Proteomics in the Light of Integral Value Transformations

Sk. S. Hassan[1]*, P. Pal Choudhury*

Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

Emails: sarimif@gmail.com & pabitrapalchoudhury@gmail.com

Abstract

In this paper, Proteomics have been studied in the light of Integral Value Transformations (IVTs) which was introduced by the Sk. S. Hassan et al in 2010. For case study, a Human olfactory receptor OR1D2 protein sequence has been taken and then different IVTs have been used to evolve OR1D2 into some other proteomic like sequences. It has been observed that some of the generated sequences have been mapped to another olfactory receptor in Human or in some other species. Also it has been corroborated through fractal dimension that some of the fundamental protein properties have been nearly intact, even after the mapping. This study will help to comprehend proteomic evolutionary network with the help of IVTs.

***Keywords*:** Olfactory Receptors (ORs), Box-counting dimension, Proteomics.

1. *Introduction:*

The study of proteins such as structures, functions and evolutions is universally known to as *Proteomics***,** was first coined in 1997 to make an analogy with *Genomics*, the study of the genes [1]. After genomics, proteomics is considered the next step in the study of biological systems. The study of proteomics is important because proteins are responsible for both the structure and the functions of all living things. Genes are simply the instructions for making proteins. Therefore, a proper quantitative understanding of proteins characteristics and their inter network are required. In this paper, an olfactory receptor OR1D2 has been considered for the proteomics study. Interestingly, on applying the IVT systematically, we have been able to show that each of the DNA sequence at various discrete time instances in IVT evolutions can be directly mapped to another specific proteomic sequences existing in different species. A number of certain fundamental properties namely percentage of accessible residues, Alpha helix (Chou & Fasman), Amino acid composition (%), Beta sheet (Chou & Fasman), Beta turn (Chou & Fasman), Coil (Deleage & Roux), Hydrophobicity (Aboderin) and Total beta strand have been considered to ensure protein properties of the IVT generated sequences. All protein plots for all the IVT generated sequences including OR1D2 using Matlab (*bioinformatics toolbox*) have been generated. Then box-counting dimensions for each of the protein plot have been calculated through BENOIT^{TM}. Since protein properties remain intact in the bijective IVT generated sequences, we claim that in the event of replacement of proteomic sequence (which may take place for various reasons like diseases), we may follow the inverse map of the bijective IVTs to get back the same with usual properties. This study will help us to ascertain potential new drugs for the treatment of disease.

[1] *Corresponding author: Sk. S. Hassan (sarimif@gmail.com)*

2. *Some Reviews and Fundamentals*

In this section, we describe very briefly about IVTs, Fractal and proteins.

2.1 Notion of Integral Value Transformation (IVT):

Let us define the Integral Value Transformations (IVTs) in ${\mathbb{N}_0}^{\mathrm{K}}$ as the following [2, 3, 4, 5]:

$IVT^{p,k}{}_j : {\mathbb{N}_0}^{\mathrm{K}} \rightarrow \mathbb{N}_0$

$IVT^{p,k}{}_j((n_1, n_2, \ldots . n_k)=$

$(f_j({a_0}^{n_1}, {a_0}^{n_2}, \ldots, {a_0}^{n_k})\ f_j({a_1}^{n_1}, {a_1}^{n_2}, \ldots, {a_1}^{n_k}) \ldots \ldots f_j({a_{l-1}}^{n_1}, {a_{l-1}}^{n_2}, \ldots, {a_{l-1}}^{n_k}))_p = m$

where $n_1 = (a_0{}^{n_1} a_1{}^{n_1} \ldots . a_{l-1}{}^{n_1})_p$, $n_2 = (a_0{}^{n_2} a_1{}^{n_2} \ldots . a_{l-1}{}^{n_2})_p$, $\ldots\ldots$ $n_k = (a_0{}^{n_k} a_1{}^{n_k} \ldots . a_{l-1}{}^{n_k})_p$

$f_j: \{0, 1, 2, \ldots, p-1\}^k \rightarrow \{0,1,2, \ldots, p-1\}$.

m is the decimal conversion from the p adic number.

Let us fix the domain of IVTs as $\mathbb{N}_0$ (k=1) and thus the above definition boils down to the following:

$$\mathrm{IVT}^{\mathrm{p},1}{}_{\mathrm{j}}\ (\mathrm{x}) = \left(\mathrm{f_j}(\mathrm{x_n})\ \mathrm{f_j}(\mathrm{x_{n-1}}) \ldots \ldots \ldots \mathrm{f_j}(\mathrm{x_1})\right)_{\mathrm{p}} = \mathrm{m}$$

where m is the decimal conversion from the p adic number, and $\mathrm{x} = (\mathrm{x_n}\ \mathrm{x_{n-1}} \ldots \ldots \mathrm{x_1})_{\mathrm{p}}$.

Now, let us denote the set of $\mathrm{IVT}^{\mathrm{p},1}{}_{\mathrm{j}}$ as

$$\mathrm{T}^{\mathrm{p},1} = \left\{ \mathrm{IVT}^{\mathrm{p},1}{}_{\mathrm{j}} : \mathbb{N}_0 \rightarrow \mathbb{N}_0 \;\middle|\; \begin{array}{c} 0 \leq \mathrm{j} < \mathrm{p^p},\ \mathrm{IVT}^{\mathrm{p},1}{}_{\mathrm{j}}\ (\mathrm{x}) = \left(\mathrm{f_j}(\mathrm{x_n})\ \mathrm{f_j}(\mathrm{x_{n-1}}) \ldots \ldots \ldots \mathrm{f_j}(\mathrm{x_1})\right)_{\mathrm{p}} = \mathrm{m} \\ \text{where m is the decimal conversion from the p adic number} \\ \text{and } \mathrm{x} = (\mathrm{x_n}\ \mathrm{x_{n-1}} \ldots \ldots \mathrm{x_1})_{\mathrm{p}} \end{array} \right\}$$

Let us define the IVT in $\mathbb{N}_0$ in 4-adic number systems. There are 256 (4^{4^1}) one variable four state CA rules. Corresponding to each of those CA rules there are 256 IVTs are there in 4 adic system in one dimension.

$\mathrm{IVT}^{4,1}{}_{\#}$ is mapping a non-negative integer to a non-negative integer.

$$\mathrm{IVT}^{4,1}{}_{\#}(\mathrm{a}) = \left((\mathrm{f_\#}(\mathrm{a_n})\mathrm{f_\#}(\mathrm{a_{n-1}}) \ldots \mathrm{f_\#}(\mathrm{a_1})\right)_4 = b$$

Where 'a' is a non-negative integer and $a = (a_n a_{n-1} \ldots a_1)_4$ and 'b' is the decimal value corresponding to the 4-adic number.

For an example, let us consider $a = 225 = (3201)_4$ and $\# = 120$ so $\mathrm{f_\#}(0) = 0;\ \mathrm{f_\#}(1) = 2;\ \mathrm{f_\#}(2) = 3\ and\ \mathrm{f_\#}(3) = 1$

Therefore, $\mathrm{IVT}^{4,1}{}_{120}(225) = (\mathrm{f_{120}}(3)(\mathrm{f_{120}}(2)(\mathrm{f_{120}}(0)(\mathrm{f_{120}}(1))_4 = (1302)_4 = 114$.

Consequently, $\mathrm{IVT}^{4,1}{}_{120}(225) = 114$.

Let us denote $\mathfrak{T}^{4,1}{}_{\#}$ as set of all $\mathrm{IVT}^{\mathrm{p,k}}{}_{\#}$ transformations. It is worth nothing that there are 4! = 24 number of Bijective functions are there in $\mathfrak{T}^{4,1}{}_{\#}$. So of the 256 ($4^{4^1}$) transformations in $\mathfrak{T}^{4,1}{}_{\#}$ four are linear and rest are nonlinear [6].

2.2 Fractal and Fractal Dimension

Our artificial world can be described easily through Euclidean geometric shapes but there are many things in nature such as shape of cloud, geometry of lightening etc. could not be described through Euclidean geometry. Many mathematicians descended the challenge for a fair enough description of natural objects but after a long

period in 1975, B. Mandelbrot took the challenge and gave the birth of a new geometry to describe nature which is known to us as 'Fractal Geometry' in short 'Fractal'. The precise definition of "Fractal" according to Benoit Mandelbrot is as a set for which the Hausdroff Besicovitch dimension strictly exceeds the topological dimension. To gain a quantitative insight of Fractal, some fractal parameters namely Fractal dimension, Hurst exponent, succolarity, lacunarity etc. are also introduced in the literature. A brief discussion follows about one of the well-known methods of calculating fractal dimension namely '*Box-Counting method'*.

Box-Counting Method: This method computes the number of cells required to entirely cover an object, with grids of cells of varying size. Practically, this is performed by superimposing regular grids over an object and by counting the number of occupied cells. The logarithm of N(r), the number of occupied cells, versus the logarithm of 1/r, where r is the size of one cell, gives a line whose gradient corresponds to the box dimension [7].

2.3 Problem in Protein Structures

Proteins are an important class of biological macromolecules present in all organisms. After the structure of DNA was discovered by James Watson and Francis Crick, who used the experimental evidence of Maurice Wilkins and Rosalind Franklin (among others), serious efforts to understand the nature of the encoding of proteins began. George postulated that a three-letter code must be employed to encode the 20 standard amino acids used by living cells to encode proteins, because 3 is the smallest integer *n* such that 4^n is at least 20 [8]. The three-dimensional structures of proteins were first determined by X-ray diffraction analysis; Perutz and Kendrew shared the 1962 Nobel Prize in Chemistry for these discoveries. At the present time, more than ten thousand protein structures were found with their atomic details. The structure of the protein is ultimately defined by its primary structure, or amino acid sequence. There are no theories or computational techniques at the moment which will allow us to predict the new protein folding by its sequence. Even, how proteins were developed during organisms' evolution is blurred. Therefore proper understanding is required in the primary structure level i.e. in the amino acids sequence level of proteins.

3. Methods and Results:

3.1 *Method of Sequence generation through IVTs:*

The domain of act of IVTs is set of non-negative positive integers. So it is required to have a numeric sequence corresponding to each of the proteomic sequence. A simple mapping *f* is used to have as defined below:

Let $\mathcal{P} = \{A, C, D, E, F, G, H, I, K, L, Q, N, P, Q, R, S, T, V, W, Y\}$ be the set of amino acid codes and $\mathcal{N} = \{0,1,2,3, \dots, 19\}$.

$$f: \mathcal{P} \to \mathcal{N} \; as \; f(A) = 0; f(C) = 1, f(D) = 2, f(E) = 3, \dots, f(W) = 18 \; and \; f(Y) = 19$$

Therefore, a protein sequence is now simply a string of twenty variables namely 0, 1, 2…19 as per coding scheme *f*.

Starting from a protein sequence to generate another proteomic like sequences, it is required to have the IVTs in a particular ${\mathfrak{I}^{p,1}}_{\#}$, which maps $\mathcal{N}$ to itself.

The list of such IVTs in $\mathfrak{T}^{p,1}{}_{\#}$ is given below in Table-I.

P-adic $\mathfrak{T}^{p,1}{}_{\#}$	$\mathfrak{T}^{2,1}{}_{\#}$	$\mathfrak{T}^{3,1}{}_{\#}$	$\mathfrak{T}^{4,1}{}_{\#}$	$\mathfrak{T}^{5,1}{}_{\#}$	$\mathfrak{T}^{6,1}{}_{\#}$	$\mathfrak{T}^{7,1}{}_{\#}$	$\mathfrak{T}^{8,1}{}_{\#}$	$\mathfrak{T}^{9,1}{}_{\#}$
#	1 2	5 11	99 114	194 214	28565 28595	297051	5135375	102907844
		21	147	294 334	28745 28805	297093	5135431	102907916
			177	414 434	28955 28985	297393	5135886	102908572
			180	694 714	29860 29890	297435	5135942	102908644
			210	894	30040 30100	299109	5138959	102913676
			225			299151	5139015	102913748
			228			299793	5139981	102915132
						299835		

Table-1: IVTs in $\mathfrak{T}^{p,1}{}_{\#}$

All the mentioned IVTs are essentially bijective in nature for the purpose of switching one to another. Now we apply Integral Value Transformations ($IVT^{p,1}{}_{\#}$) systematically [3, 9, 10] :-

Firstly, Divide the whole one dimensional initial sheet of proteomic sequence (numeric sequence) of length *n* and divided it into *r* multiple blocks. We designate the initial sequence as $S(t_0)$.

Secondly, we apply bijective domain preservative transformations (need not to be all distinct) taken from $\mathfrak{T}^{p,1}{}_{\#}$ (for different p starting from 2 to 19) over each of the *r* different blocks of $S(t_0)$. Thereby, we call this case as *Hybrid Application of IVTs.* In other words, we are getting $S(t_1)$ from $S(t_0)$ through hybrid application of IVTs. Next, we follow this step successively as long as we wish to iterate.

The results, on applying the proposed systematic technique of application of IVTs on OR1D2 are enumerated in the following subsections.

3.2. *Results*

Here we discuss the results on applying different IVTs in two following cases.

3.2.1: On Applying $IVT^{p,1}{}_{\#}$

The proteomic sequence of OR1D2 is 312 long (sequence shown below in Text-1). Choose r =50**,** so there are 7 blocks are there. The following IVTs are used to generate $S(t_1)$ as shown in Table-2.

MDGGNQSEGSEFLLLGMSESPEQQRILFWMFLSMYLVTVVGNVLIILAIS

SDSRLHTPVYFFLANLSFTDLFFVTNTIPKMLVNLQSHNKAISYAGCLTQ

LYFLVSLVALDNLILAVMAYDRYVAICCPLHYTTAMSPKLCILLLSLCWV

LSVLYGLIHTLLMTRVTFCGSRKIHYIFCEMYVLLRMACSNIQINHTVLI

ATGCFIFLIPFGFVIISYVLIIRAILRIPSVSKKYKAFSTCASHLGAVSL

FYGTLCMVYLKPLHTYSVKDSVATVMYAVVTPMMNPFIYSLRNKDMHGAL

GRLLDKHFKRLT

Text-1: Protein Sequence of OR1D2.

BLOCK	$S(t_1)$ in 2 adic IVT	$S(t_1)$ in 3 adic IVT	$S(t_1)$ in 4 adic IVT
Block-1	$IVT^{2,1}{}_{1}$	$IVT^{3,1}{}_{5}$	$IVT^{4,1}{}_{99}$
Block-2	$IVT^{2,1}{}_{1}$	$IVT^{3,1}{}_{5}$	$IVT^{4,1}{}_{114}$
Block-3	$IVT^{2,1}{}_{2}$	$IVT^{3,1}{}_{11}$	$IVT^{4,1}{}_{147}$
Block-4	$IVT^{2,1}{}_{1}$	$IVT^{3,1}{}_{11}$	$IVT^{4,1}{}_{177}$
Block-5	$IVT^{2,1}{}_{2}$	$IVT^{3,1}{}_{21}$	$IVT^{4,1}{}_{180}$
Block-6	$IVT^{2,1}{}_{2}$	$IVT^{3,1}{}_{21}$	$IVT^{4,1}{}_{210}$
Block-7	$IVT^{2,1}{}_{2}$	$IVT^{3,1}{}_{21}$	$IVT^{4,1}{}_{225}$
BLOCK	$S(t_1)$ **in 5 adic IVT**	$S(t_1)$ **in 6 adic IVT**	$S(t_1)$ **in 7 adic IVT**
Block-1	$IVT^{5,1}{}_{194}$	$IVT^{6,1}{}_{28565}$	$IVT^{7,1}{}_{297051}$
Block-2	$IVT^{5,1}{}_{214}$	$IVT^{6,1}{}_{28595}$	$IVT^{7,1}{}_{297093}$
Block-3	$IVT^{5,1}{}_{294}$	$IVT^{6,1}{}_{28745}$	$IVT^{7,1}{}_{297393}$
Block-4	$IVT^{5,1}{}_{334}$	$IVT^{6,1}{}_{28805}$	$IVT^{7,1}{}_{297435}$
Block-5	$IVT^{5,1}{}_{414}$	$IVT^{6,1}{}_{28985}$	$IVT^{7,1}{}_{299109}$
Block-6	$IVT^{5,1}{}_{434}$	$IVT^{6,1}{}_{28955}$	$IVT^{7,1}{}_{299151}$
Block-7	$IVT^{5,1}{}_{694}$	$IVT^{6,1}{}_{29860}$	$IVT^{7,1}{}_{299793}$
BLOCK	$S(t_1)$ **in8 adic IVT**	$S(t_1)$ **in 9 adic IVT**	
Block-1	$IVT^{8,1}{}_{5135375}$	$IVT^{9,1}{}_{102907844}$	
Block-2	$IVT^{8,1}{}_{5135431}$	$IVT^{9,1}{}_{102907916}$	
Block-3	$IVT^{8,1}{}_{5135886}$	$IVT^{9,1}{}_{102908572}$	
Block-4	$IVT^{8,1}{}_{5135942}$	$IVT^{9,1}{}_{102908644}$	
Block-5	$IVT^{8,1}{}_{5138959}$	$IVT^{9,1}{}_{102913676}$	
Block-6	$IVT^{8,1}{}_{5139015}$	$IVT^{9,1}{}_{102913748}$	
Block-7	$IVT^{8,1}{}_{5139981}$	$IVT^{9,1}{}_{102915132}$	

Table-2: IVTs from $\mathfrak{T}^{p,1}{}_{\#}$ used for generation of $S(t_1)$

Similarly, others $S(t_i)$ can be generated applying the IVTs in different blocks of the $S(t_{i-1})$ as tabulated in *supl. met.-I*. We have generated 90 such $S(t_i)$s corresponding to OR1D2 in each $\mathfrak{T}^{p,1}{}_{\#}$ system (for p=2, 3… 19) (available in *supl. met.-II*).

All these generated sequences have been blast in the NCBI database for significant similarity. The blast result is shown in *supl. met.-III*.

Most of the generated sequences are mapped to olfactory receptors (specifically more closed to OR1D2) in different organisms like *homo sapiens*, *pan troglodytes*, *lagothrix lagotricha* and etc. Some of the sequences are not, due to the fact that they are more conserved sequence than OR1D2.
Also we have been observed that some of the protein primary structural properties (listed below) are intact with respect to the two dimensional protein plot graphs (using bioinformatics toolbox of Mtlab-R2010b) for each of the generated sequences.
The protein properties which we have considered here are as follows:

- Prop-1: Accessible residues (%)
- Prop-2: Alpha helix (Chou & Fasman)
- Prop-3: Amino acid composition (%)
- Prop-4: Beta sheet (Chou & Fasman)
- Prop-5: Beta turn (Chou & Fasman)
- Prop-6: Coil (Deleage & Roux)
- Prop-7: Hydrophobicity (Aboderin)
- Prop-8: Total beta strand

Corresponding to each property of the $S(t_i)$, we have had eight protein plot graphs from which we have calculated box counting dimensions using BENOIT™.
The data for OR1D2 sequence are stated below in the table-III. The rest all data are available in the *supl. met-IV*.

Sequence	Property	Box-counting dimension
OR1D2	Prop1	1.91092
	Prop2	1.91103
	Prop3	1.90855

	Prop4	1.91141
Sequence	**Property**	**Box-counting dimension**
	Prop5	1.91095
	Prop6	1.91348
	Prop7	1.90989
	Prop8	1.91071

Table-3: Box-counting dimension for protein plots of OR1D2

We have observed that box-counting dimensions for all the eight protein plots corresponding to each of the protein property for all the generated sequences $S(t_i)$s are almost same to the same of OR1D2. The data for all the box counting dimension of protein plots for the $S(t_i)$ generated through the $\mathfrak{T}^{2,1}{}_{\#}$ system is shown below. Hereby we can come to a conclusion that these IVTs preserve the protein properties of the strings. It is to be noted that all these IVTs are bijective; therefore one can switch from one protein to another protein through the IVTs without encumbering the protein properties.

Sequence	**Property**	**Box-counting dimension**
$S(t_1)$	Prop1	1.92694
	Prop2	1.91117
	Prop3	1.90976
	Prop4	1.91111
	Prop5	1.9113
	Prop6	1.93038
	Prop7	1.91021
	Prop8	1.91144
$S(t_2)$	Prop1	1.91124
	Prop2	1.91099
	Prop3	1.91389
	Prop4	1.90948
	Prop5	1.91064
	Prop6	1.93051
	Prop7	1.91398
	Prop8	1.90983
$S(t_3)$	Prop1	1.91045
	Prop2	1.91049
	Prop3	1.90994
	Prop4	1.91299
	Prop5	1.92765
	Prop6	1.91648
	Prop7	1.92813
	Prop8	1.91448
$S(t_4)$	Prop1	1.91294
	Prop2	1.91495
	Prop3	1.91084
	Prop4	1.9108
	Prop5	1.91155
	Prop6	1.91577
	Prop7	1.9281
	Prop8	1.93043
$S(t_5)$	Prop1	1.91443
	Prop2	1.91431
	Prop3	1.91259
	Prop4	1.93055
	Prop5	1.92909
	Prop6	1.91638
	Prop7	1.92901
	Prop8	1.91676
$S(t_6)$	Prop1	1.92863
	Prop2	1.928
	Prop3	1.91431
	Prop4	1.9295
	Prop5	1.91133
	Prop6	1.91751
	Prop7	1.91379
	Prop8	1.91292
$S(t_7)$	Prop1	1.91421
	Prop2	1.928
	Prop3	1.9142
	Prop4	1.91614
	Prop5	1.9101
	Prop6	1.91402
	Prop7	1.9108
	Prop8	1.91314
$S(t_8)$	Prop1	1.9104
	Prop2	1.91378
	Prop3	1.91039
	Prop4	1.91287
	Prop5	1.91177
	Prop6	1.91392
	Prop7	1.90987
	Prop8	1.91378
$S(t_9)$	Prop1	1.91428
	Prop2	1.91129
	Prop3	1.91367
	Prop4	1.91337
	Prop5	1.91263

	Prop6	1.91431
	Prop7	1.91084
	Prop8	1.91413
$S(t_{10})$	Prop1	1.91082
	Prop2	1.9108
	Prop3	1.91081
	Prop4	1.91337
	Prop5	1.91263
	Prop6	1.91514
	Prop7	1.91084
	Prop8	1.9176

Table-4: Box-counting dimension for all protein plots of $S(t_i)$ in $\mathfrak{T}^{2,1}{}_{\#}$

Most of the $S(t_i)$, IVT generated sequences preserve the all eight protein properties. It is to be noted that in the case $\mathfrak{T}^{2,1}{}_{\#}$ system, the $S(t_1)$ and $S(t_2)$ are both mapped to G-protein-coupled receptor in OR1D2 in human. Also they follow all the protein properties as in OR1D2.
But interestingly, there are many $S(t_i)$ in different $\mathfrak{T}^{p,1}{}_{\#}$ systems, which do not map significantly in any organisms but they retain the protein properties as in OR1D2. One of the main reasons is that most of the sequences are conserved (restricted to a few amino acids) whereas OR1D2 is not so. Some of the $S(t_i)$ are not mapped to any of the ORs in any organism although the box-counting dimension for all the protein plots are intact as it is in OR1D2. It is our strong conviction that these $S(t_i)$ serve the purpose for replacement of OR1D2 in genetic evolutionary phase. In the next section we are going to discuss the case on applying the bijective IVTs from $\mathfrak{T}^{20,1}{}_{\#}$.

3.2.2 On Applying $IVT^{20,1}{}_{\#}$

We have chosen a few bijective IVTs (available in *supl. met.-I)* from $\mathfrak{T}^{20,1}{}_{\#}$ system to generate $S(t_i)$ from the protein code for OR1D2 (methodology is discussed in 3.1). Here all the $S(t_i)$ have been blasted in NCBI and they all mapped to G-protein-coupled receptor, OR MOR30-1, hypothetical protein, conserved hypothetical protein etc. in different organisms ranging from human to Plasmodium species (data shown in *supl. met.-III*). The box counting dimension is still intact for all the protein plots for all the IVT generated sequence in $\mathfrak{T}^{20,1}{}_{\#}$ system as shown in Figure-I (raw data shown in *supl. met-IV)*.

Sequence	**Property**	**Box-counting dimension**
$S(t_1)$	Prop1	1.90836
	Prop2	1.91371
	Prop3	1.92937
	Prop4	1.91313
	Prop5	1.92746
	Prop6	1.9128
	Prop7	1.91234
	Prop8	1.91291
$S(t_2)$	Prop1	1.91418
	Prop2	1.91204
	Prop3	1.91182
	Prop4	1.91205
	Prop5	1.91418
	Prop6	1.92998
	Prop7	1.9099
	Prop8	1.91351
$S(t_3)$	Prop1	1.91459
	Prop2	1.91308
	Prop3	1.91151
	Prop4	1.91464
	Prop5	1.91434
	Prop6	1.91216
	Prop7	1.91306
	Prop8	1.91321
$S(t_4)$	Prop1	1.91087
	Prop2	1.91468
	Prop3	1.90957
	Prop4	1.90991
	Prop5	1.92755
	Prop6	1.9159
	Prop7	1.9104
	Prop8	1.91369
$S(t_5)$	Prop1	1.91448
	Prop2	1.91485
	Prop3	1.92691
	Prop4	1.914
	Prop5	1.9123
	Prop6	1.91203
	Prop7	1.92751
	Prop8	1.92845
$S(t_6)$	Prop1	1.91315
	Prop2	1.91176
	Prop3	1.91169
	Prop4	1.91317

	Prop5	1.91348
	Prop6	1.91507
	Prop7	1.91141
	Prop8	1.92879
$S(t_7)$	Prop1	1.91258
	Prop2	1.91057
	Prop3	1.91388
	Prop4	1.91508
	Prop5	1.92907
	Prop6	1.91605
	Prop7	1.91244
	Prop8	1.91098
$S(t_8)$	Prop1	1.92725
	Prop2	1.92767
	Prop3	1.91331
	Prop4	1.91074
	Prop5	1.91459
	Prop6	1.91608

	Prop7	1.90883
	Prop8	1.91143
$S(t_9)$	Prop1	1.90984
	Prop2	1.92917
	Prop3	1.9154
	Prop4	1.91098
	Prop5	1.91336
	Prop6	1.91545
	Prop7	1.91013
	Prop8	1.92845
$S(t_{10})$	Prop1	1.91286
	Prop2	1.91425
	Prop3	1.91506
	Prop4	1.91402
	Prop5	1.92938
	Prop6	1.91632
	Prop7	1.91337
	Prop8	1.9125

Table-5: Box-counting dimension for all protein plots of $\mathbf{S(t_i)}$ in $\mathfrak{T}^{20,1}{}_{\#}$

It is noted that the number of bijective, domain preservative IVTs is increased as p increased in $\mathfrak{T}^{p,1}{}_{\#}$. Consequently the sequential conservation is inversely proportional to p.

4. Summary and Concluding Remarks:

In summary, we have seen that IVTs steer a given OR sequence of a species to another of the same or different (most likely) species, preserving the protein properties of the original sequence. This methodology will be helpful to mimic the genomic evolution procedure artificially, which is required for genetic replacement therapy. IVTs may also be considered to be a platform to comprehend the morphological connections among the various species.

A naïve question to the Biologists can be raised as in the following:

Suppose, we are given an olfactory receptor *or1* of a species *s1*which help it to identify the odors *x1, x2…*

Now, we apply the proposed methodology to *or1* and obtain a new olfactory receptor *or2* (supposedly) of species *s2*.

So, does *or2* help *s2* in identifying the same odors *x1, x2*,...?

In near future, we are really interested to explore the underlying biological methodology that governs the entire process.

Acknowledgement: Authors would like to thanks to their visiting students ***Anjan Pal*** and **Snehasish Banerjee** for their enormous help in writing computer programs.

References:

[1] Crick, Francis (1988). "Chapter 8: The genetic code". What mad pursuit: a personal view of scientific discovery. *New York: Basic Books*. pp. 89–101.

[2] Hassan Sk. S., Pal Choudhury Pabitra, Singh R. Das, S. and Nayak B K (2010) Collatz Function like Integral Value Transformations, *Alexandria Journal of Mathematics*, Vol 1(2), 31-35.

[3] Hassan Sk. S., Pal Choudhury. Pabitra, Guha R., Chakraborty, S., Goswami, A. (2011) DNA Sequence Evolution through Integral Value Transformations. (In press of the journal *Interdisciplinary Sciences--Computational Life Sciences*, Springer).

[4] Hassan, Sk. S. et al. (2010) Underlying Mathematics in Diversification of Human Olfactory Receptors in Different Loci. *Nature Precedings: npre.2010.5475.1*

[5] Hassan, Sk. S. et al. (2011) Quantification of miRNAs and Their Networks in the light of Integral Value Transformations. *Nature Precedings: npre.2011.6035.2*

[6] Hassan, Sk. S. et al. (2011) One Dimensional p-adic Integral Value Transformations, *arXiv:1106.3586*

[7] B. Mandelbrot, The Fractal Geometry of Nature. W.H. Freeman and Company. *ISBN 0-7167-1186-9.*

Briggs, John (1992).

[8] Yu Xia and Michael Levitt (2004) Simulating protein evolution in sequence and structure space, *Current Opinion in Structural Biology*, 14:202–207

[9] Carlo Cattani, (2010) Fractals and hidden symmetries in DNA, *Mathematical Problems in Engineering* Volume 2010.

[10] Yu Zu-Guo *et al* (2002) Fractals in DNA sequence analysis, *Chinese Phys.* Vol 11 Num. 12.