COMPUTER-AIDED DRUG RESISTANCE CALCULATOR
A RATIONAL AND RAPID POINT-OF-CARE DIAGNOSTIC DEVICE

NWANKWO NORBERT and NJOKU NGOZIKA

EXECUTIVE SUMMARY: We have earlier invented a rational and rapid point-of-care diagnostic device called Computer Aided Drug Resistance Calculator (CADRC). Patent Publication for this device has been made and application is still in progress. It has been identified that variations in the physio-chemical and structural properties (phenotypic characteristics) of the target proteins of drugs that result in drug resistance also translate into alterations in the amino acid sequences or mutations (genotypic features). CADRC engages a Digital Signal Processing (DSP) technique called Information Spectrum Method (ISM) to translate these genotypic features of protein targets of drug back to phenotypic characteristics using all the Amino Acid Scales (AAS) involved. These phenotypic attributes (in this case, drug resistance) presented as numerical sequences (signals) are further processed and quantified using Discrete Fourier Transform (DFT). CADRC uses computer algorithm to immediately deliver drug resistance and in numerical terms, enabling the healthcare provider-use to quickly make informed decision on appropriate antibiotic of use. In this presentation, we propose that the CADRC's algorithm be centrally located for quality control monitoring, upgrade and update while the healthcare service providers will be afforded, through their Internet-enabled phones, office computers e.t.c, opportunity to deposit and process sequences derived from their samples and also retrieve calculated resistance. Fluconazole-resistant candida albican is used to demonstrate the functioning of the CADRC. CADRC is employable on all drugs as they are known to be either proteins, have protein targets or proteins encoding them. In this submission, a rational, highly sensitive, reliable and rapid point-of-care diagnostic device called CADRC, which is effective in tackling the emerging antimicrobial resistance is presented. The procedure engaged has earlier fetched us another International Award-InnoCentive Award Winning Solver for computationally "Assessing Vaccine Potency" (ID:9933477).

Introduction: We have preliminarily invented a rational and rapid point-of–care diagnostic device which computes and present resistance offered to various therapeutic agents by their target proteins. This is called Computer-Aided Drug Resistance Calculator (CADRC) [1]. Patent publication for the device has been made (US20150370964) and Patent Application is still in progress. CARDC originated

from [2]. Its version (a proposal) was also published [3]. The entire results (ten figures and a table) presented in [3] were obtained from [2].

Alterations in the physio-chemical and structural properties (phenotypic behavior) of the target proteins of the drugs initiated by organisms and viruses in an attempt to overcome the harmful effects of these therapeutic agents are known to also translate to mutations or protein residue alterations (genotypic manifestations) [2,3]. As demonstrated in the Patent submission [1], CADRC operates by first engaging all the Amino Acid Scales (AASs) [4] to translate these genotypic information back into phenotypic characteristics and in this instance, drug resistance.

Amino Acid Scales [4] are set of twenty (20) numerical values that describe the level of participation of each of the 20 essential amino acids in the phenotypic (physio-chemical and structural) interactions [5]. There are about 565 Amino Acid Scales [6] and some of them are deposited in a database [4]. Twenty-two (22) of these AASs have been engaged to study the resistance offered to Amprenavir by its target protein, the HIV-I protease enzyme [6].

The translated phenotypic properties obtained by interchanging the alphabetic codes of the protein sequences with their corresponding values of a particular AAS, which is also called signals are then processed using Discrete Fourier Transform (DFT)-based technique called Informational Spectrum Method (ISM) [7,8]. This procedure presents the processed phenotypic characteristics of the drug target proteins in numerical terms. This enables the healthcare providers and other users to quickly make more informed decisions on appropriate therapeutic agents (in this case antibiotic of use).

Additionally, CADRC engages a computer algorithm to carry out the ISM-based processes and as a result, delivers the computed drug resistance in seconds. Consequently, the drug resistance outcomes are delivered almost immediately for used by the healthcare providers. This is unlike other preliminary procedures such as Culture and Sensitivity Antimicrobial Resistance Assay, which takes 48 hours for incubation only [9]. In the Patent Submission [1], we explained the functioning of the CADRC using Amprenavir, an antiretroviral agent and in target protein, HIV-I protease Enzyme. HIV-I Protease Enzyme consist of 99 amino acids [10]. Here, we will engage the Fluconazole-resistance *Candida albican*.

The Fourier Transform-based DSP technique engaged in the CADRC is the foundation of other technologies like Radar, Image processing [6]. Radar is the apparatus that controls all the aircrafts in

the space. This is one technology that has reliably served mankind for ages. This goes to authenticate the robustness and reliability of the device. Additionally, to study the sensitivity of the device, ten (10) randomly generated and highly hetergenous sequences were employed in the Patent Submission [1]. The sequences were incorporated into the dataset for an Amino Acid Scale designated Ra, and analyzed. The device reacted by producing resistance as high as 100% while the sequences of the HIV-I protease Enzyme demonstrated resistance range of 0-23%.

CADRC does not require any reagent, equipment but computers and sequence information. Notwithstanding, the sequences can only be obtained from the sequencing facilities that will require deposits of the samples (microbes) and a sequencing technique such as Polymer Chain Reaction (PCR). PCR sequencing exercises are expensive. This appears to be the only limitation of the techniques. However, other cheaper sequencing techniques can be used.

In order to ensure accuracy of the result generated by the device, appropriate monitoring of the results ansapparatus as well as its updates and upgrade will be taken into consideration, the computer program of the CADRC [1] will be centrally and securely kept. Similar program exists [11]. The healthcare providers will be afforded with a facility in their phones, laptops, office computers e.t.c to deposit their sequences and run the CADRC program in order to generate the resistance, which will be immediately and numerically presented.

Currently, available Antimicrobial Resistance Diagnostic apparatus including culture and sensitivity test are known to engage or consume lots of resources (reagents, equipment e.t.c) and time [9]. They are envisage to introduce handling error such as unwanted microbes [1]. In some poorer nations, some laboratory technicians may hand over falsified results to their patients simply because unaffordable resources may involved in carrying out these experiments such as in the case of Gonorrhea. CADRC requires only the sequence information which will be processed before the watching eyes of the patients in the Internet-enabled phones, laptops and computers belonging to the Healthcare Providers. As a result, CADRC does not give room for the falsification of results.

In the subsequent sections, the functioning of the CADRC as submitted [1] and its engagement in the determination characterization and quantification of resistance offered to Fluconazole-resistant *Candida albican* are presented.

COMPUTER-AIDED DRUG RESISTANCE CALCULATOR: HOW DOES IT WORK: In this section we explain how Computer-Aided Drug Resistance Calculator works as earlier presented [1].

Procedure: (a) Materials: As demonstrated in in the Patent Submission [1], the Consensus Sequence of the HIV-1 protease enzyme, consisting of 99 amino acids, which is the target protein for Amprenavir, an antiretroviral agent, is retrieved from a database [10]. The 295 mutations, preliminarily investigated using 22 Amino Acid Scales [6] were constructed based on the consensus sequence. To validate the reliability of the CADRC, 10 randomly generated sequences were obtained. An example of the consensus, mutated and randomly generated sequences is displayed in Figure 1, courtesy of [1].

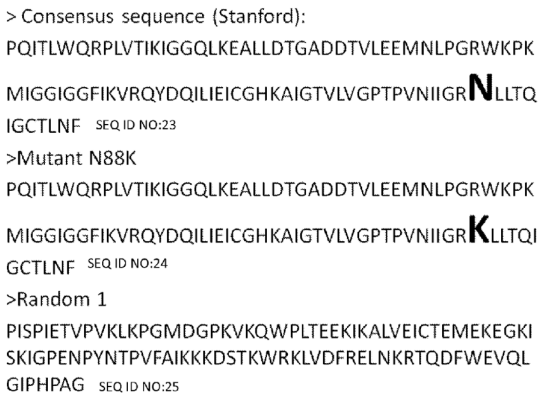

Fig 1: Showing example of consensus sequence and mutated sequences of the HIV-1 protease enzyme, the target protein for Amprenavir; as well as the randomly generated sequence used in the patent Submission [1].

(b) The Informational Spectrum Method: All the sequences are processed using ISM procedure [1-3,5,7,8], which is briefly described below:

Step 1: Conversion of the Alphabetic codes of the Amino Acids of the protein residues involved into numbers using the Amino Acid Scales: As in the submission [1] the alphabetic codes of all the sequences comprising of the 295 mutated and 10 randomly generated sequences are translated into numerical sequences using the 22 AASs involved. This brings to 305, the sequences processed. Each sequence is 99 amino acid length. As a result, there are 664, 290 (305 x 99 x 22) computations and 6,710 (305 x 22) plots.

Step 2: Decomposition and Processing of the Numerical Sequences (Signal) using Discrete Fourier Transform (DFT): Each of the 305 sequences is then decomposed, processed and analyzed using DFT. This process results in a plot of the Absolute values of the DFT-derived Amplitudes (on the y-axis) and the Frequency or Position of the Amino Acid (x-axis). This is called

Informational Spectrum (IS). The Amplitudes depicts the magnitude of interaction at various frequencies or protein residue positions.

Step 3: Point wise Multiplication of the DFT-derived outcomes (Common Informational Spectrum): In order to ascertain the common point of interaction for the sequences, point-wise multiplication is carried out. This process, which is termed the Common Informational Spectrum (CIS) identifies the common positions of interactions from which contributions by the susceptible and resistant sequences are calculated and aggregated. The values are normalized such that the sum of amplitudes is equal to one [1]. This is to ease calculation. In other cases, normalization is achieved by deriving percentage values of the amplitude [2,3]. The common position is called Consensus Frequency (CF). In the Patent Submission [1], we stated that aggregation of the results of all mutations using all AASs engaged is needed to obtain the actual resistance. The procedure described in [1] is an established procedure for calculating drug resistance.

Figures 1-3 are the CIS of all the mutants of position 88 of the HIV-1 protease enzyme; IS of the susceptible (N88A) and resistant sequence (N88F) as obtained from [1].

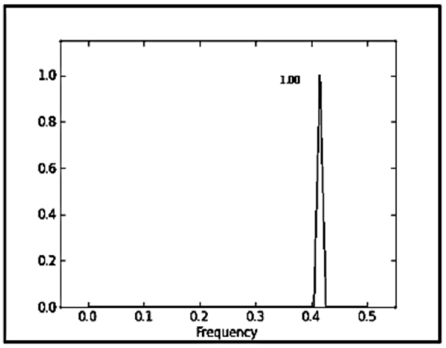

Figure 2: Showing the Common Information Spectrum (CIS) of all the mutants of position 88 of the HIV-1 Protease Enzyme (using AAS cased Ra) that resisted Amprenavir in the Patent Submission [1]. The consensus frequency (CF) is at 0.4149. (courtesy of [1]).

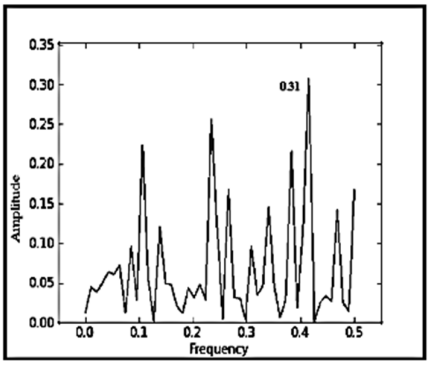

Figure 3A: showing the Informational Spectrum of the susceptible sequence N88A, which demonstrated an amplitude (level of interaction) of 0.31 (100%) at the consensus frequency of 0.4149, courtesy of [1].

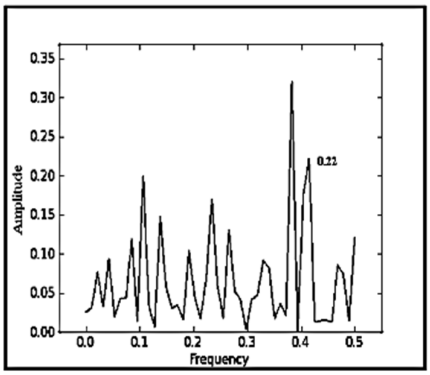

Figure 3B: showing the Informational Spectrum of the susceptible sequence N88F, which demonstrated an amplitude (level of interaction) of 0.22 (71%) at the consensus frequency of 0.4149, leaving a resistance of 29%, courtesy of [1].

Step 4: The Calculation of the Actual Resistance: In the Patent Submission [1] we aggregated the resistance offered by the 295 mutated sequences using all AASs, and arrived at a tentative result, 5.86%. This was achieved by subtracting the amplitude magnitude of interaction demonstrated by the resistance strains from that of the susceptible isolates. This was then averaged to obtain the actual resistance offered to Amprenavir by the HIV-I protease enzyme as 5.86%. As recorded in the Patent Submission [1], and as shown in figure 3, the mutant N88A for example, which came from the most susceptible strain presented an amplitude of 0.31 (100%). The resistant mutant, N88F provided a reduced susceptibility (amplitude) of 0.22 (71%). The calculated resistance is then obtained as 29%.

CADRC: The Fluconazole–Resistant Candida Albican Experience: Global, worrisome and abundant incidences of systemic fungal intention arising from immuno-deficiency as a result of Organ and Bone Marrow Transplant, HIV/AIDS, and Cancer therapy have been reported [12]. Additionally invasive fungal infections have been identified an a serious menace in Hematology, Oncology and Intensive Care [13]. *Candida albican* remains one of the primary opportunistic organisms associated with this health hazard.

Triazole antifungal including Fluconazole had preliminarily served an salvage therapy for fungal infection until the emergence of resistance [12]. As a result, Fluconazole-resistant *candida albican* is now listed as one of the 18 drug-resistant bacteria of highest concern in the National Action Plan for combating Antibiotic-Resistant Bacteria.

Triazole are known to inhibit fungal growth by interfering with Ergosterol, a cell wall component, through the blockage of the enzymatic activities of CYP51A1 that is encoded by ERG11 gene [12,13].

Mutations in the ERG11 gene, responsible for fluconazole resistance are therefore used as an example in this submission. Reduced affinity of the azole to the target enzyme, CYP1A1 resulting from point mutations in the ERG11 have been verified as a mechanism of azole resistance [12]. As a result, an AAS that governs affinity called Electron-Ion Interaction Potential (EIIP) is employed. About 60 mutations have been recognized to result in reduce susceptibility. Y132F, Y132H, K143R and K143Q have been found to confer high level of resistance [12,13]. K143Q, K143R, and Y132F are recognized to yield Minimal Inhibitory Concentration (MIC) greater than 256 microgram/ml. while Y132H affords 64 microgram/ml [13].

Currently, Fluconzole-resistant *Candida albican*, has no known susceptible strains with mutations at positions 132 and 143 and as a result, consensus sequences is engaged here. We had preliminarily engaged consensus sequences as susceptible sequence [2,3]. This is because they are naïve both in activity and alignment. Both the consensus and the mutated sequences of the ERG11 are therefore subjected to ISM procedure using EIIP. Unlike in the Patent Submission [1] the DFT-derived results are not normalize in the Plot or Figures.

UNIPROT ID: P10613

MAIVETVIDGINYFLSLSVTQQISILLGVPFVYNLVWQYLYSLRKDRAPLVFYWIPWFGS
AASYGQQPYEFFESCRQKYGDVFSFMLLGKIMTVYLGPKGHEFVFNAKLSDVSAEDAYKH
LTTPVFGKGVIYDCPNSRLMEQ**K**KFAKFALTTDSFKRYVPKIREEILNYFVTDESFKLKE
KTHGVANVMKTQPEITIFTASRSLFGDEMRRIFDRSFAQLYSDLDKGFTPINFVFPNLPL
PHYWRRDAAQKKISATYMKEIKSRRERGDIDPNRDLIDSLLIHSTYKDGVKMTDQEIANL
LIGILMGGQHTSASTSAWFLLHLGEKPHLQDVIYQEWEL LKEKGGDLNDLTYEDLQKLP
SVNNTIKETLRMHMPLHSIFRKVTNPLRIPETNYIVPKGHYVLVSPGYAHTSERYFDNPE
DFDPTRWDTAAAKANSVSFNSSDEVDYGFGKVSKGVSSPYLPFGGGRHRCIGEQFAYVQL
GTILTTFVYNLRWTIDGYKVPDPDYSSMVVLPTEPAEIIWEKRETCMF

MUTATED SEQUENCES: (K143R)

MAIVETVIDGINYFLSLSVTQQISILLGVPFVYNLVWQYLYSLRKDRAPLVFYWIPWFGS
AASYGQQPYEFFESCRQKYGDVFSFMLLGKIMTVYLGPKGHEFVFNAKLSDVSAEDAYKH
LTTPVFGKGVIYDCPNSRLMEQ**R**KFAKFALTTDSFKRYVPKIREEILNYFVTDESFKLKE
KTHGVANVMKTQPEITIFTASRSLFGDEMRRIFDRSFAQLYSDLDKGFTPINFVFPNLPL
PHYWRRDAAQKKISATYMKEIKSRRERGDIDPNRDLIDSLLIHSTYKDGVKMTDQEIANL
LIGILMGGQHTSASTSAWFLLHLGEKPHLQDVIYQEWEL LKEKGGDLNDLTYEDLQKLP
SVNNTIKETLRMHMPLHSIFRKVTNPLRIPETNYIVPKGHYVLVSPGYAHTSERYFDNPE
DFDPTRWDTAAAKANSVSFNSSDEVDYGFGKVSKGVSSPYLPFGGGRHRCIGEQFAYVQL
GTILTTFVYNLRWTIDGYKVPDPDYSSMVVLPTEPAEIIWEKRETCMF

Figure 4: Showing the Consensus and Mutated (K143R) sequences of the ERG11, the gene encoding Fluconazole-resistant *Candida albican*. (uniprot ID:P10613)

Result: As presented in Figure 5, the Common Information Spectrum of the consensus and mutated sequence, which is the product of point-wise multiplication demonstrates amplitude of 4.191 (2.054 x 2.04) at the consensus frequency position 154. Normalization using percentages gives the maximum amplitude of 4.191 as 100%.

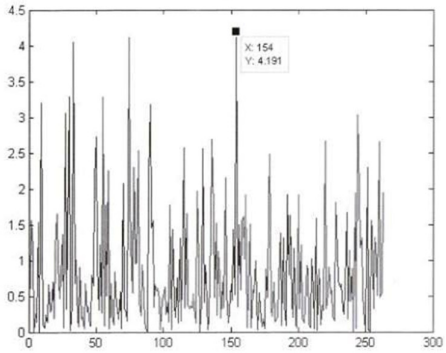

Figure 5: Showing the Common Informational Spectrum (CIS) of the consensus Consensus and mutated (K143R) of the ERG11 protein, demonstrating amplitudes of 4.191 at the Consensus Frequency (CF) position 154

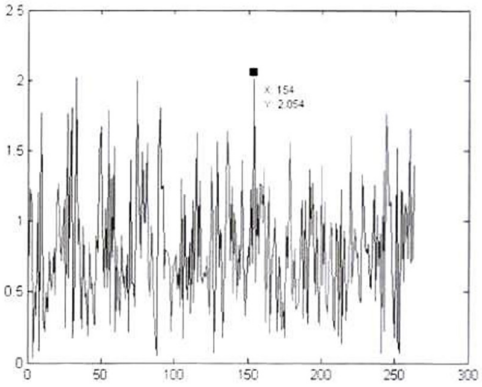

Figure 6A: Showing the Informational Spectrum of the Consensus Sequence with amplitude of 2.054 (100%) at the CF position 154.

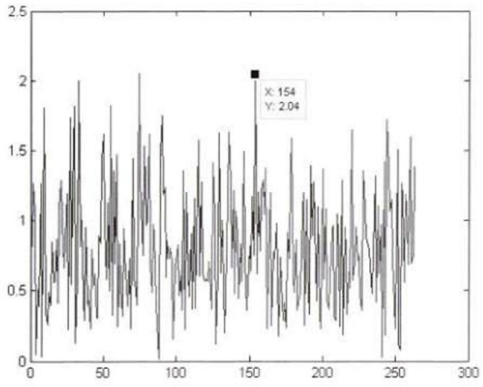

Figure 6B: Showing the IS of the mutated sequence (K143R) with amplitude of 2.04 (99.3%) at the CF position 154, indicating a resistance of 0.7%.

Figure 6A and 6B are the Informational Spectra of the consensus and mutated (K143R) sequences, respectively. The consensus sequence showed amplitude of 2.054 or 100% while the mutated (K143R) displayed amplitude of 2.04 (99.3%) leaving a resistance of 0.7%. Therefore, by means of one mutation (K143R) and one parameter EIIP, the resistance obtained is 0.7%. Total resistance offered to Fluconazole by *Candida albican* can be obtained as preliminarily carried out [1] through aggregation of all resistance derived from all ERG11 mutations and parameters involved.

In this same manner therefore, resistance offered by any target protein to any bio-active substance including antibiotics are therefore calculated using this procedure. This is because, it has been established that all bio-active substances, irrespective of composition (alkaloids, Anthraquinones, Azole derivatives, e.t.c such as Fluconazole) have either protein targets or genes/proteins encoding them [14] and as such they carry as much biological information as the bio-active substances [15].

In essence, where there is need for the drug resistance assessment to be carried out using any microbe, including those listed in the National Action Plan, the procedure will entail first the

extraction of the protein sequence information of the drug target proteins or genes/proteins encoding, from the submitted sample. By engaging Internet-enabled phones, gadgets such as laptop and computers, the sequence is submitted to the centrally-manned CARDC program and ran. The CADRC-based calculated drug resistance will immediately be delivered numerically to the healthcare provider in order to immediately help them with informed decision on antibiotic use for the patient.

Conclusion: Drugs including antibiotics such as the penicillin, e.t.c, which have served as salvage therapy for various diseases have later been rendered ineffective as a result of resistance offered by the targeted organisms. Except for rapid designing and development of new antibiotics, Medical Practitioners would have face terrible time treating several ailments. This is even financially challenging. Another related global menace is the spread of drug resistance. Because of high-rate of cross border activities, spread of drug resistance has become a global concern. For example, by the time the plasmid (mobile gene) of the New Delhi Metallo-B-Lactarnase (NDM) associated with resistance to all known antibiotics was identified, the carrier, a Swedish national of Indian descent who picked up the NDM - encoded klebselia pneumonia in New Delhi was already in Sweden [16].

To help curb these global menaces as initiated by National Action Plan for combating Antibiotics Resistant Bacteria, we submit a rapid point-of-need diagnostic device that computes and numerically display drug resistance offered by any organism to any therapeutic agent. It is called Computer-Aided Drug Resistance Calculator (CADRC, US20150370964). This is an already established device for calculating drug resistance [1] that anchors on a technique called Digital Signal Processing, which has reliably served mankind for years such as in Radar technology.

This highly sensitive, rapid and rational point-of-care diagnostic device that transforms genotypic information to phenotyptic characteristics assesses and presents resistance to all drugs numerically and immediately, and which can centrally be located for quality control monitoring is therefore presented for use in tackling the menace of Antimicrobial Resistance (AMR). The procedure has earlier earned us InnoCentive Award for “Assessing Vaccine Potency” Computationally. A prototype of CADRC can be built in less six months.

Enclosed: (1) Innocentive Challenges: Affidavit (2) Collaboration Contract (3) Patent Publication: Computer Aided Drug Resistance Calculator Calculating Drug Resistance Using Amprenavir as a Case Study (US20150370964).

Reference

[1] Nwankwo N.I, inventor; NWANKWO NORBERTIKECHUKWU, assignee. 2015 “Computer Aided Drug Resistance Calculator Calculating Drug Resistance Using Amprenavir as a Case Study.” United States Patent Publication No: 20150370964 Al.

[2] Nwankwo N. 2012. “Signal processing-based Bioinformatics methods for characterization and identification of Bio-functionalities of proteins”. Openthesis. 123-146.

[3] Nwankwo N, Seker H. 2010. "A signal processing-based bioinformatics approach to assessing drug resistance: human immunodeficiency virus as a case study". Conf. Proc. IEEE Eng Med BiolSoc, 2010:1836-1839.

[4] Tomii K, Kanehisa M. 1996. “Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins," Protein Engineering, 9(1):27-36.

[5] Nwankwo N. 2013. “A Digital Signal Processing-based Bioinformatics Approach to Identifying the Origins of HIV-1 non B subtypes infecting US Army Personnel serving abroad”, Curr HIV Res. 11(4): 271-80.

[6] Hoj L, Kjaer J, Winther O, *et al*. 2008. "In silico identification of physicochemical properties at mutating positions relevant to reducing susceptibility to amprenavir," XVII International HIV Drug Resistance Workshop, Poster No. 113.

[7] Veljkovic V, Cosic I, Dimitrijevic B, *et al*. 1985. “Is it possible to analyze DNA and Protein sequence by the method of Digital Signal Processing”. IEEE Trans Biomed Eng. 32(5):337-341

[8] Veljkovic V, Glisic S, Veljkovic N, *et al*.2014. Influenza vaccine as prevention for cardiovascular diseases: Possible molecular mechanism. Vaccine , http://dx.doi.Org/10.1016/j.vaccine.2014.07.007

[9] W.O.F.A.H (OIE). 2008. “Laboratory Methodologies for Bacterial Antimicrobial Susceptibility Test”.

[10] Jain E, Bairoch A, Duvaud S, *et al*. 2009. “Infrastructure for the life sciences: design and implementation of the UNIPROT website" BMC Bioinformatics. 10:136-154.

[11] Open web sever for information spectrum method-based phylogenetic analysis (ISTREE), available : www.istree.bioprotection.org, assessed 23/11/2016

[12] Xiang M, Liu J, Ni P, *et al*. 2003, “ERG11 mutations associated with azole resistance in clinical isolates of *candida albican”*. FEMS Yeast Res. 13: 386-393

[13] Wang H, Kong F, Sorell T, *et al*. 2009. “Rapid detection of ERG11 gene mutations in clinical *canidida albicans* isolates with reduced susceptibility to fluconazole by rolling circle amplification and DNA sequencing”. BMC microbiology 9: 167

[14] Lin H, Burton D, Li L, *et al*. 2015. “Gain-of-function mutation identify Amino Acid within transmembrance domains of the yeast vacuolar transporter Zrc-1 that determine metal specificity” Biochem J. 422 (2): 273-283

[15] Nwankwo N. 2015. “Direct Computerized translation of biological data into information is now feasible: The gains of Digital signal processing-based Bio-information techniques”. Virol-mycol.4:2

[16] Yong D, Toleman M. A, Giske C. G, *et al*. 2009. “Characterization of a New Metallo-B-lactamase Gene, bla NDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic structure in Klebsiella pneuminae Type 14 from India”. Antimicrob. Agents chemother. 53(12): 5046-5054