ISSN : 0974 - 7435 Volume 8 Issue 11

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 8(11), 2013 [1478-1482]

Comparison of ncRNAs related to two diseases using z-curve method

Yang Yan-ling[1,2]

[1]Shandong Provincial Key Laboratory of Functional Macromolecular Biophysics, DeZhou University, DeZhou (CHINA)

[2]College of physics and electronic information engineering, DeZhou University, De Zhou 253023(CHINA)

E-mail: yyl000831@163.com

ABSTRACT

At the present time, it is proved that non- protein- coding RNAs(ncRNAs) play very important roles in the cellular process of translation and relates to Human Severe Diseases. From the NONCODE database 15 disease-related ncRNA sequences are selected in this paper, which play regulator roles and relate to Alzheimer Disease and Cancer. Based on Z-curve theory, we map and analyze the Z curves of the studied sequences. By comparing the curves, the statistical features of the Z-curve are obtained. We find that there are almost same curves and base content for one kind of ncRNAs playing same roles. And there are differences of Z_n-n curve in different organisms. The conclusion is that the Z-curve is connected with the type or function of ncRNA and organisms.

KEYWORDS

Non-coding RNA;
Alzheimer disease;
Cancer;
Role;
Z-curve;
Mutation.

INTRODUCTION

Non-coding regions, in other words, non-protein-coding regions, it is widely accepted now, are those regions in genome which are not coding for any proteins but play regulator roles in the celluar process[1]. It has been proved that the number of non-coding sequences in the human genome is more than 97% and even their many functions are unknown today. The same conditions occur in other organisms, including animals and plants[2]. It is proved that Non-coding sequences play important roles in the process of translation in organisms ranging from bacteria to mammals. At the present time the research on non-coding region and its function is still a hot field all over the world. Especially the study on ncRNAs is becoming more and more important and has been made great progress already[3].

Increasing evidence stands in support of the notion that some ncRNAs are closely related to many Human Severe Diseases, such as Alzheimer disease, Cancer, Aplastic anemia, Beckwith-Wiedemann Syndrome, etc. Among these diseases, Alzheimer disease has become the fourth-biggest cause of the illness threaten the old men, next below the cancers, heart diseases and cerebrovascular diseases[4].

Traditionally, most RNA molecules were thought to function as mediators carrying the information from the gene to the translational machinery. However, since the late 1990s, it has been widely acknowledged that other types of untranslated RNA molecules are present in many different organisms ranging from bacteria to mammals, and are affecting a large variety of processes

including plasmid replication, phage development, chromosome structure, DNA transcription, RNA processing and modification, development control and more. All of these functional RNAs have a common term—noncoding RNAs (ncRNAs)[5].

It is proved that some ncRNAs (such as BC200 RNA, bicRNA) are closely related to Alzheimer disease and Cancer. But the features and functions of the many recently identified ncRNAs remain mostly unknown[6].

Alzheimer disease is a progressive degenerative disorder of the brain characterized by a slow, progressive decline in cognitive function and behavior. As the disease advances, persons with Alzheimer disease have tough time with daily usage of things like using the phone, cooking, handling money, or driving the car[7]. The disease is more common in elder population. It is estimated that Alzheimer disease affects 15 million people worldwide and approximately 4 million Americans[8].

The neuropathologic hallmarks of the disorder are amyloid-rich senile plaques, neurofibrillary tangles, and neuronal degeneration. It has reported that three genes with autosomal dominant mutations have been identified that may lead to Alzheimer symptoms in carriers before they reach age 60. The pathogenesis of Alzheimer disease has been unknown up to the present time. Furthermore, the clinical features of this disease overlaps with common signs of aging, and other types of dementia, hence the diagnosis of Alzheimer disease remains quite difficult to us[9].

Cancer is a class of diseases in which a group of cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). These three malignant properties of cancers differentiate them from benign tumors, which are self-limited, do not invade or metastasize[10]. Cancer may affect people at all ages, even fetuses, but the risk for most varieties increases with age. Cancer causes about 13% of all deaths[11].

Nearly all cancers are caused by abnormalities in the genetic material of the transformed cells. These abnormalities may be due to the effects of carcinogens, such as tobacco smoke, radiation, chemicals, or infectious agents. Other cancer-promoting genetic abnormalities may be randomly acquired through errors in DNA replication, or are inherited, and thus present in all cells from birth. The heritability of cancers is usually affected by complex interactions between carcinogens and the host's genome. New aspects of the genetics of cancer pathogenesis, such as DNA methylation, and microRNAs (is ncRNAs) are increasingly recognized as important[12].

However, there are so many unsolved problems in ncRNAs field and many of these ncRNAs still have uncharacterized functions. One of the largest challenges is identifying ncRNAs and their unknown functions. In this paper, we make use the ZCURVE method to analysis disease-related ncRNAs. ZCURVE is a geometrical approach to study DNA sequences.

MATERIAL AND METHOD

Material

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the custom paper size (21 cm * 28.5 cm).

The NONCODE database is an integrated knowledge database designed for the analysis of non-coding RNAs (ncRNAs). In the updated version of NONCODE (NONCODE v2.0), the number of collected ncRNAs has reached 212 527, including a wide range of microRNAs, Piwi-interacting RNAs and mRNA-like ncRNAs[13]. All classes of reported ncRNAs are included.

In this paper, we select 15 BC200 RNA sequences from NONCODE database, which relate to Alzheimer disease and Cancer and come from different organisms. The accession numbers of the sequences in this database and sequence length are as follows:

1.n289 (122bp); 2.n637 (200bp); 3.n751 (205bp); 4.n752 (204bp); 5.n753 (198bp); 6.n754 (203bp); 7.n755 (197bp); 8.n756 (204bp); 9.n757 (200bp); 10.n758 (200bp); 11.n759 (197bp); 12.n760 (195bp); 13.n761 (196bp); 14.n4617 (126bp); 15.n4817 (200bp).

Where, the selected sequences belong to class BC200RNA but 1, 14 and 15 are named BC200 and other sequences are named BC200-alpha. Their cellular roles are regulators, but their sequence length is different and coming from different organisms, respectively.

Method

FULL PAPER

The *Z*-curve is a three-dimensional curve that provides a unique representation of a DNA sequence in that the DNA sequence and the *Z*-curve can each be uniquely reconstructed from the other. Therefore, the *Z*-curve contains all the information that the corresponding DNA sequence carries. The resulting curve has a zigzag shape, hence the name *Z*-curve[14].

Consider a DNA sequence read from the 5' to the 3'-end with N bases. Inspect the sequence one base at a time, beginning from the first base. Let the number of the inspecting steps is denoted by n, i.e., n = 1, 2... N. In the nth step, count the cumulative numbers of the bases A, C, G and T, occurring in the subsequence from the first to the nth base in the DNA sequence inspected. Denoting the cumulative occurring numbers of the bases A, C, G and T in the above subsequence by A_n, C_n, G_n and T_n, respectively. There is a relation as follows:

$$A_n + C_n + G_n + T_n = n$$

The Z curve is composed of a series of nodes P_0, P_1, P_2,…,P_N with straight lines, whose coordinates x_n ÿ y_n and z_n (n=0, 1, 2,…N, where N is the length of the DNA sequence being studied) are uniquely determined by the Z-transform of DNA sequence. Furthermore, the three independent components, and have a clear biological meaning, respectively[15].

One of the advantages of the *Z*-curve is its intuitiveness; the entire *Z*-curve of a genome can be viewed on a computer screen or on paper, regardless of genome length, thus allowing both global and local compositional features of genomes to be easily grasped. It offers an intuitive and convenient approach to study DNA sequences geometrically. For more detailed information about the Z-curve defined, please refer to references[15-17].

We make use the Z-plotter and Origin7.5 software to analysis selected ncRNAs. Based on the Z-curve theory, some global and local features of ncRNAs sequence can be detected in a perceivable way.

RESULT AND DISCUSSION

Result

Some typical curves of studied sequences are shown and compared in Figure 1 to 8.

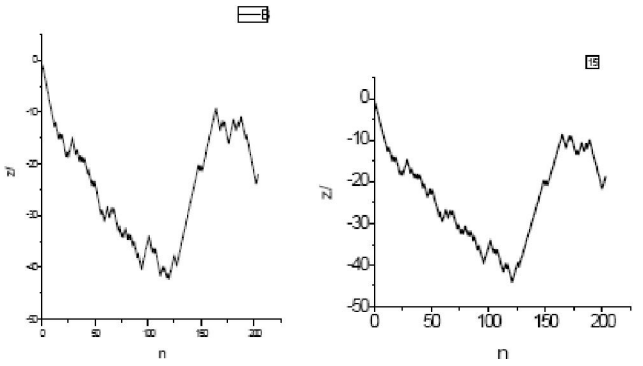

Figure 1 : The Z_n-n curves of BC200 RNA sequences 2 and 15

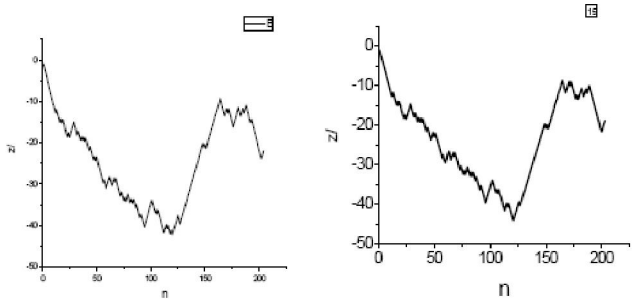

Figure 2 : The Z'_n-n curves of BC200 RNA sequences 2 and 15

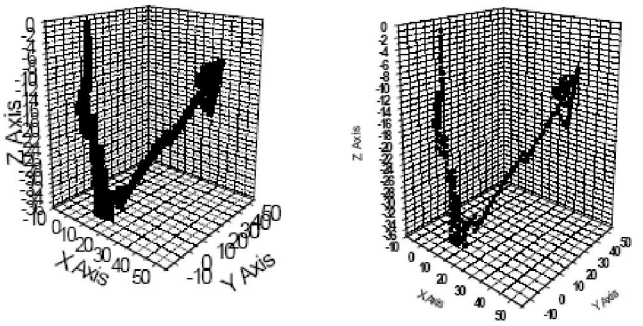

Figure 3 : The 3D curves of BC200 RNA sequences 2 and 15

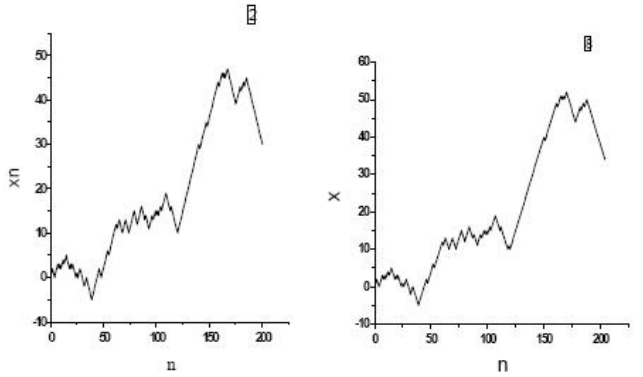

Figure 4 : The X_n-n curves of BC200-alpha sequences 2 and 8

Figure 1, 2and 3 are the z_n , z'_n and 3D curves of sequence 2 and 15, respectively. The curves are al-

most same in global and local. The sequences 2 and 15 come from human, respectively. Figure 4, 5 and 6 are the x_n, $1/2\,(x_n \pm y_n)$ and curves of sequence 2 and 8, respectively. The curves are very similar in global and local. The sequences come from human and African green monkey, respectively. Figure 7 is the y_n curves of 8 and 15. The curves are very similar in global and local. The sequences come from African green monkey and human, respectively. Figure 8 is the z_n curves of sequence 8 and 15. The curves are obviously similar in global but different in local. In the previous figure the value of z_n is lesser than zero. But in the area (from 160 to 190bp) of next figure, the value of z_n is greater than zero and at others locations z_n <0. The same conditions occur in sequence 2 and 8.

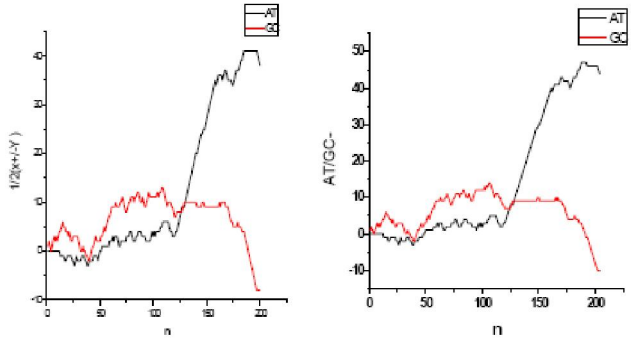

Figure 5 : The AT-and GC-curves of sequences 2 and 8

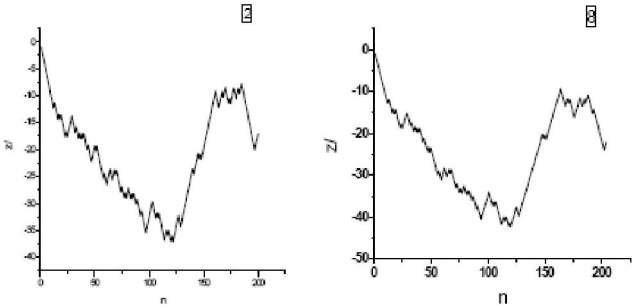

Figure 6 : The Z'_n-n curves for sequences 2 and 8

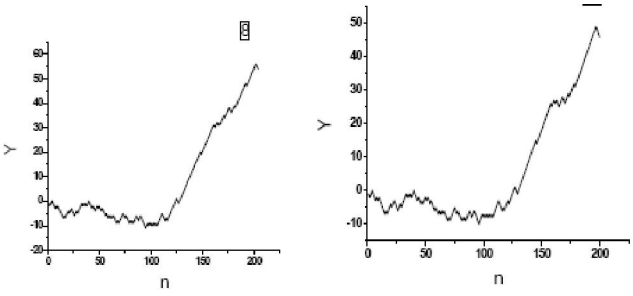

Figure 7 : The Y_n-n curves of sequences 8 and 15

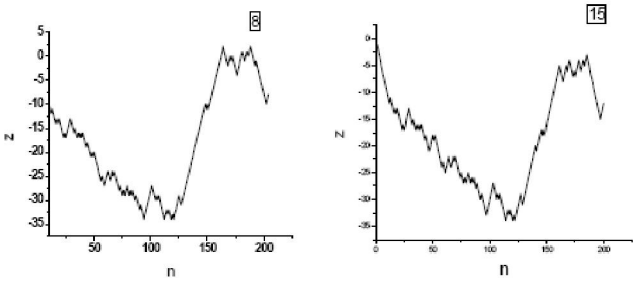

Figure 8 : The Z_n-n curves of BC200 RNA sequences 8 and 15

In addition, z_n and z'_n curves are almost same for all studied sequences. And the two type curves show a global minimum at the position of about 120bp. Furthermore, there is a sudden jump in the z'_n curve at the position about 130-160bp (see Fig.6 and 8).

Then we calculate the base content of A, C, G, T and GC in the studied sequences (see Table).

In sequences 2 and 15 the results are 33%, 29%, 25%, 14%, and 0.53, respectively. Nevertheless in sequence 8 the results are 34%, 28%, 24%, 13% and 0.52%, respectively. This fact indicates that there is minor difference on base content in different type ncRNA sequences.

Discussion

We can see obviously that all corresponding curves of sequences coming from same organism are almost no disparity, not only having same shapes but also same tendency (as shown in Figure1, 2 and 3). However, for sequences coming from different organism, some curves are different. That is to say, the value of z_n is different in the z_n curves. This fact indicates that the Z-curves of ncRNA sequences are connected with organism[18].

On top, in the z'_n curves of studied BC200 RNA and BC200-alpha RNA sequences, the value of z'_n is lesser than zero (see Figure2 and 6). The result shows that this type of ncRNA is a stable structure[18]. And in the z'_n curve the sudden jump indicates an A+T rich region at the position about 130-160bp. It might imply a transfer of foreign DNA sequence from other species and may be the position of the Integron Island[19].

Besides, in z_n and z'_n curves the global minimum at the position of about 120bp may be the function position of BC200RNAs.

FULL PAPER

In addition, we think the obvious disparity in some z curves (see Figure 8) is based on the base content in the selected ncRNA sequences.

CONCLUSION

In the z'_n curves of studied sequences z'_n <0 is regarded as a criterion to identify BC200 RNA.

The difference of z_n indicates that the Z-curves of studied sequences have some connection with organisms.

The sudden jump in the z'_n curve indicates a transfer of foreign DNA sequence from other species. The sudden jump in the z'_n curve may be the position of the Integron Island.

ACKNOWLEDGEMENTS

This work was supported by Chinese National Key Fundamental Research Project and Shandong Fundamental Research Project (Grant No. 90403120 and ZR2011FL011).

REFERENCES

[1] J.C.Castle, C.D.Armour, M.Lower, D.Haynor, M.Biery, H.Bouzek, R.Chen, S.Jackson, J.M.Johnson, C.A.Rohl et al; Digital genome-wide ncRNA expression, including SnoRNAs, across 11 human tissues using polyAneutral amplification. PLoS One, **5(7)**, e11779 **(2010)**.

[2] J.R.Neilson, G.X.Zheng, C.B.Burge, P.A.Sharp; Dynamic regulation of miRNA expression in ordered stages of cellular development. Genes & Dev., 21**,** 578–589 **(2007)**.

[3] T.Zhang, Z.H.Zhang, L.J.Ling, B.C.Shi, R.S.Chen; Conservation analysis of small RNA genes in Escherichia coli. Bioinformatics, **20**, 599-603 **(2004)**.

[4] D.D.Zhang, G.Q.Liu, Y.Qin et al.; Common chaperone activity in the G-domain of trGTPase protects L11–L12 interaction on the ribosom. Nucleic Acids Research,doi: 10.1093/nar/gks833, **(2012)**.

[5] R.Hershberg, S.Altuvia, H.Margalit; A survey of small RNA-encoding genes in Escherichia coli. Nucleic Acids Res, **31**, 1813-1820 **(2003)**.

[6] P.Boffetta, E.Couto, J.Wichmann et al.; Fruit and vegetable intake and overall cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). J Natl Cancer Inst, **8(102)**, 529-37 **(2010)**.

[7] G.Storz; An expanding universe of noncoding RNAs. Science, **296**, 1260-1263 **(2002)**.

[8] S.R.Eddy; Non-coding RNA genes and the modern RNA world. Nature Rev. Genet., **2**, 919-929 **(2001)**.

[9] G.Casadesus, M.A.Smith, S.Basu, J.Hua, D.E.Capobianco, S.L.Siedlak, X.Zhu; Increased isoprostane and prostaglandin are prominent in neurons in Alzheimer disease. Mol Neurodegener**, 22**, 2 **(2007).**

[10] K.E.Webster, J.R.Merory, J.E.Wittwer; Gait variability in community dwelling adults with Alzheimer disease. Alzheimer Dis Assoc Disord., **20(1)**, 37-40 **(2006)**.

[11] E.E.Tuppo, H.R.Arias; The role of inflammation in Alzheimer's disease.Int J Biochem Cell Biol, **37(2)**, 289-305 **(2005)**.

[12] E.E.Tuppo, H.R.Arias; The role of inflammation in Alzheimer's disease. Int J Biochem Cell Biol, **37(2)**, 289-305 **(2005)**.

[13] T.K.Cheung, H.H.Xia, B.C.Wong; Helicobacter pylori eradication for gastric cancer prevention. J. Gastroenterol, **42(17)**, 10–5 **(2007)** .

[14] D.Dingli, M.A.Nowak; Cancer biology: infectious tumour cells. Natur, **443(7107)**, 35–6 **(2006)**.

[15] C.N.Liu, B.Bai, Y.Geir Skogerbo, L.Cai, Deng, W.Zhang, Y.Bu, D.B.Zhao, Y.Chen; NONECODE: an integrated knowledge database of non-coding RNAs. Nucleic Acids Res., **33**, D112-D115 **(2005)**.

[16] R.Zhang, C.T.Zhang; Z curves, an intuitive tool for visualizing and analyzing DNA sequences. J.Biomol. Struc. Dyn., 767-782 **(1994)**.

[17] Zhang, C. T., R. Zhang and H. Y. Ou. (2003) The z curve database: a graphic representation of genome sequences. Bioinformatics, 19, 593-599.

[18] F.B.Guo, H.Y.Ou, C.T.Zhang; ZCURVE: A new system for recognizing protein-coding genes in bacterial and archaeal genomes., Nucleic Acids Res., **31**, 1780-1789 **(2003)**.

[19] R.Zhang, C.T.Zhang; Identification of replication origins in archaeal genomes based on the Z-curve method. Archaea, **1(5)**, 335-346 **(2005)**.