpusiRNA1, cpusiRNA2 and

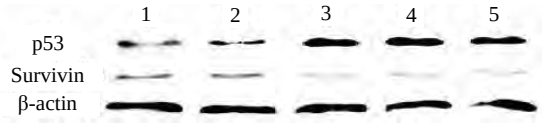


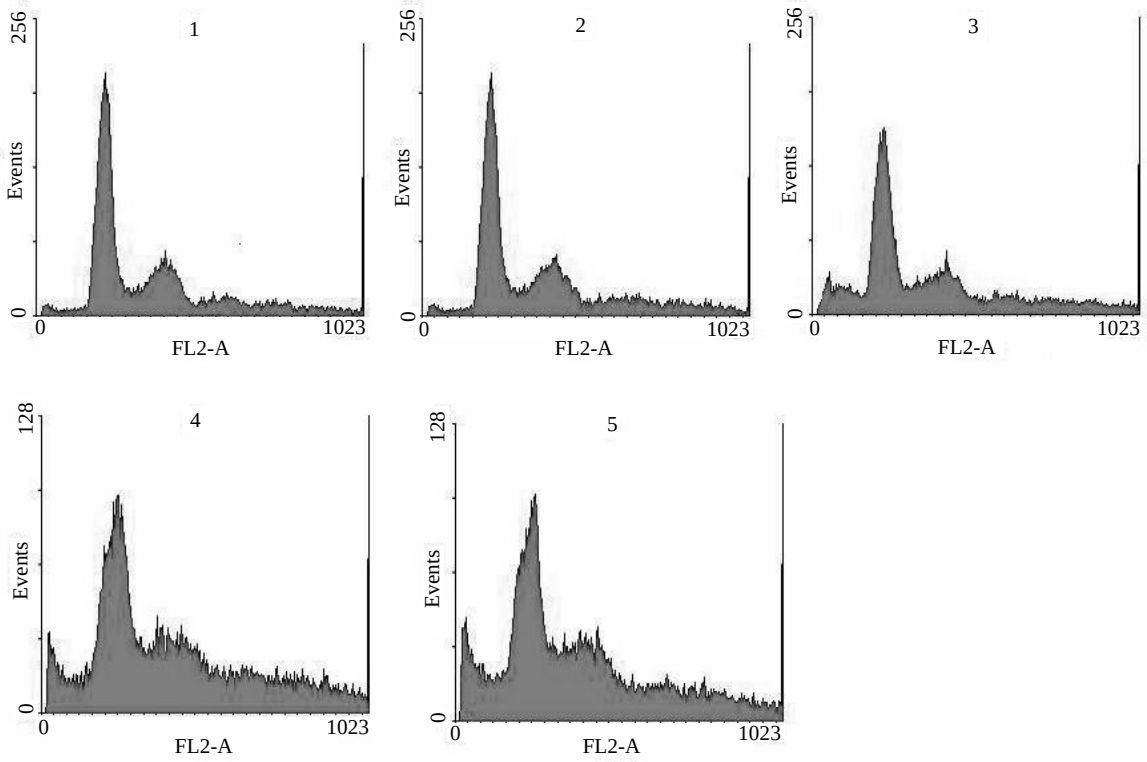
Figure 3. Western blot showed the change of survivin and p53 protein. 1. HepG2 cells transfected with 100 nM of siRNA cpusiRNA4 after 48 hours; 2. Untransfected HepG2 cells; 3. 4. 5. HepG2 cells transfected with 100 nM of siRNA cpusiRNA1 cpusiRNA2 and cpusiRNA3 after 48 hours

cpusiRNA3 respectively, flow cytometry analysis of their cell cycle displayed that the cells transfected with 100nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 respectively with a marked reduction in G2/M phase by 8.3%, 16.2% and 21.5%, the apoptosis index was as high as 13.4%, 26.6% and 32.1%. And the control groups transfected with 200nM of siRNA cpusiRNA4 is almost the same as the untransfected HepG2 (**Figure 4**).

4. DISCUSSION

Tumorigenesis is a multiple factor course. The high expressed inhibited-apoptosis genes in tumor cells can suppress apoptosis and make tumor cells avoid to be cleared and identified by the immunity system. Moreover, the inhibited apoptosis genes can also play a role during the course of drug resistance of tumor cells. Survivin provides a new research direction to cancer therapy for itscharacters of high conservation and only expression in tumor cells. Small interfering RNA technology possesses a high ability to specifically silence particular genes. Therefore, it can be used as a powerful tool in researches on the functions of genes and genetic therapy for carcinoma. Attention has been paid to RNAi in the field of researches on gene functions.

In our study, the result of MTT assay demonstrated that the chemically synthesized siRNA cpusiRNA1,



Figuer 4. Result of cell cycle detected by flow cytometry. 1 HepG2 cells transfected with 200 nM of siRNA cpusiRNA4 after 48 hours; 2 Untransfected HepG2 cells; 3 HepG2 cells transfected with 200 nM of siRNA cpusiRNA1 after 48 hours; 4 HepG2 cells transfected with 200 nM of siRNA cpusiRNA2 after 48 hours; 5 HepG2 cells transfected with 200 nM of siRNA cpusiRNA3 after 48 hours

cpusiRNA2 and cpusiRNA3 can effectively inhibit the-growth of the HepG2 cells. Their IC_{50} to the HepG2 cells is about 1/187, 1/232 and 1/127 less than that of Cis-platin. The result of RT-PCR demonstrated not only that survivin gene was exerted a knockdown, but also that the mRNA of p53 gene increased at the level of transcription when the cells were transfected with 100 nM of siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 at 48 h in HepG2 cells. The result of Western blot demonstrated that survivin gene was exerted a knockdown and p53 gene was up-regulation in HepG2 cells at the level of protein. Flow cytometry analysis revealed that the HepG2 cells transfected with the siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 presented an obvious AP peak, a marked reduction in G2/M phase and an increase in apoptosis index compared to the control groups. We also observed that HepG2 cells transfected with the siRNA cpusiRNA1, cpusiRNA2 and cpusiRNA3 grew slowly, as compared with the control groups. The above mentioned findings confirm that chemically synthesized siRNAs can specifically block survivin gene expression, induce cell apoptosis, and inhibit the growth of carcinoma cells.

The results of our study not only confirm that the inhibitory activity of survivin gene on the growth of human hepatocellular carcinoma cell line HepG2 could be realized by inducing cell apoptosis, but also confirm that the survivin gene interfering on the growth of human hepatocellular carcinoma cell line HepG2 could up-regulate p53 gene expression to inhibit the growth of HepG2 cells. Besides, RNAi alone could block survivin gene expression to induce a remarkable increase in cell apoptosis. This unique effect of survivin provides new evidence for its antiapoptotic effects on HepG2. In summary, survivin gene can be regarded as a very good target gene in genetic therapy for carcinomas. RNAi of survivin gene is a promising approach in treating carcinomas.

REFERENCES

[1] H. Caldas, L. E. Honsey and R. A. Altura. (2005) Survivin a novel Survivin splice variant expressed in malignancies. Mol Cancer, 4(1): 11−23.

[2] W. Salz, D. Eisenberg and J. Plescia. (2005) A survivin gene signature predicts aggressive tumor behavior. Cancer Res, 65: 3531−3534.

[3] T. Liu and D. Grossman. (2004) Rapid induction of mitochondria events and caspase-independent apoptosis in Survivin-targeted melanoma cells. Oncogene, 23 (1): 39−48.

[4] M. Stevenson. (2004) Therapeutic potential of RNA interference. N Engl J Med, 351 (17): 1772−1777.

[5] D. C. Altieri; (2003) Survivin versatile modulation of cell division and apoptosis in cancer. Oncogene, 22: 8581–8589.

[6] X. Ling, R. J. Bernacki, M. G. Brattain and F. Li. (2004) Induction of survivin expression by taxol (paclitaxel) is an early event, which is independent of taxol-mediated G2/M arrest. J Biol Chem;

279: (5)196−203.
[7] C. D. Novina and P. A. Sharp. (2004) The RNAi revolution. Nature, 430(6996): 161−164.
[8] P. D. Zamore. (2002) Ancient pathways p rogrammed by small RNAs. Science, 296 (5571): 1265−1269.
[9] G. J. Hannon. (2002) RNA interference. Nature, 418: 2442−2511.
[10] M. Pennati, M. Binda, M. De Cesare, G. Pratesi, M. Folini, L. Citti, M. G. Daidone, F. Zunino and N. Zaffaroni. (2004) Ribozyme-mediated down-regulation of survivin expression sensitizes human melanoma cells to topotecan in vitro and in vivo. Carcinogenesis; 25: 1129−1136.
[11] M. Izquierdo. (2005) Short interfering RNAs as a tool for cancer gene therapy. Cancer Gene Ther, 12: 217−227.
[12] M. Kappler, M. Bache, F. Bartel, M. Kotzsch, M. Panian, P. Wurl, K. Blumke, H. Schmidt, A. Meye and H. Taubert. (2004) Knockdown of survivin expression by small interfering RNA reduces the clonogenic survival of human sarcoma cell lines independently of p53. Cancer Gene Ther, 11: 186−193.
[13] M .Kappler, H. Taubert, F. Bartel, K. Blumke, M. Panian, H. Schmidt, J. Dunst and M. Bache. (2005) Radiosensitization, after a combinedtreatment of survivin siRNA and irradiation, is correlated with the activation of caspases 3 and 7 in a wtp53 sarcoma cell line, but not in a mt-p53 sarcoma cell line. Oncol Rep, 13: 167−172.
[14] M. Chawla-Sarkar, S. I. Bae, F. J. Reu, B. S. Jacobs, D. J. Lindner and E. C. Borden. (2004) Downregulation of Bcl-2, FLIP or IAPs (XIAP and survivin) by siRNAs sensitizes resistant melanoma cells to Apo2L/TRAIL-induced apoptosis. Cell Death Differ, 11: 915−923.
[15] S. Coma, V. Noe, C. Lavarino, J. Adan, M. Rivas, M. Lopez-Matas, R. Pagan, F. Mitjans, S. Vilaro, J. Piulats and C. J. Ciudad; (2004) Use of siRNAs and antisense oligonucleo -tides against survivin RNA to inhibit steps leading to tumor angiogenesis. Oligonucleotides, 14: 100−113.
[16] S. M. Elbashir, W. Lendeckel and T. Tuschl. (2002) RNA interference is mediated by 21-and 22-nucleotide RNAs. Genes Dev, 15: 188−200.
[17] Corresponding author; Prof. Xi Tao, Research Centre of Biotechnology, China Pharmaceutical University, NanJing, 210009, China.

J. Biomedical Science and Engineering, 2009, 2, 63-69

The pattern of co-existed posttranslational modifications-A case study

Zheng-Rong Yang

School of Biosciences, University of Exeter, Correspondence should be addressed to Zheng- Rong Yang (Z.R.Yang@exeter.ac.uk)

Received August 20th, 2008; revised December 1st, 2008; accepted December 5th, 2008

ABSTRACT

Posttranslational modifications are a class of important cellular activities in various biochemical processes including signalling transduction, gene/metabolite networks, and disease development. It has been found that multiple posttranslational modifications with the same or different modification residues can co-exist in the same protein and this co-occurrence is critical to signalling networks in cells. Although some biological studies have spotted this phenomenon, little bioinformatics study has been carried out for understanding its mechanism. Four data sets were downloaded from NCBI for the study. The joint probabilities of any two neighbouring posttranslational modification sites of different modification residues were analyzed. The Bayesian probabilistic network was derived for visualizing the relationship between a target modification and the contributing modifications as the predictive factors.

Keywords: posttranslational modification, bioinformatics, pattern analysis, posttranslational modification, pattern analysis, probability model

1. INTRODUCTION

Posttranslational modifications (PTMs) are a chemical process of modifying a protein's chemical or structural properties after the translation of the protein has been completed. Posttranslational modifications are closely related with signalling networks of molecules in cells. The modifications include attaching a chemical, chemical structural change of amino acids, or protein structural change. The modification will also alter the functions of a protein in both ways, adding functions or removing functions, for instance, phosphorylation and dephosphorylation, carboxylation and decarboxylation. Because of these changes, proteins will carry different *signals* for functioning in cells. Posttranslational modifications are then the focus of many signalling transduction network studies. For instance, properly folded and posttranslational modified endoplasmic reticulum is related with stress and will lead to different pathological states [1]. Poly (ADP-ribosyl)ation is related with DNA repair and cell cycle checkpoint pathways as the unique signal for protein function modulations [2]. Posttranslational modifications are the mediators for the transporters for multiple functions of human copper-transporting ATPases [3]. In the study of various chronic diseases, it is found that protein 3-nitrotyrosine (nitration) plays an important role in pathological conditions [4]. Together mutations and aberrant mRNA splicing, hyperphosphorylation will lead to a number of neurodegenerative disorders [5]. S-Nitrosation has been recently found to have similar function as phosphorylation and acetylation because of its association with various pathological cell reactions in signalling networks [6]. In cell-cycle control, differentiation, metabolism, stress response and programmed cell-death, the FOXO subgroup of forkhead transcription factors have been found being tightly controlled by phosphorylation, acetylation and ubiquitination [7].

In biological experiments, it has been found that the co-occurrence of posttranslational modifications is critical for many cellular functions in recent a few years. For instance, it has been found that the necessary condition for stable transcriptional activity of p53 is the cooperation of multiple posttranslational modifications such as phosphorylation, acetylation, and ubiquitination [8]. In studying the complex pattern of posttranslational modifications and its impact on cellular processes, it has been found that lysine acetylation, arginine/lysine methylation and serine/threonine phosphorylation will work together cooperatively for regulating the high mobility group proteins [9]. In the experiments with human cancer specimens, it has been found that the extent of acetylation, formylation and methylation is higher in cultured cells [10]. It has also been found that proteins with multiple posttranslational modifications may make contribution to similar signalling functions [11]. In studying DNA repair, apoptosis and senescence, it has been found that the interplay between multiple protein modifications, including phosphorylation, ubiquitylation, acetylation and sumoylation is critical for properly propagating DNA damage signals [12] and the interplay between methylation and acetylation has been found important for activating p53 by responding to DNA damage signals

[13]. It is even found that there is crosstalk between different posttranslational modifications [14]. In glycogen syntheses kinase-3, it has been found that O-GlcNAcylation O-phosphate is interplaying for cellular regulation [15]. The interplay has been also found in the steroid receptor coactivators [16]. In the study of transcriptional programming, it is found that interplay between posttranslational modifications exists in H3 termini [17]. In histone, it has been found that there are multiple arginine posttranslational modifications which are critical for some disease development. Also in histone, the interplay between sumoylation and either acetylation or ubiquitylation has been observed contributing to complex functions of proteins [18]. A recent study has used laboratory method to identify co-occurrence posttranslational modifications [19]. A computational method was proposed to predict the interplay between phosphorylation sites and O-GlcNAc sites based on peptides around modification residues [20,21]. The analysis was based on the prediction results from various PTM prediction tools and was based on peptide information only. Moreover, the method only focused on the competition mechanism between phosphorylation sites and O-GlcNAc modifications at the same residues.

The patterns of co-occurrence of posttranslational modifications are so far unclear or have not yet emerged through large scale studies. Bioinformatics study will help revealing those patterns and will benefit many desirable cellular engineering processes, i.e. disease control and prevention based on handling signalling pathways subjectively. This study is aimed to analyse the patterns of co-occurrence of multiple posttranslational modifications and visualizing their relationship through a probabilistic analysis.

Computational approaches, such as structural bioinformatics [22], molecular docking [23], molecular packing [24,25], pharmacophore modelling [26], Mote Carlo simulated annealing approach [27], diffusion-controlled reaction simulation [28], bio-macromolecular internal collective motion simulation [29], QSAR [30,31], protein subcellular location prediction [32,33], identification of membrane proteins and their types [34], identification of enzymes and their functional classes [35], identification of GPCR and their types [36], identification of proteases and their types [37], protein cleavage site prediction [38,39,40], and signal peptide prediction [41,42], can timely provide very useful information and insights for both basic research and drug design and are widely welcome by science community. The present study is devoted to develop a computational approach for studying co-occurrence of posttranslational modifications.

2. APPROACH

The first thing we need to do is to scan all the sequences in a data set to find all neighbouring modifications. We use frequency as the joint probability to measure the quantitative property of co-occurrence of modifications first. Through analyzing the frequency of two modifications, the likelihood that a pair of modifications occurs can be quantified.

However, the frequency analysis only shows how likely two modifications can occur simultaneously in the same sequence. For instance, we may observe that the probability for hydroxyproline and hydroxylysine to occur simultaneously in the same sequence is 18.6%. But this does not indicate how likely hydroxylysine depends on hydroxyproline. In other words, if we have observed a hydroxyproline in a sequence, how likely can we find a hydroxylysine in the same sequence as the neighbouring modification? We first define the joint probability of two different modification residues as the frequency defined as

$$P(X,Y)=\frac{\text{the number that } X \text{ and } Y \text{ occur simutaneously}}{\text{the number of neighbouring modifications}} \tag{1}$$

below Here, X and Y are two different modification residues. We then define the marginal probability as the frequency for one modification residue to occur in a data set

$$P(X)=\frac{\text{the number of } X \text{ modification}}{\text{the number of all modifications}} \tag{2}$$

Using the product theory in probability, we have Here

$$P(X,Y)=P(X\mid Y)P(Y)=P(Y\mid X)P(X) \tag{3}$$

$P(X\mid Y)$ reads out as the conditional probability for X to occur given that Y has happened. Based on the above calculation, we will have two conditional probabilities, either the probability of observing X if Y has been observed or the probability of observing Y if X has been observed. Based on the conditional probabilities and the marginal probabilities, we can use the Bayes rule to determine the posterior probabilities which are commonly used for decision-making. The Bayes rule is defined as

$$P(X_i\mid Y)=\frac{P(Y\mid X_i)P(X_i)}{\sum_m P(Y\mid X_m)P(X_m)} \tag{4}$$

Here we treat Y as a target modification, for instance a hyrdoxyproline residue. X_i is the ith potential contributing modification for the target modification Y. The posterior probability indicates if Y has been observed, what is the probability that it has a neighbour X_i. It quantifies quantitatively how possible X_i is most probable neighbouring modification of Y. Note that $\sum_i P(X_i\mid Y)=1$. The posterior probabilities are then used to draw networks to illustrate the relationship between a target modification residue and other modification residues. This probabilistic network here is a simple

type of Bayesian networks [43].

3. DATA AND EXPERIMENTAL DESIGN

Two classes of posttranslational modifications are used for the study, *i.e.* hydroxylation and methylation. Both have two most common modification residues. The hydroxylation mainly functions at a lysine residue or a proline residue while methylation mainly functions at a lysine residue or an arginine residue. Both have ample experimentally verified data for the study. Four keywords, hydroxyproline, hydroxylysine, methyllysine and methylarginine were used to scan NCBI database to download sequences. All the identical sequences were removed from the study. All three types of phosphorylations were grouped together named as phosphorylation. All the amidation activities were also grouped into one type of modification. Various acetylation modifications are grouped together. Because there are only two poly (methylaminopropyl) lysine sites, they are treated as methyllysine. Two methy-hydroxylysine residues are treated separated as one methyllysine and one hydroxylysine.

Table 1 shows the statistics of these four data sets. There are 10, 17, 10, and 8 different modification residues in the hydroxylysine, hydroxyproline, methylarginine, and methyllysine data sets, respectively. Here modification residue means a specific posttranslational modification activity at residues in proteins, for instance a hydroxyproline residue means a proline which can be hydroxylated and has been confirmed in experiments. The percentages of sites per sequence are 15.3, 7.1, 7.3, and 5.8 for the hydroxylysine, hydroxyproline, methylarginine, and methyllysine data sets, respectively. The hydroxyproline data set has the double number of modification residues compared with other three data sets. The details of multiple modifications are listed in **Table 2**. The abbreviations are seen in **Table 3**.

Table 4 shows the distribution of neighbouring PTMs of different modification residues. It can be seen that at least 25% (and up to 35%) of neighbouring PTMs are of different modification residues.

Table 1. The statistics of four data sets

Data sets	Sequences	Sites	PTM types	Sites per seq
Hydroxylysine	38	581	10	15.3
Hydroxyproline	199	1405	17	7.1
Methylarginine	23	167	10	7.3
Methyllysine	65	376	8	5.8

Table 2. The details of modifications in four data sets

Hydroxylysine	Hydroxyproline	Methylarginine	Methyllysine
hydroxyproline	hydroxyproline	phosphorylation	phosphorylation
hydroxylysine	hydroxylysine	methylarginine	acetylation
Amidation	hydroxyphenylalanine	methylhistidine	citrulline
Allysine	allysine	methylglutamine	methyllysine
phosphorylation	amidation	thioglycine	methylarginine
methyllysine	acetylation	acetylation	amidation
bromotryptophan	carboxylation	methyllysine	methylhistidine
hydroxyphenylalanine	hydroxyasparagine	citrulline	methylalanine
hydroxyarginine	bromotryptophan	methylcysteine	
proteolytic	methylcysteine	deamidation	
	phosphorylation		
	methyllysine		
	sulfoation		
	didehydrobutyrine		
	chlorotryptophan		
	didehydroalanine		
	hydroxyvaline		

Table 3. The abbreviations

AC	acetylation	AK	allysine
AM	amidation	BR	bromotryptophan
CA	carboxylation	CH	chlorotryptophan
CT	citrullination	DA	didehydroalanine
DM	deamidation	DP	D-phenylalanine
DT	D-tryptophan	DY	didehydrobutyrine
HK	hydroxylysine	HP	hydroxyproline
HR	hydroxyarginine	HN	hydroxyasparagine
HV	hydroxyvaline	HF	hydroxyphenylalanine
MA	methylalanine	MC	methylcysteine
MQ	methylglutamine	MH	methylhistidine
MK	methyllysine	MR	methylarginine
NT	nitration	PH	phosphorylation
PR	proteolytic	SU	sulfoation
TH	thioglycine		

Among these neighbouring PTMs of different modification residues, 33%, 68%, 78%, 84% have the distance less than 10 residues for the methylarginine, the methyllysine, the hydroxylysine, and the hydroxyproline data sets, respectively as seen in **Table 5**. Because of this, two PTMs of different modification residues may likely share similar structure (at least an overlapped local structure) for binding. This suggests that the cooperative activities of PTMs of different modification residues are critical to cellular signalling/functioning.

Based on the collected sequences, we produce a program to search for all the posttranslational modification residues in four data sets. The sites must have the notation as </site_type="modified">, </experiment="experimental evidence, …">, and </note_type=X>. Here X can be various types, for instance, 4-hydroxyproline, 3-hydroxyproline, 5-hydroxyproline, etc. For each of four posttranslational modifications, we find all the involved posttranslational modification sites. The sequences in different data sets are analyzed separately although there are some overlaps among them.

4. RESULTS

Table 6 shows the frequencies as the joint probabilities of nine types of modifications for the hydroxylysine data set. Proteolytic is removed as there is only one such site in the data set. The highest joint probability is 60.2% for two hydroxyproline residues to be neighbours. However, the joint probability for two hydroxylysine residues to be neighbours is only 6.81%. The co-occurrence probability for these two types of hydroxylation is 18.8%. These two probabilities indicate that for every hydroxylysine residue, the probability for it to have a hydroxyproline as the neighbour is three times higher than the probability for it to have the same hydroxylysine residue as the neighbour. The joint probabilities for a hydroxylysine to have an amidation, allysine, phosphorylation residue, bromotryptophan, hydroxyphenylalanine, and hydroxyarginine residue as the neighbour are 0.55%, 0.55%, 0.18%, 0.74%, and 0.18%, respectively. This means that except for the same type of modification,

Table 4. The distribution of neighbouring PTMs of different modification residues

	sites	Percentage over total
Hydroxylysine	147	25%
Hydroxyproline	434	31%
Methylarginine	40	24%
Methyllysine	132	35%

Table 5. The times for different modifications to occur in one sequence

	1-2	3-4	5-6	7-8	9-10	>11	Sum
Hydroxylysine	32	32	33	3	14	33	147
Hydroxyproline	175	97	46	18	29	69	434
Methylarginine	3	2	2	2	4	27	40
Methyllysine	48	24	5	11	2	42	132

Table 6. The joint probability as frequency of co-occurred modifications for the hydroxylysine data set

	HP	HK	AM	AK	PH	MK	BR	HF	HR
HP	60.2	18.8	0	0	0	0	0.18	0	0
HK	18.8	6.81	0.55	0.55	0.18	0	0.18	0.74	0.18
AM	0	0.55	0	0	0	0	0	0	0
AK	0	0.55	0	0	0	0	0	0	0
PH	0	0.18	0	0	4.05	5.52	0	0	0
MK	0	0	0	0	5.52	1.47	0	0	0
BR	0.18	0.18	0	0	0	0	0	0	0.18
HF	0	0.74	0	0	0	0	0	0.37	0
HR	0	0.18	0	0	0	0	0.18	0	0

hydroxylysine has a high correlation with allysine, amidation, and hydroxyphenylalanine modifications.

Figure 1 visualises the probabilistic relationship among different types of modifications in the hydroxylysine data set using the posterior probabilities, where all the posterior probabilities less than 10% are omitted for simplicity. The network demonstrates that hydroxylysine only depends on hydroxyproline (24%). However, hydroxylysine has great impacts on five modification residues, allysine (100%), amidation (100%), hydroxyphenylalanine (67%), hydroxyarginine (50%), and bromotryptophan (33%). Phosphorylation and methyllysine modification residues are independent from the hydroxylysine block. They are mutually correlated to each other.

Table 7 shows the joint probabilities for the hydroxyproline data set. It can be seen that the probability for two hydroxyproline residues to be neighbours is 57.7%. The likelihood for a hydroxyproline to have a hydroxylysine

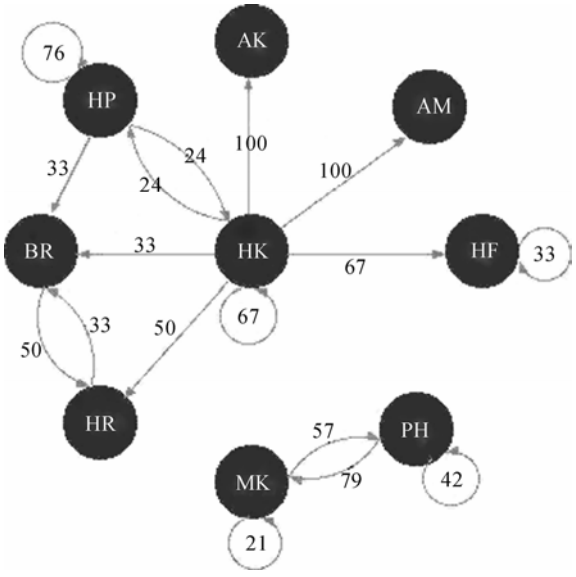


Figure 1. The probabilistic network as the relationship among modification residues in the hydroxylysine data set. The values represent the posterior probabilities. The arcs mean the directions. For instance, the arc from HK to AM with a number 100 means P(HK|AM)=100%. In other words, the probability of observing a hydroxylysine residue nearby an observed amidation residue is almost certain

as the neighbour is 8.46%. However, the likelihood for a hydroxyproline to have a hydroxyphenylalanine as the neighbour is 18.2%. This means that a hydroxyproline is more likely to have a hydroxyphenylalanine to co-occur rather than hydroxylysine. The other two important modifications for hydroxyproline are carboxylation and amidation. The probability for a hydroxyproline residue and an amidation residue to be neighbours is 4.06% and the probability for a hydroxyproline residue and a carboxylation residue to be neighbours is 1.91%. The other co-occurred modifications with joint probabilities larger than 0.2% are acetylation and bromotryptophan.

The probabilistic relationship among different modifications shown in **Figure 2** is built for the hydroxyproline data set using the posterior probabilities. All the posterior probabilities less than 10% are not shown. In the probabilistic network, it can be seen that the most contributing modifications for hydroxyproline is hydroxyphenylalanine (20%). However, hydroxyproline has contributed to 11 other modification residues. For instance, the posterior probability P(HP|AC) is 100% meaning that whenever we have found an acetylation residue, it is certain that there is a hydroxyproline residue nearby. The posterior probability P(HP|HP) is 63% while P(HK|HP)=9%, P(AK|HP)=1%, P(AM|HP)=4%, P(AC|HP)=1%, P(CA|HP)=2%, P(BR|HP)=1%.

Table 7. The joint probability (larger than 0.2%) as frequency of co-occurred modifications for the hydroxyproline data set

	HP	HK	HF	AM	AC	CA	BR
HP	57.7	8.46	18.2	4.06	0.41	1.91	0.5
HK	8.46	1.58	0	0.08	0	0	0.08
HF	18.2	0	0.25	0	0	0	0
AM	4.06	0.08	0	0.17	0	0.17	0.41
AC	0.41	0	0	0	0	0	0
CA	1.91	0	0	0.17	0	2.24	0.33
BR	0.5	0.08	0	0.41	0	0.33	0

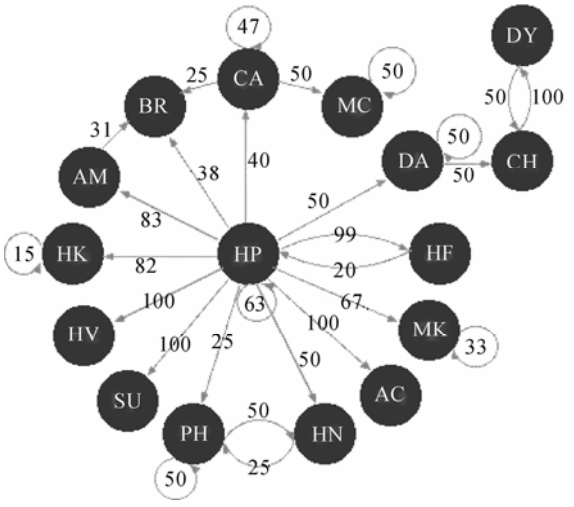


Figure 2. The probabilistic network as the relationship among modifications in the hydroxyproline data set. The values represent the posterior probabilities

Table 8 shows the frequencies of modifications in the methylarginine data set. The likelihood for two methylarginine residues to co-occur as neighbours is 40.3%. The interesting phenomenon is that the co-occurrence probability between methylarginine and methyllysine residues (as neighbours) is 0%. The contributing modifications to methylargine residues are phosphorylation (11.1%), acetylation (2.78%), methylhistidine (1.39%), and citrulline (1.39%). This is not as expected as it is thought that both two dominant methylation modifications should be highly correlated.

Shown in **Figure 3** is the probabilistic relationship among the modifications derived from the methylarginine data set. The network shows that the most contributing modifications to methylarginine modification is phosphorylation (19%), i.e. P(PH|MR)=19%. For any observed methylarginine residue in a protein, the probability of observing a phosphorylation residue is 19%. Meanwhile, methylarginine residue can be an important pre-request for other modification residues. For instance, the probability of observing a methylarginine residue if a methylhistidine residue has been observed is 100% and the probability that a methylarginine residue is standing near an observed acetylation residue is 33%.100%.

Table 8. The joint probability as frequency of co-occurred modifications for the methylarginine data set

	PH	MR	MH	MG	TH	AC	MK	CT
PH	26.4	11.1	0	0	0	2.78	1.39	0.69
MR	11.1	40.3	1.39	0.69	0.69	2.78	0	1.39
MH	0	1.39	0	0	0	0	0	0
MG	0	0.69	0	0	0.69	0	0	0
TH	0	0.69	0	0.69	0	0	0	0
AC	2.78	2.78	0	0	0	2.08	0.69	0
MK	1.39	0	0	0	0	0.69	0.69	0
CT	0.69	1.39	0	0	0	0	0	2.78

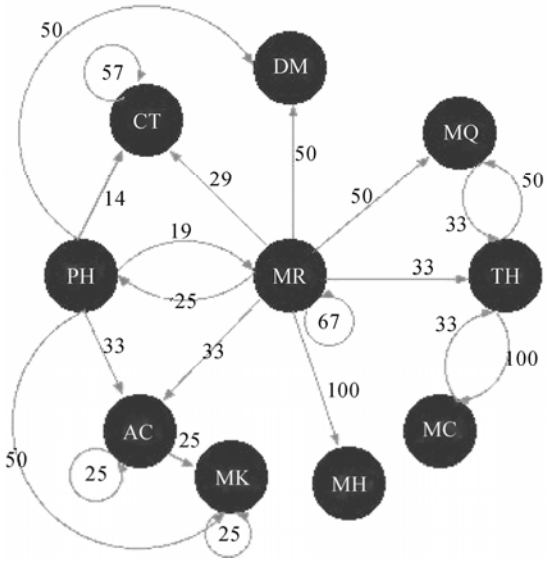


Figure 3. The probabilistic network as the relationship among modifications in the methylarginine data set. The values represent the posterior probabilities

Table 9 shows the frequencies for the methyllysine data set. The probability for two methyllysine residues to be neighbours is 28.3%, which is not dominantly high. Interestingly, we have found the contributing modifications for methyllysine are acetylation (15.11%) and phosphorylation (10.61%). It is again difficult to find the evidence that two methylation modifications are highly correlated.

Figure 4 shows the probabilistic network as the relationship among the modifications in the methyllysine data set. Here, the most contributing modifications to methyllysine are phosphorylation (22%) and acetylation (31%). The methyllysine residue has a high correlation with methylarginine, methyhistine, and methylalanine. All have the posterior probabilities as 100%.

5. CONCLUSION

This paper has studied the co-occurrence pattern of two types of posttranslational modifications with four modification residues. The study aims to reveal how posttranslational modifications are correlated to each other, i.e. how one posttranslational modification contributes to the others. It has been found that the hydroxylysine and hydroxyproline residues are not the most mutually dependent modification residues, so are the methylarginine and methyllysine residues. We have found that the hydroxyllysine residues depend on the hydroxyproline residues with a posterior probability 24% and the hydroxyproline residues are the unique major contributing modification residue for the hydroxylysine residues. However, we have found the hydroxyproline residues nearly do not depend on the hydroxylysine residues. Instead, the hydroxyphenylalanine residues are the contributing modification residue to the the hydroxyproline residues with a posterior probability 20%. Among the methylarginine residues and the methyllysine residues, we have found that the phosphorylation residues are the main player for both of these two modification residues. In addition, the acetylation residues are needed for the methyllysine residues as well. Surprisingly, two different methylation residues also do not rely on each other. Although the study is limited to two modification classes with four modification residues, it is expected that this method can be generalized to a wide range of multiple posttranslational modification pattern discovery.

Table 9. The joint probability as frequency of co-occurred modifications for the methyllysine data set

	PH	AC	CT	MK	MR	AM	MH	MA
PH	8.04	11.58	0	10.61	0	0.32	0	0
AC	11.58	28.3	0.96	15.11	1.93	0	0	0
CT	0	0.96	0	0.96	0	0	0	0
MK	10.61	15.11	0.96	21.22	0	0	0.32	0.64
MR	0	1.93	0	0	0	0	0	0
AM	0.32	0	0	0	0	0	0	0
MH	0	0	0	0.32	0	0	0	0
MA	0	0	0	0.64	0	0	0	0

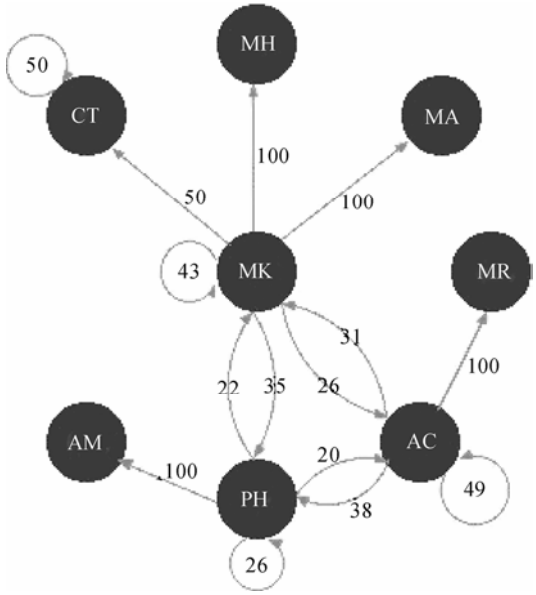


Figure 4. The probabilistic network as the relationship among modifications in the methyllysine data set. The values represent the posterior probabilities

REFERENCES

[1] E. Lai, T. Teodoro, and A. Volchuk, (2007) Endoplasmic reticulum stress: signaling the unfolded protein response. Physiology (Bethesda), 22, 193−201.

[2] J. F. Haince, S. Kozlov, V. L. Dawson, T. M. Dawson, M. J. Hendzel, M.F. Lavin, and G. G. Poirier, (2007) Ataxia telangiectasia mutated (ATM) signaling network is modulated by a novel poly (ADP-ribose)-dependent pathway in the early response to DNA- damaging agents. J Biol Chem., 282, 16441−16453.

[3] S. Lutsenko, A. Gupta, , J. L. Burkhead, and V. Zuzel, (2008) Cellular multitasking: The dual role of human Cu-ATPases in cofactor delivery and intracellular copper balance. Arch Biochem Biophys., (in press).

[4] J. M. Souza, G. Peluffo, and R. Radi, (2008) Protein tyrosine nitration-Functional alteration or just a biomarker? Free Radic Biol Med., (in press).

[5] J. Z. Wang, and F. Liu, (2008) Microtubule-associated protein tau in development, degeneration and protection of neurons. Prog Neurobiol., 85, 148−175.

[6] V. V. Sumbayev and I. M. Yasinska, (2008) Protein S-nitrosation in signal transduction: assays for specific qualitative and quantitative analysis. Methods Enzymol, 440, 209−219.

[7] T. Obsil, and V. Obsilova, (2008) Structure/function relationships underlying regulation of FOXO transcription factors. Oncogene, 27, 2263−2275.

[8] A. Scoumanne and X. Chen, (2008) Protein methylation: a new mechanism of p53 tumor suppressor regulation. Histol Histopathol., 23, 1143−1149.

[9] Q. Zhang, and Y. Wang, (2008) High mobility group proteins and their post-translational modifications. Biochim Biophys Acta., (in press).

[10] J. R. Wisniewski, A. Zougman, S. Kruger, P. Ziołkowski, M. Pudełko, M. Bębenek and M. Mann, (2008) Constitutive and dynamic phosphorylation and acetylation sites on NUCKS, a hypermodified nuclear protein, studied by quantitative proteomics. Proteins, (in press).

[11] C. A. Galea, A. Nourse, Y. Wang, S. G. Sivakolundu, W. T. Heller and R. W. Kriwacki, (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. J Mol Biol. 376, 827−838.

[12] M. S. Huen and J. Chen, (2008) The DNA damage response pathways: at the crossroad of protein modifications. Cell Res., 18, 8−16.

[13] M. S. Huen and J. Chen, (2008) The DNA damage response pathways: at the crossroad of protein modifications. Cell Res.,

18, 8−16.

[14] T. Hunter, (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol Cell., 28, 730−738.

[15] Z. Wang, A. Pandey and G. W. Hart, (2007) Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. Mol Cell Proteomics, 6, 1365−1379.

[16] S. Li and Y. Shang (2007) Regulation of SRC family coactivators by post-translational modifications. Cell Signal, 19, 1101− 1112.

[17] P. R. Thompson and W. Fast, (2006) Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? ACS Chem Biol., 21, 433−41.

[18] D. Nathan, K. Ingvarsdottir, D. E. Sterner, G. R. Bylebyl, M. Dokmanovic, J. A. Dorsey, K. A. Whelan, M. Krsmanovic, W. S. Lane, P. B. Meluh, E. S. Johnson, and S. L.Berger, (2006) Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. Genes Dev., 20, 966−976.

[19] J. Seo, J. Jeong, Y. M. Kim, N. Hwang, E. Paek and K. J. Lee, (2008) Strategy for comprehensive identification of post-translational modifications in cellular proteins, including low abundant modifications: application to glyceraldehyde-3-phosphate dehydrogenase. J Proteome Res., 7, 587−602.

[20] A. Ahmad, T. S. Khan, D. C. Hoessli, E. Walker-Nasir, A. Kaleem, A. R. Shakoori and S. Nasir-ud-Din, (2008) In silico modulation of HMGN-1 binding to histones and gene expression by interplay of phosphorylation and O-GlcNAc modification. Protein Pept Lett., 15, 193−199.

[21] A. Kaleem, D. C. Hoessli, I. Ahmad, E. Walker-Nasir, A. Nasim, A. R. Shakoori and S. Nasir-ud-Din, (2008) Immediate-early gene regulation by interplay between different post-translational modifications on human histone H3. J Cell Biochem., 103, 835−851.

[22] K. C. Chou, (2004) Review: Structural bioinformatics and its impact to biomedical science. Current Medicinal Chemistry, 11, 2105−2134.

[23] K. C. Chou, D. Q. Wei and W. Z. Zhong (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, Vol.310, 675). Biochem Biophys Res Comm, 308, 148−151.

[24] K. C. Chou, G. Nemethy and H. A. Scheraga, (1984) Energetic approach to packing of a-helices: 2. General treatment of nonequivalent and nonregular helices. Journal of American Chemical Society, 106, 3161−3170.

[25] K. C. Chou, G. M. Maggiora, G. Nemethy and H. A. Scheraga, (1988) Energetics of the structure of the four-alpha-helix bundle in proteins. Proceedings of National Academy of Sciences, USA, 85**,** 4295−4299.

[26] S. Sirois, D. Q. Wei, Q. S. Du and K. C. Chou, (2004) Virtual-Screening for SARS-CoV Protease Based on KZ7088 Pharmacophore Points. J Chem Inf Comput Sci, 44, 1111−1122.

[27] K. C. Chou, (1992) Energy-optimized structure of antifreeze protein and its binding mechanism. Journal of Molecular Biology, 223, 509−517.

[28] K. C. Chou and G. P. Zhou, (1982) Role of the protein outside active site on the diffusion-controlled reaction of enzyme. Journal of American Chemical Society, 104, 1409−1413.

[29] K. C. Chou, (1988) Review: Low-frequency collective motion in biomacromolecules and its biological functions. Biophysical Chemistry, 30, 3−48.

[30] Q. S. Du, R. B. Huang, Y. T. Wei, L. Q. Du and K. C. Chou, (2008) Multiple Field Three Dimensional Quantitative Structure- Activity Relationship (MF-3D-QSAR). Journal of Computational Chemistry, 29, 211−219.

[31] H. Gonzalez-Díaz, Y. Gonzalez-Díaz, L. Santana, F. M. Ubeira and E. Uriarte, (2008) Proteomics, networks, and connectivity indices. Proteomics, 8, 750–778.

[32] K. C. Chou and H. B. Shen, (2007) Review: Recent progresses in protein subcellular location prediction. Analytical Biochemistry, 370, 1−16.

[33] K. C. Chou and H. B. Shen, (2008) Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms. Nature Protocols, 3, 153−162.

[34] K. C. Chou and H. B. Shen, (2007) MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Comm, 360, 339−345.

[35] H. B. Shen, and K. C. Chou, (2007) EzyPred: A top-down approach for predicting enzyme functional classes and subclasses. Biochem Biophys Res Comm, 364, 53−59.

[36] K. C. Chou, (2005) Prediction of G-protein-coupled receptor classes. Journal of Proteome Research, 4, 1413−1418.

[37] K. C. Chou and H. B. Shen, (2008) ProtIdent: A web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochem Biophys Res Comm, 376, 321−325.

[38] K. C. Chou, (1993) A vectorized sequence-coupling model for predicting HIV protease cleavage sites in proteins. J Biol Chem, 268, 16938−16948.

[39] K. C. Chou, (1996) Review: Prediction of HIV protease cleavage sites in proteins. Analytical Biochemistry, 233, 1−14.

[40] H. B. Shen and K. C. Chou, (2008) HIVcleave: a web-server for predicting HIV protease cleavage sites in proteins. Analytical Biochemistry, 375, 388−390.

[41] K. C. Chou and H. B. Shen, (2007) Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. Biochem Biophys Res Comm, 357, 633−640.

[42] H. B. Shen and K. C. Chou, (2007) Signal-3L: a 3-layer approach for predicting signal peptide. Biochem Biophys Res Comm, 363, 297−303.

[43] P. Pourret and B. Naim, (2008) Marcot, Bayesian Networks: A Practical Guide to Applications. Chichester, UK: Wiley.

J. Biomedical Science and Engineering, 2009, 2, 70-75

Computer-Assisted analysis of subcellular localization signals and post-translational modifications of human prion proteins

Fatemeh Moosawi, Hassan Mohabatkar*

Department of Biology, College of Sciences, Shiraz University, Shiraz, Iran, *Corresponding author to Fatemeh Moosawi (mohabat@shirazu.ac.ir), Telephone: +98711-6137426, Mobile: 09177157459, Fax: +98711-2280926.

Received August 26th, 2008; revised December 19th, 2008; accepted December 7th, 2008

ABSTRACT

In the present work, computational analyses were applied to study the subcellular localization and posttranslational modifications of human prion proteins (PrPs). The tentative location of prion protein was determined to be in the nucleolus inside the nucleus by the following bioinformatics tools: Hum-PLoc, Euk-PLoc and Nuc-PLoc. Based on our results signal peptides with average of 22 base pairs in N-terminal were identified in human PrPs. This theoretical study demonstrates that PrP is post-translationally modified by: 1) attachment of two N-linked complex carbohydrate moieties (N181 and N197), 2) attachmet of glycosylphosphatidylinositol (GPI) at serine 230 and 3) formation of two disulfide bonds between "6–22" and "179–214" cysteines. Furthermore, ten protein kinase phosphorylation sites were predicted in human PrP. The above-noted phosphorylation was carried out by PKC and CK2. By using bioinformatics tools, we have shown that computationally human PrPs locate particularly into the nucleolus.

Keywords: Prion protein; Subcellular localization; Signal peptides; Post-translational Modifications; Bioinformatics

1. INTRODUCTION

Prion diseases or transmissible spongiform encephalopathies (TSE) are fatal neurodegenerative conditions in humans and animals that originate spontaneously, genetically or by infection [1]. Human TSE diseases include sporadic, genetic, iatrogenic and variant Creutzfeldt-Jakob disease (CJD) and sporadic or familial fatal insomnia. Animal counterparts are scrapie in sheep and goats, bovine spongiform encephalopathy, and chronic wasting disease of mule deer and elk [2,3]. The critical pathogenetic events in TSE diseases are conformational changes of the physiological host prion protein (PrPc) into an insoluble form (PrPsc) [4].

Prions are devoid of nucleic acid and seem to be composed exclusively of protein. The normal, cellular PrP is converted into PrPsc through a process whereby a portion of its α-helical and coil structure is refolded into β-sheet [5,6]. This structural transition is accompanied by profound changes in the physicochemical properties of the PrP [7]. At the molecular level, PrP is a sialoglycoprotein of 253 amino acids in human [8]. The C-terminal end is a signal peptide allowing the anchoring of the protein to the plasma membrane via a glycosylphosphatidylinositol residue in the early steps of maturation in the endoplasmic reticulum [9]. The ultimate destination is then the plasma membrane where PrP can be released by phospholipase or protease treatment [10]. Finally, PrP cycles between the plasma membrane and the endocytic pathway. During infection, PrPsc is thought to derive from PrPc after exposure to the plasma membrane [11]. It has been suggested that PrPc can bind a putative ''protein X'' that may function as a molecular chaperone in the formation of PrPsc [12].

The cellular role of the normal host protein PrP is still unknown. Oesch has characterized membrane associated proteins that interact with PrP [13,14]. Immunocytochemical studies reveal that PrPsc and PrPc are present, not only at the plasma membrane or in cytoplasmic compartments [15,16,17], but also in the nuclear compartment, particularly into the nucleolus [18]. The inconsistency in the nuclear localization of PrP between non-infected and infected cells may account for important pathogenetical mechanisms [19].

The posttranslational modification of amino acids expands the range of functions of the protein by attaching it to other biochemical functional groups such as acetate, phosphate lipids and carbohydrates. This occurs by altering the chemical nature of an amino acid or by making structural changes, like the formation of disulfide bridges. The molecular mechanism by which the PrPsc is formed and causes infectivity or neurodegeneration is not known. In an emerging view, post-translational modifications play roles in the transformation of PrPc to PrPsc [20]. Moreover, Post-translational modification of the scrapie prion protein is thought to account for the unusual fea-

tures of this protein [21].

All eukaryotic cells are compartmentalized into separate membrane-bound organelles and require tightly regulated transport of proteins and lipids between these compartments. The function of a protein is closely correlated with its subcellular location. With the rapid increase in new protein sequences entering into data banks, we are confronted with a challenge. Proteins are classified, according to their subcellular locations, into the following 18 groups: cell wall, centriole, chloroplast, cyanelle, cytoplasm, cytoskeleton, endoplasmic reticulum, extracell, Golgi apparatus, hydrogenosome, lysosome, mitochondria, nucleus, peroxisome, plasma membrane, plastid, spindle pole body, and vacuole [22].

Determination of subcellular location of a protein is essential for understanding its biochemical function. These data are hard to obtain experimentally but have become especially significant since many protein sequences are still lacking detailed functional information. To address this rarity of data, many computational analysis methods have been developed. However, these methods have varying levels of accuracy and perform differently based on the sequences that are presented to the underlying algorithm. Giving the huge number of uncharacterized protein sequences, computer-aided analysis of posttranslational modifications and translocation signals from amino acid sequence becomes a necessity.

In this study, we have analyzed subcellular localization, signals peptides and posttranslational modifications of human PrPs.

2. MATERIALS AND METHODS

2.1. Amino Acids Sequence

The sequences of human PrPs were obtained from http://www.ncbi.nlm.nih.gov/sites/Entrez and http//beta.uniprot.org. Accession numbers of human PrPs are shown in **Table 1**.

2.2. Consensus Sequence and Percentage of Different Amino Acids

The consensus sequence was achieved by using Multalin 5.4.1 server and the percentage of different amino acids was calculated by expasy server.

Table 1. Accession numbers of human PrPs

AAD46098	AAG21693	AAO83636	AAC05365
AAC78725	AAV38303	AAB59442	AAB59443
AAA60182	AAO83635	A2A2V1	AAR21603
AAH12844	AAH22532	AAS80162	AAA19664
A1YVW6	ABM85428	ABD63004	ABM82244
ABL75508	BAA00011	CAG46836	CAD62016
CAB75503	CAM27320	CAA56283	CAI19053
CAI19053	CAA58442	CAG46869	EAX10449
EAX10450	NP_000302	NP_001073592	NP_001073591
NP_898902	NP_001073590	O75942	P04156
P23907	Q6FGR8	Q5QPB4	Q53YK7
Q6FGN5	Q540C4	Q6SES1	Q5U0K3
Q86XR1			

2.3. Prediction of Signal Peptides

Signal-CF server was employed to study the signal peptides. The web interface to the Signal-CF tool was accessible at http://www.chou.med.harvard.edu/shen. This server is called Signal-CF, where C stands for "coupling" and F for "fusion", meaning that Signal-CF is formed by incorporating the subsite coupling effects along a protein sequence and by fusing the results derived from many width-different scaled windows through a voting system. Signal-CF is featured by high success prediction rates with short computational time, and hence is particularly useful for the analysis of large-scale datasets [23].

2.4. Prediction of Subcellular Localization

Several computational tools for predicting the subcellular localization of a protein are available. In this study, Hum-PLoc, Euk-PLoc and Nuc-PLoc have been utilized to study the localization of prion protein sequences.

Hum-PLoc is a server that analyzes the subcellular localization of human proteins among the following 12 locations: centriole, cytoplasm, cytoskeleton, endoplasmic reticulum, extracell, Golgi apparatus, lysosome, microsome, mitochondrion, nucleus, peroxisome, and plasma membrane [24]. The web interface to this tool is present at http://www URL: http://www.chou.med.harvard.edu/shen.

Euk-PLoc is available as a web-server at http://202.120.37.186/bioinf/euk. A new benchmark dataset is constructed that covers the following 18 localizations: cell wall, centriole, chloroplast, cyanelle, cytoplasm, cytoskeleton, endoplasmic reticulum, extracell, Golgi apparatus, hydrogenosome, lysosome, mitochondria, nucleus, peroxisome, plasma membrane, plastid, spindle pole body, and vacuole [25].

A new classifier, called Nuc-PLoc, has been developed that can be exploited to recognize nuclear proteins among the following nine subnuclear locations: chromatin, heterochromatin, nuclear envelope, nuclear matrix, nuclear pore complex, nuclear speckle, nucleolus, nucleoplasm and nuclear promyelocytic leukemia (PML) body. As a user-friendly web-server, Nuc-PLoc is accessible at http://chou.med.harvard.edu/ bioinf/Nuc-PLoc [25].

2.5. Analysis of Posttranslational Modifications

N-myristoylation, N-glycosylation, protein kinase C, casein kinase II and Serine, threonine, tyrosine phosphorylation sites were predicted. Expasy which is available at www.expasy.ch/tools was applied for this purpose. Big-PI server was utilized to study the glycosylphosphatidylinositol (GPI) anchor signal [26]. The web server http://clavius.bc.edu/~clotelab/DiANNA was chosen for prediction of disulfide bonds [27].

3. RESULTS

3.1. Sequences and Signal Peptides

Number of amino acids and molecular weight of human consensus sequence of prion protein were 253 and 27661.1 respectively. The consensus sequence of human PrPs is shown in **Figure 1**.

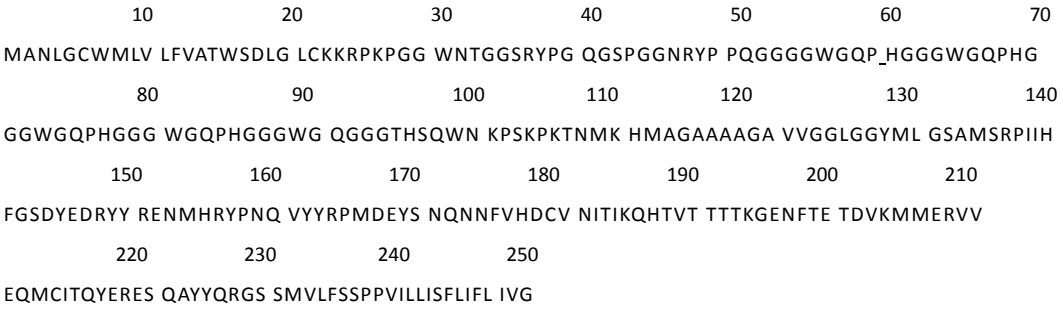


Figure 1. Consensus sequence of the human prion protein

Percentage of different amino acids in the protein was calculated (**Table 2**). The most prevalent amino acid was glycine (45 residues), and the least one was cysteine (4 residues).

To identify functional signal peptides in the human PrP, 49 FASTA format sequences of prion protein input in Signal-CF sever. Signal peptide sequences of human PrPs were sorted in 4 groups based on their length (**Table 3**).

3.2. Prediction of Subcellular Localization

The subcellular distribution of PrP proteins was verified by Euk-Ploc, Hum-Ploc and Nuc-ploc. These results showed a strong tendency of the protein to nucleus and especially to nucleolus.

3.3. Prediction of Post-translation Modifications

It is interesting that the post-translational modifications alone, or in combination with amino acid changes, play dominant roles in the pathogenic transformation of PrP(C) to PrP (SC). According to our analysis 2 asparagines in positions 181 and 197 were predicted to be glycosylated. Results also showed that threonine in positions 107, 183, 190, 191, 192, and 193 and serine in position 132 were predicted to be kinase C phosphorylated. Serine in position 143, and threonines in positions 201 and 206 were expected to be casein kinase II phosphorylated and no glycine was predicted to be myristoylated. Disulfide bridges in cysteine residues at positions 6–22 and 179–214 were predicted.

GPI anchors, which allow the attachment of proteins to the extracellular leaflet of the plasma membrane, were also analyzed. Glycosylphosphoatidylinositol (GPI) lipid anchoring is a common posttranslational modification known mainly in extracellular eukaryotic proteins. Attachment of the GPI moiety to the carboxyl terminus (omega site) of a polypeptide happens following proteolytic cleavage of a C-terminal propeptide (**Figure 2**).

The best predicted site was G229 and the second best was S230 (underlined). Furthermore, potential phosphorylation sites of serine, threonine and tyrosine in the human PrPs were determined (**Table 4**).

Table 2. Residue composition for consensus sequence of human PrPs

A	%4	10	C	%1.6	4	D	%2.4	6	E	%3.6	9	F	%2.8	7
G	%17.8	45	H	%4.0	10	I	%3.6	9	K	%4.0	10	L	%4.7	12
M	%4.7	12	N	%4.7	12	P	%6.7	17	Q	%5.9	15	R	%4.3	11
S	%5.9	15	T	%5.1	13	V	%5.5	14	W	%3.6	9	Y	%5.1	13

Table 3. Categories of signal peptides of human PrPs

Position of signal peptides	1-14	1-15	1-22	1-24
Number of signal peptides	1	8	38	2

Table 4. Predicted phosphorylation sites of consensus sequence of human PrPs

Amino acid	Phosphorylation position			
Ser	43	143	23	231
Thr	191	192	163	-
Tyr	145	149	169	225

MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQGG
GGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQGGGTHSQWNKPSKP
KTNMKHMAGAAAAGAVVGGLGGYMLGSAMSRPIIHFGSDYEDRYYRENMHRYPN
QVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQM
CITQYERESQAYYQRGS SMVLFSSPPV ILLISFLIFL IVG

Figure 2. GPI lipid anchoring signals sequence

4. DISCUSION

The goal of this investigation was to apply bioinformatics methods to study the subcellular localizations, signal peptides and posttranslational modifications of human PrPs.

4.1. Identification of Signal Peptides in PrP

Based on our results, there were signal peptides with average 22 bp in N-terminal of human PrPs. According to a survey, conducted by Alexandre and his colleagues, PrP does not contain a nuclear localization signal and that, in normal conditions, PrP cannot be released in the cytosolic compartment remaining membrane bound till its degradation and in infected cells, PrP can interact with a molecule to form a complex able to be released in the cytosol and then targeted enters the nucleus [18]. However in another investigation, the presence of two independent nuclear localization signals in the N-terminal region of PrP was observed. The first signal included residues 23–28, and the second one included residues 101–106 of PrP [29].

Protein signals have become crucial tools for researchers to construct new drugs which are expected to enter a particular organelle to correct a specific defect. For example, by adding a specific tag to a desired protein, one can tag it for excretion, making it much easier to harvest. To use such a tool successfully, first one has to identify the signal sequences. Since the number of nascent protein sequences entering databanks is rapidly increasing, it is time consuming and expensive to identify the signal peptides entirely by experiments [28].

4.2. Determinants of Subcellular localization of PrP

Protein subcellular localization prediction has been widely studied (reviewed in [30,31]). Available servers differ in many aspects including the computational method, the type and diversity of protein characteristics, the localization coverage, the target organism(s) and the reliability. Servers can be grouped into 4 general classes based upon the protein characteristics that are considered: amino acid composition and order based predictors [32,33,34], sorting signal predictors [35,36], homology based predictors [37,38] and hybrid methods that integrate several sources of information to predict localization [39,40]. Nowadays, the importance of developing a powerful high-throughput tool to predict protein subcellular location has become obvious [41].

In the present study, the tentative location of prion protein was determined to be in the nucleolus inside the nucleus by bioinformatics tools, Hum-PLoc, Euk-PLoc and Nuc-PLoc. There are different opinions regarding the subcellular localization of PrP. Stahl and his colleagues considered a signal peptide at the C-terminus of prion protein allowing the anchoring of the protein to the plasma membrane via a glycosylphosphatidylinositol residue in the early steps of maturation in the endoplasmic reticulum [8]. Oesch has characterized membrane-associated proteins that interact with PrP [12,13]. More recently, it has been shown that the 37-kDA laminin receptor interacts with PrP [42].

A number of cellular proteins, among them the nuclear lectin CBP35, was identified that bound to the predicted RNA stem-loop structure of PrP RNA. CBP35 could also be detected in purified infections prions, [43]. Moreover, the presence of PrP in the nucleus and its subnuclear location in the nucleolus has been reported [17,44].

In addition, Gu and his coworkers demonstrated that nuclear accumulation of PrP fragments was mediated by nuclear localization signals in the N-terminal domain of PrP that became functional under certain conditions and might contribute to the pathogenesis of certain prion disorders [29].

4.3. Post-translational Modifications of Consensus Sequence of PrP

Our analysis shows that PrP is post-translationally modified by the attachment of two N-linked complex carbohydrate moieties (N181 and N197) and a GPI anchor at serine 230 as well as by the formation of a disulfide bond between 6−22 and 179−214 cysteins.

Glycosylation is one of the most complex and ubiquitous post-translational modifications of proteins in eukaryotic cells. It is a dynamic enzymatic process in which saccharides are attached to proteins or lipoproteins, usually on serine (S), threonine, asparagine, and tryptophan residues. Glycosylation, like phosphorylation, is clinically important because of its role in a wide variety of cellular, developmental and immunological processes, including protein folding, protein trafficking and localization, cell-cell interactions, and epitope recognition [45,46,47,48,49,50]. The number of glycosylation sites in our work is in agreement with the results obtained by molecular cloning of a PrP cDNA [20]. It has already been shown also that addition of one or two N-glycans causes retention of the N-terminal PrP fragment in the endoplasmic reticulum in a partially aggregated form, and a small amount is secreted into the medium. Presence of two glycans in the N-terminal fragment is more conducive to proper folding and secretion into the medium than one glycan, which largely remains in the ER [29].

In GPI anchors, a hydrophobic phosphatidylinositol group is linked to a residue at or near the C-terminus of a protein through a carbohydrate-containing linker. GPI anchor addition is both structurally and functionally related to another important post-translational modification, prenylation, in which hydrophobic farnesyl or geranylgeranyl moieties are added to C-terminal cysteine residues of target proteins. Additionally, GPI anchors proteins to the cell membrane [51]. Although we determined the nucleus as the tentative location for prion proteins, this fact also should be taken in mind that according to a previous study GPI anchor and the N-glycans function in a complicated way to reduce the tendency of PrP for localization in nucleus [30].

In our study, 10 protein kinases phosphorylation sites were predicted in the human PrPs. The addition of a phosphate molecule to a polar R group of an amino acid residue can turn a hydrophobic portion of a protein into a polar and extremely hydrophilic part. In this way, it can introduce a conformational change in the structure via interaction with other hydrophobic and hydrophilic residues in the protein. Moreover, phosphorylation may modulate PrP biological activity. Regarding the bonding states of cysteine also, it has been found out that it plays important functional and structural roles in proteins. Particularly, disulfide bond formation is one of the most important factors influencing the three-dimensional fold of proteins [52].

In conclusion, this study can help in better understanding of signal peptides of prion proteins. Generally, our results indicate the role that bioinformatics can play in analysis of proteins modification and localization.

ACKNOWLEDGMENT

Support of this study by Shiraz University is acknowledged.

REFERENCES

[1] G. G. Kovacs and H. Budka, (2008) Prion Diseases: From Protein to Cell Pathology. The American Journa of Pathology, 172(3): 555−565.

[2] R. G. Will, J. W. Ironside, M. Zeidler, S. N. Cousens, K. Estibiro, A. Alperovitch, S. Poser, M. Pocchiari, A. Hofman and P. G. Smitch, (1996) A new variant of Creutsfeldt-Jakob disease in the UK. Lancet, 347: 921−925.

[3] M. E. Bruce, (2006) New variant Creutsfeldt-Jakob disease and bovine spongiform encephalopathy. Nature medicine, 6: 258−259.

[4] I. Loredana, N. Beatriz, Z. Andrea, C. Franco, A. Raquel, S. Marco, B. Simona, I. Marcello, L. Quanguo, V. Vito, L. Mei, F. Franco, C. Salvatore, F. Antonio and P. Maurizio, (2006) Scrapie infectivity is quickly cleared in tissues of orally-infected farmed fish. BMC Veterinary Research, 2: 21.

[5] D. A. Lysek, C. Schorn, L. G. Nivon, V. Esteve-Moya, B. Christen, L. Calzolai, C. Von Schroetter, F. Fiorito, T. Herrmann, P. Guntert and K. (2005) Wuthrich, Prion protein NMR structure of cats, dogs, pigs and sheep. Proceedings of the National Academy of Sciences of the United States of America, 102: 640−645.

[6] K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. Huang, R. J. Fletterick and F. E. Cohen, *et al*. (1993) Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. Proceedings of National Academy of Sciences of the United States of America, 90(23): 10962−10966.

[7] B. P. Stantley, (1998) Proceedings of the National Academy of Sciences of the United States of America, 95: 13363−13383.

[8] H. A. Kretzschmar, L. E. Stowring, D. Westaway, W. H. Stubblebine, S. B. Prusiner and S. J. DeArmond, (1986) Molecular cloning of a human prion protein cDNA. DNA, 5: 315–324.

[9] N. Stahl, D. R. Borchelt, K. Hsiao and S. B. Prusiner, (1987) Scrapie prion protein contains a phosphatidylinositol glycolipid. Cell, 51: 229−240.

[10] N. Stahl, D. R. Borchelt and S. B. Prusiner, (1990) Differential release ofcellular and scrapie prion proteins fromcellular membranes by phosphatidylinositol-specific phospholipase C. Biochemistry, 29: 5405−5412.

[11] B. Caughey and G. J. Raymond, (1991) The scrapie-associated form of PrP is made from a cell surface precursor that is both protease and phospholipase-sensitive. The Journal of biological chemistry, 266: 18217–18223.

[12] G. C. Telling, M. Scott, J. Mastrianni, R. Gabizon, M. Torchia, F. E. Cohen, S. J. DeArmond and S. B. Prusiner, (1995) Prion propagation in mice expressing human and chimeric PrP transgenes implicates the interaction of cellular PrP with another protein. Cell, 83: 79−90.

[13] B. Oesch, D. B. Teplow, N. Stahl, D. Serban, L. E. Hood and S. B. Prusiner, (1990) Identification of cellular proteins binding to the scrapie prion. Biochemistry, 29: 5848–5855.

[14] B. Oesch, (1994) Characterization of PrP binding proteins. Philosophical transactions of the Royal Society of London. Series B, Biological, 343: 443–445.

[15] B. Caughey, K. Neary, R. Buller, D. Ernst, L. Perry, B. Chesebro and R. Race, (1990) Normal and scrapie-associated forms of prion protein differ in their sensitivities to phospholipase and proteases in intact neuroblastoma cells. Journal of virology, 64: 1093−1101.

[16] A. Taraboulos, M. Rogers, D. R. Borchelt, M. P. McKinley, M. Scott, D. Serban and S. B. Prusiner, (1990) Acquisition of proteaseresistance by prion proteins in scrapie-infected cells does notrequire asparagine-linked glycosylation. Proceedings of the National Academy of Sciences of the United States of America, 87: 8262−8266.

[17] M. P. McKinley, A. Taraboulos, L. Kenaga, D. Serban, A. Stieber, S. J. DeArmond, S. B. Prusiner and N. Gonatas, (1991) Ultrastructural localization of scrapie prion proteins in cytoplasmic vesicles of infected cultured cells. Laboratory investigation, 65: 622−630.

[18] K. Pfeifer, M. Bachmann, H. C. Schroder, J. Forrest and W. E. G. Muller, (1993) Kinetics of expression of prion protein in uninfected and scrapieinfected N2 (a) mouse neuroblastoma cells. Cell biochemistry and function, 11: 1−11.

[19] J. Alexandre, M. Franck, M. Jean, C. Bertrand, P. D. Jean and D. Dominique, (1998) Search for a Nuclear Localization Signal in the Prion Protein. Molecular and Cellular Neuroscience, 11: 127−133.

[20] L. Otvos and M. Cudic, (2002) Post-translational modifications in prion proteins. Current protein and peptide science, 3(6): 643−652.

[21] T. Haraguchi, S. Fisher, S. Olofsson, T. Endo, D. Groth, A. Tarentino, D. R. Borchelt, D. Teplow, L. Hood, and A. Burlingame, *et al.* (1989) Asparagine-linked glycosylation of the scrapie and cellular prion proteins. Archives of biochemistry and biophysics, 274(1): 1−13.

[22] H. B. Shen, J. Yang and K. C. Chou, (2007) Euk-PLoc: an ensemble classifier for large-scale eukaryotic protein subcellular location prediction. Amino Acids, 33(1): 57−67.

[23] K. C. Chou and H. B. Shen, (2007) Signal-CF: a subsite-coupled and window fusing approach for predicting signal peptides. Biochemical and biophysical research communications, 357(3): 633−640.

[24] K. C. Chou, and H. B. Shen, (2006) Hum-PLoc: a novel ensemble classifier for predicting human protein subcellular localization. Biochemical and biophysical research communications, 347(1): 150−157.

[25] H. B. Shen, and K. C. Chou, (2007) Nuc-PLoc: a new web-server for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM. Protein engineering, design and selection, 20(11): 561−567.

[26] B. Eisenhaber, G. Schneider, M. Wildpaner and F. Eisenhaber, (2004) A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for Aspergillus nidulans, Candida albicans, Neurospora crassa, Saccharomyces cerevisiae and Schizosaccharomyces pombe. Journal of molecular biology, 337(2**)**: 243−253.

[27] F. Ferrè, and P. Clote, (2005) Clote· DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Research, 33: 230−232.

[28] K. C. Chou, (2001) Using subsite coupling to predict signal

peptides. Protein Engineering, 2: 75−79.
[29] Y. Gu, J. Hinnerwisch, R. Fredricks, S. Kalepu, R. S. Mishra and N. Singh, (2002) Identification of cryptic nuclear localization signals in the prion protein. Neurobiology of Disease, 12: 133−149.
[30] Z. P. Feng, (2002) An overview on predicting the subcellular location of a protein. In silico biology, 2: 291−303.
[31] A. Reinhardt and T. Hubbard, (1998) Using neural networks for prediction of the subcellular location of proteins. Nucleic acids research, 26: 2230−2236.
[32] S. Hua and Z. Sun, (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics, 17: 721−728.
[33] K. C. Chou, (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins, 43: 246−255.
[34] J. D. Bendtsen, H. Nielsen, G. Heijne and S. Brunak, (2004) Improved prediction of signal peptides: SignalP 3.0. Journal of molecular biology, 340: 783−795.
[35] O. Emanuelsson, H. Nielsen, S. Brunak and G. Heijne, (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of molecular biology, 300: 1005−1016.
[36] E. M. Marcotte, I. Xenarios, A. M. Der Bliek and D. Eisenberg, (2000) Localizing proteins in the cell from their phylogenetic profiles. Proceedings of the National Academy of Sciences of the United States of America, 97: 12115−12120.
[37] Z. Lu, D. Szafron, R. Greiner, P. Lu, D. S. Wishart, B. Poulin, J. Anvik, C. Macdonell and R. Eisner, (2004) Predicting subcellular localization of proteins using machine-learned classifiers. Bioinformatics, 20: 547−556.
[38] K. Nakai and M. Kanehisa, (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics, 14: 897−911.
[39] A. Drawid and M. Gerstein, (2000) A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome. Journal of molecular biology, 301: 1059−1075.
[40] M. S. Scott, D. Y. Thomas, and M. T. Hallett, (2004) Predicting subcellular localization via protein motif co-occurrence. Genome research, 14: 1957−1966.
[41] Y. D. Cai, and K. C. Chou, (2004) Predicting subcellular localization of proteins in a hybridization space. Bioinformatics, 20(7): 1151−1156.
[42] R. Rieger, F. Edenhofer, C. I. Lasme´zas, and S. Weiss, (1997) The human 37-kDA laminin receptor precursor interacts with the prion protein in eukaryotic cells. Nature medicine, 3: 1383−1388.
[43] W. E. Müller, U. Scheffer, S. Perovic, J. Forrest and H. C. Schröder, (1997) Interaction of prion protein mRNA with CBP35 and other cellular proteins Possible implications for prion replication and age-dependent changes. Archives of gerontology and geriatrics , 25(1): 41−58.
[44] J. F. Bazan, R. J. Fletterick, M. P. McKinley and S. B. Prusiner, (1987) Pre-dicted secondary structure and membrane topology of the scrapie prion protein. Protein engineering, 2: 125−135.
[45] R. Haltiwanger and J. Lowe, (2004) role of glycosylation in development. Annual Review of Biochemistry, 73: 491−537.
[46] P. Mentesana and J. Konopka, (2001) Mutational analysis of the role of N glycosylation in alpha-factor receptor function. Biochemistry, 40(32): 9685−9694.
[47] K. Pilobello and L. Mahal, (2007) Deciphering the glycocode: the complexity and analytical challenge of glycomics. Current opinion in chemical biology, 11(3): 300−305.
[48] S. Miyamoto, (2006) Clinical applications of glycomic approaches for the detection of cancer and other diseases. Current opinion in molecular therapeutics, 8: 507−513.
[49] R. Gupta and S. Brunak, (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. Pacific Symposium on Biocomputing, 310−322.
[50] C. W. Von der Lieth, A. Bohne-Lang, K. K. Lohmann and M. Frank, (2004) Bioinformatics for glycomics: Status, methods, requirements and perspectives. Briefings in Bioinformatics, 5(2)**:** 164−178.
[51] B. Eisenhaber, P. Bork and F. Eisenhaber, (1999) Prediction of Potential GPI-modification Sites in Protein Sequences. Journal of molecular biology, 292: 741−758.
[52] S. M. Muskal, S. R. Holbrook and S. H. Kim, Prediction of the disulfide-bonding state of cysteine in proteins. Protein Engineering 1990, 3(8): 667−672.

Call for Papers

Journal of Biomedical Science and Engineering (JBiSE)

www. scirp. org/journal/jbise

JBiSE, publishes research and review articles in all important aspects of biology, medicine, engineering, and their intersection. Both experimental and theoretical papers are acceptable provided they report important findings, novel insights, or useful techniques in these areas. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed in hard copy.

Subject Coverage

Bioelectrical and neural engineering

Bioinformatics

Medical applications of computer modeling

Biomedical modeling

Biomedical image processing & visualization

Real-time health monitoring systems

Biomechanics and bio-transport

Patten recognition and medical diagnosis

Biomedical effects of electromagnetic radiation

Safety of wireless communication devices

Biomedical devices, sensors, and nano technologies

NMR/CT/ECG technologies and EM field simulation

Physiological signal processing

Medical data mining

Other related topics

Editors-in-Chief

Editorial Board

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

www. scirp. org/journal/jbise
Email: jbise@scirp.org

ISSN 1937-6871 (Print), 1937-688X (Online)

Journal of Biomedical Science and Engineering, 2009, 2.
Published Online February 2009 in SciRes (http://www.scirp.org/journal/jbise).

TABLE OF CONTENTS

Volume 2, Number 1, February 2009

977193768700104