Interdiscip Sci Comput Life Sci (2014) 6: 279–284
DOI: 10.1007/s12539-012-0044-x

PDB Explorer – A Web Based Algorithm for Protein Annotation Viewer and 3D Visualization

Anuraj Nayarisseri*, Rakesh Kumar Shardiwal, Mukesh Yadav,
Neha Kanungo, Pooja Singh, Pratik Shah, Sheaza Ahmed
(In silico Research Laboratory, Eminent Biosciences, Indore, M.P., 452010, India)

Received 13 July 2012 / Revised 19 March 2012 / Accepted 11 September 2013

Abstract: The PDB file format, is a text format characterizing the three dimensional structures of macro molecules available in the Protein Data Bank (PDB). Determined protein structure are found in coalition with other molecules or ions such as nucleic acids, water, ions, Drug molecules and so on, which therefore can be described in the PDB format and have been deposited in PDB database. PDB is a machine generated file, it's not human readable format, to read this file we need any computational tool to understand it. The objective of our present study is to develop a free online software for retrieval, visualization and reading of annotation of a protein 3D structure which is available in PDB database. Main aim is to create PDB file in human readable format, i.e., the information in PDB file is converted in readable sentences. It displays all possible information from a PDB file including 3D structure of that file. Programming languages and scripting languages like Perl, CSS, Javascript, Ajax, and HTML have been used for the development of PDB Explorer. The PDB Explorer directly parses the PDB file, calling methods for parsed element secondary structure element, atoms, coordinates etc. PDB Explorer is freely available at http://www.pdbexplorer.eminentbio.com/home with no requirement of log-in.
Key words: PDB explorer, protein 3D structure visualization tools, PDB parser tools, PDB.

1 Background

Protein Data Bank is a public repository for 3 dimensional Structures of Protein and other biological macromolecules [1]. The file format (.pdb) is a textual file format which represents the three dimensional structures of molecules held in the Protein Data Bank. Protein structures are made in x-ray crystallography, NMR spectroscopy and have been deposited in protein data bank. Some protein structures are modeled using theoretical method as well and is made available in protein data bank. The information in protein data bank concern to proteins, and the pdb format accordingly procure for rich description and annotation of protein properties. However, proteins are generally crystallized with other molecules or ions such as ions, water, nucleic acids, drug molecules and so on, which accordingly can be described in the pdb format as well [1, 2].

2 PDB file format

Protein 3D structure data file created in the Protein Data Bank in PDB format. The File represents coordinate, sequence, secondary structure and citation data. The information included in the PDB format are prefixed with a record tag followed by individual items of data. The data in PDB records are classified into Structure PDB records, unstructured PDB records and semi-structured pdb records [2]. A typical PDB file starts with HEADER, TITLE and AUTHOR details [3, 4]. The second set of line includes REMARK Records. Remark records contain free-form annotation, but they also accommodate standardized information. The third set of lines include SEQRES Records which gives the sequences of the three peptide chains like A Chain, B Chain, C Chain etc. Forth set of lines include ATOM Records Describe the coordinates of the atoms that are part of the protein. ATOM line above describes the alpha-N atom of the first residue of peptide chain A; the first three floating point numbers are its x, y and z coordinates and are in units of Angstroms (Å). The next three columns are the occupancy, temperature factor, and the element name, of the molecule respectively [5].

2.1 Scope of PDB explorer

Normal computer applications not usually support to

*Corresponding author.
E-mail: anuraj@eminentbio.com

to read a PDB file. An academic biologist can hardly understand a PDB file as the existing applications are complicated and command line based. Here we introduce PDB Explorer an online server which is a user friendly tool for reading the features of PDB file and visualization of three dimensional structure of any protein which is made available in the Protein Data Bank.

2.2 Technology used

PDB Explorer is developed using CGI Perl with the MVC based architecture. GUI of PDB Explorer is developed in Hyper Text Markup Language [6] and Cascading Style Sheets [7]. User can retrieve any protein 3D structure from PDB explorer by using the appropriate pdb accession number. After retrieving the protein 3D Structure one has to upload in to PDB Explorer to analyse it. PDB Explorer will give the complete annotation of the protein 3D Structure which is made in the .pdb file including the Molecule name, Organism name, Methods used for the modeling, Crystal Structure Resolution, Title of the research, Authors, year of publication etc.

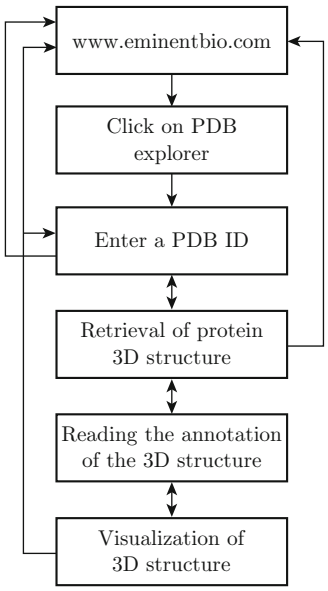


Fig. 1 Web architechture

2.3 Implementation Web architecture of PDB explorer

Model-View-Controller web framework technology has been used to develop PDB-Explorer. The model represents enterprise data and the business rules that govern access to and updates of this data. The view renders the model into a form suitable for interaction, typically a user interface element [8]. The controller receives user input and initiates a response by making calls on model objects. A controller accepts input from the user and instructs the model and a view port to perform actions based on that input [8] [9]. In Model-View-Controller (MVC), the model represents the information (the data) of the application; the view corresponds to elements of the user interface such as text, checkbox items, and so forth; and the controller manages the communication of data and the business rules used to manipulate the data to and from the model [8].

The web interface is developed in HTML and CSS [10] [11]. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists etc as well as for links, quotes, and other items. It allows images and objects to be embedded and can be used to create interactive forms [10]. CSS is a style sheet language used to describe the presentation semantics (that is, the look and formatting) of a document written in a markup language [11].

Here we used Bioperl module to read the the pdb file. A BioPerl module is a reusable package specified in a collection of module libraries whose name is similar to the name of the package with a file extension (.pm) [12]. For reading the annotation of the pdb file we developed a new perl module to parse the pdb file to the html interface and has been called as pdbexplorer.pm.

The Parser file which we developed will read the annotation of any pdb file which the user has uploaded. The parser will read the Header which contains the molecule name, number of amino acids it contains, amino acids chain, the secondary structure information, OBSLTE, Title, split, caveat, compound, source, keywords, Experimental Data, NuMMDL, MDL type, Authors, RevDat, Sparsde, Journal, Remark etc. This will read the above data via HTML and will return to the result page of PDB explorer (Fig. 2). Once the Annotation reading has been done, the PDBExplorer will allow the user to visualize the 3D structure of the particular PDB file using Java Script via JMOL [13]. Jmol is an algorithm being written in Java a very portable and faster application for 3D visualization of any macromolecule. Although Jmol suits to do everything that rasmol does, it parses pdb files disparately. This sometimes results in Jmol being unable to work out any of the atom types in a pdb file which rasmol can read correctly [13].

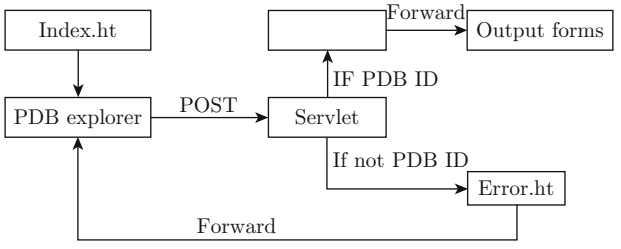


Fig. 2 Class architecture of our PDB explorer

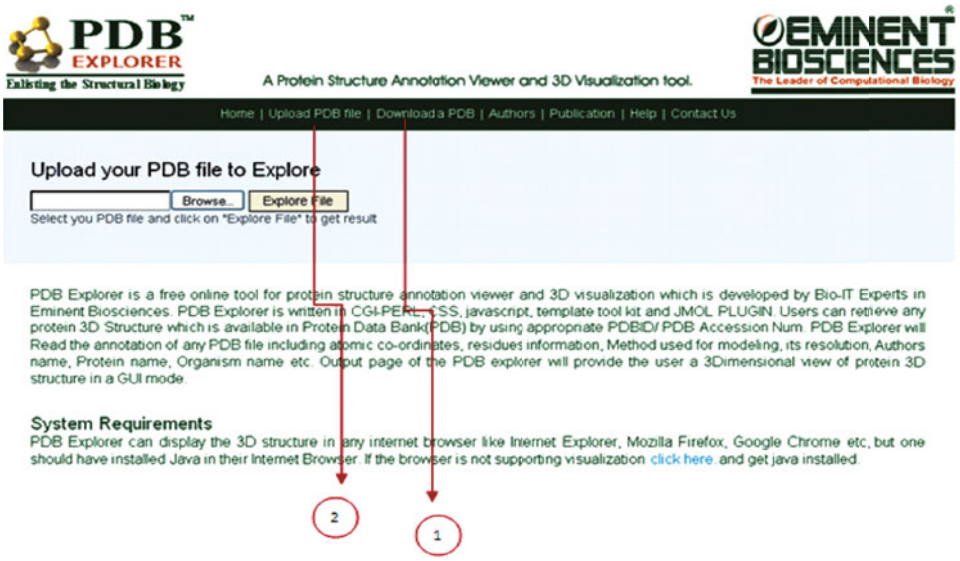


Fig. 3 Webinterface of PDB Explorer. (1. shows how to download a PDB file; 2. Shows show to upload a PDB file to PDB Explorer for Visualization of 3D.)

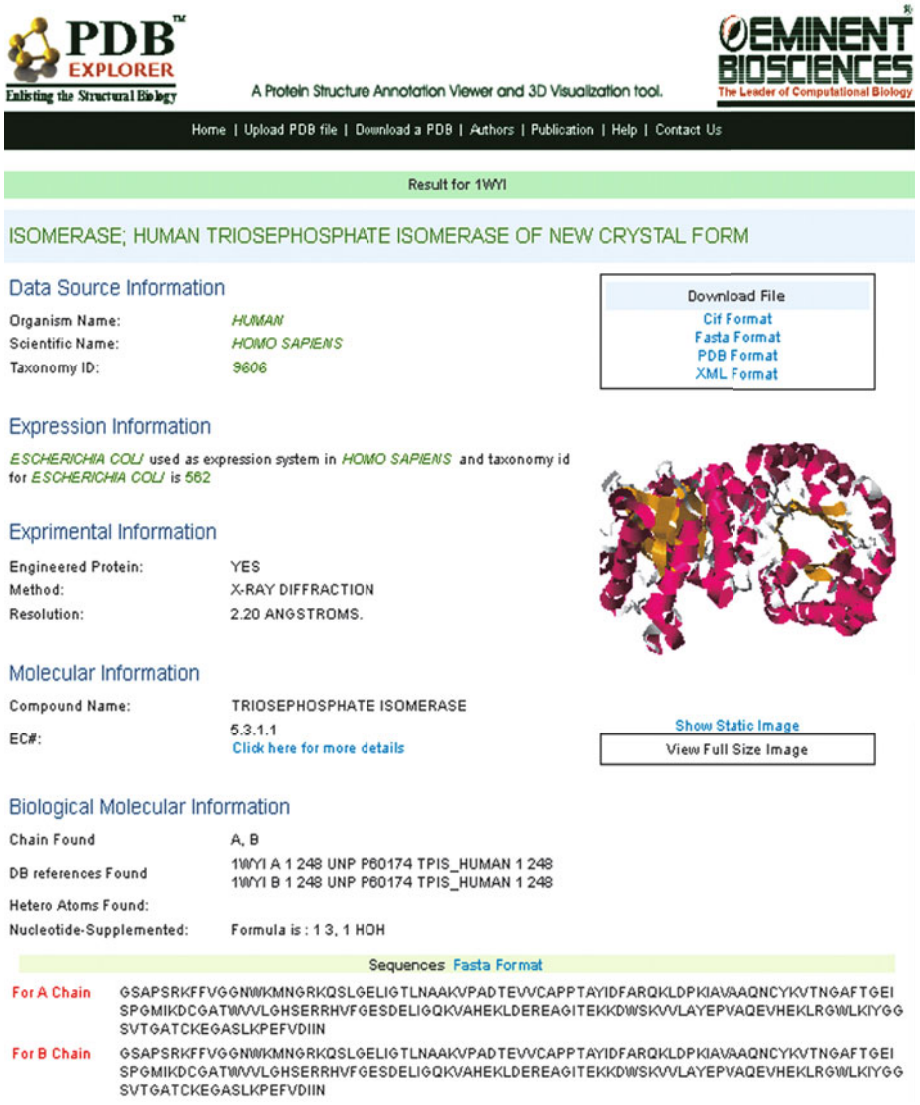
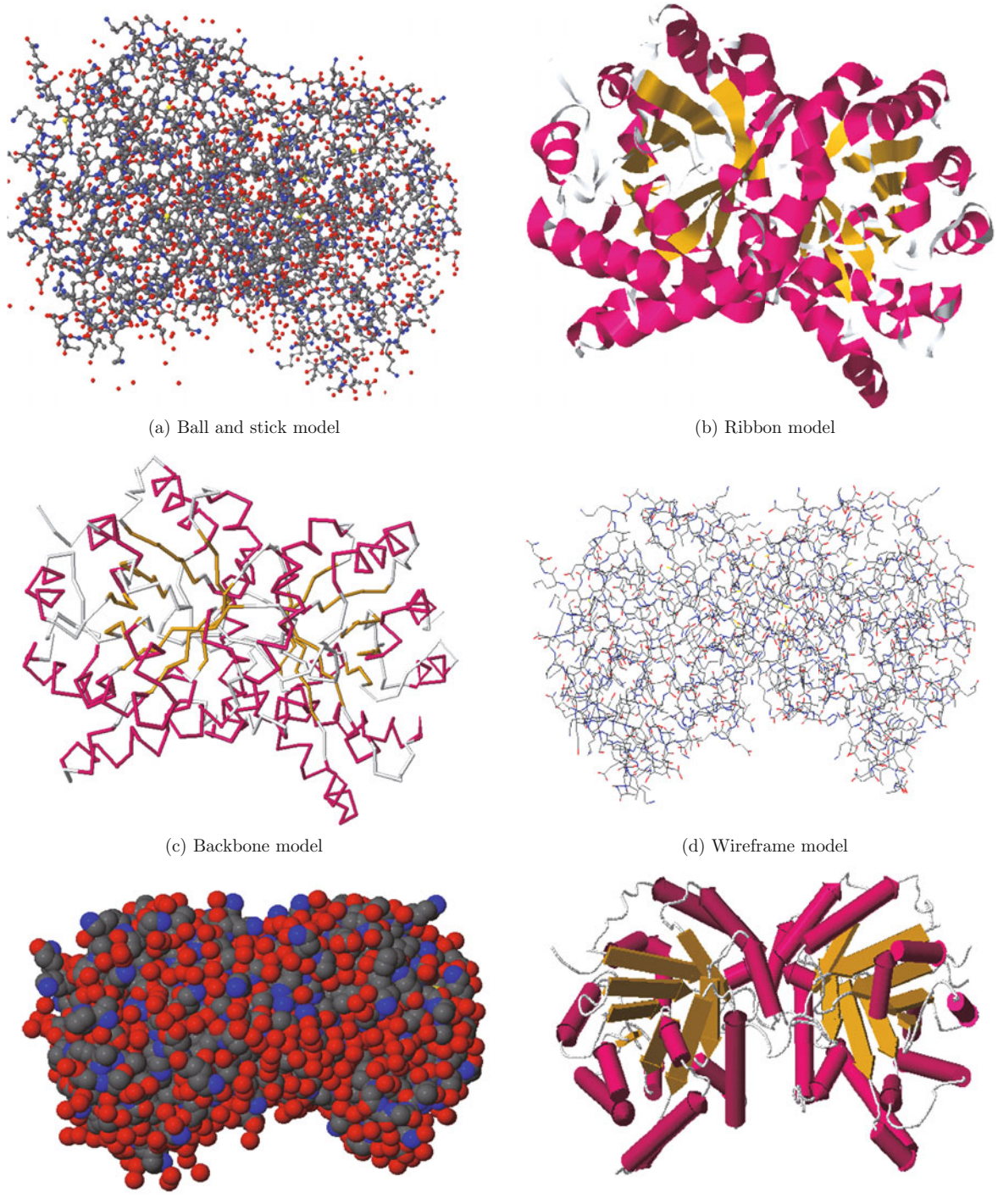


Fig. 4 PDB Explorer Showing the annotation of PDB file 1WYI, Crystal Strcture of Human Triosephoshpate Isomerase.



(a) Ball and stick model

(b) Ribbon model

(c) Backbone model

(d) Wireframe model

(e) Spacefill model

(f) Rocket model

Fig. 5 PDB explorer-visualization of PDB file: 1WYI, human triose phosphate isomerase protein 3D structure in various models.

3 Results and discussions

Here we advert how we represent the molecular visualization from the protein 3D structure file. The user can very easily browse and upload the pdb file to the PDB Explorer. PDB Explorer will highlight in different aspects of the 3D. Each and every mode or aspects of 3D have its own advantage and disadvantages. Ball and sticks model helps to visualize all the atoms and chemical bonds as a direct link between atoms by linking balls as atoms and sticks as bonds [14] [15]. This has been extremely popular in computational chemistry and is still widely used these days. The balls have colors; Black represents carbon(C); Red Oxygen (O); Blue, Nitrogen (N); and white, Hydrogen (H) (Fig. 5(a) and (b)) [14]. The user can understand the secondary na-

ture of the protein using Ribbon Model [16] [17]. The secondary structure is predicted from the amino acids atomic coordinates present in the pdb file. Here the (Fig. 5(c) and (d)) represents the secondary nature of the protein Human triosephosphate isomerase [PDB ID: 1WYI] using ribbon model. This model has been created using hermite curves. For this we used JMOL algorithm (Fig. 5(a) and (b)).

The backbone model will help the user to understand the protein molecule as a chain and will help to visualize the amino acids positions in the chain. This model has been created using alpha carbon atom, nitrogen atom and oxygen atom in the protein Structure [18] (Fig. 5(c) and (d)).

The Spacefill model will help the user to understand the volume of a protein molecule occupies, but it required the information in respect to how the chains are formed and how amino acids are connected each other as well (Fig. 5(e) and (f)) [19]. Spacefill model represent the van der waals model. This Promote to visualize the volume of a protein molecule it occupies. IT gives an overall view of the molecule of a protein in a 3D View. This model is build using van der waals radius hence the viewer can analyse the relative size of the atoms making up the protein molecule (Fig. 5(e) and (f)) [19] [20].

3.1 Performance tests results

We have tested PDB Explorer using Windows, Ubuntu Linux, Red Hat Linux, Mac-OS by various browsers like Internet explorer, Mozilla Firefox, Netscape navigator, Google Chrome, Apple-safari etc. The tools chosen for the tests were Jmol 12.1.15 and Jmol 12.1.16 based on java 3D. We have tested the same with 100s of Protein 3D file and few are listed in table B1. The rendering speed test of PDB Explorer had better performance as compared to the other program. For the visualization performance one has to install any versions of java for web browser.

Table 1 B1-list of PDB files which tested with PDB-explorer

PDB ID	Size	Resolution	Residues	N of Atoms
1WYI	367 KB	2.20	248	3736
2KBC	1.08 MB	NMR	25	668
1I6N	228KB	1.80	278	2180
1P0N	413 KB	2.80	349	4517
1HTI	335 KB	2.80	248	3736
2K1V	1.12 MB	NMR	27	696
2JK2	350 KB	1.70	250	3706
2VOM	680 KB	1.85	250	7372
1I60	225 KB	1.60	278	2180
1K52	149 KB	1.80	72	1122

3.2 Availability & requirements

Project Name: PDB Explorer – A web based algorithm for Protein Annotation viewer and 3D Visualization.

Project home page: http://www.pdbexplorer.eminentbio.com/home

Operating system(s): Platform independent web server.

License: PDB Explorer is License free software.

4 Conclusion

Here we concluded the high performance of PDB Explorer for protein 3D Annotation analysing and protein 3D Visualization. There are many tools available for protein 3D visualization. Some are command line based and some are web based. Many web based tools do not perform well due to conflicts in java 3D. Here we provide the user a high speed user friendly 3D visualization support even if they input huge molecule (pdb file). In order to make PDB Explorer better we are trying to add a lot more features to it.

References

[1] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. Nucleic Acids Res. Jan. 1, 28(1), 235-242.

[2] John Westbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick, Helen M. Berman. 2005. Bioinformatics 21(7), 988-992.

[3] Meyer, E.F. 1997. Protein Sci. Jul., 6(7), 1591-1597.

[4] Westbrook, J., Ito, N., Nakamura, H., Henrick, K., Berman, H.M. 2005. Bioinformatics. Apr. 1. 21(7), 988-992. Epub 2004 Oct 27.

[5] Meyer, E.F. Jr. 1971. Interactive computer display for the three dimensional study of macromolecular structures. Nature Jul 23, 232(5308), 255-257.

[6] Berners-Lee, Tim, Connolly, Daniel (June 1993). IETF IIIR Working Group. Retrieved 18 September 2010.

[7] Meyer, Eric A. Cascading Style Sheets 2.0 Programmer's Reference, McGraw-Hill Osborne Media, ISBN 0-07-213178-0.

[8] Huang, Y.S., Horton, M., Vilhjálmsson, B.J., Seren, U., Meng, D., Meyer, C., Ali Amer, M., Borevitz, J.O., Bergelson, J., Nordborg, M. 2011. Oxford Journal 2011, May 23

[9] Trygve, M.H. Reenskaug. XEROX PARC 1978-1979.

[10] Lynch, P.J., Horton, S.J. 1998. Biocommun 25(2), 20-25.

[11] Brelstaff, G., Moehrs, S., Anedda, P., Tuveri, M., Zanetti, G.J. 2001. Med Internet Res. Jan-Mar 3(1), E8.

[12] Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert,

J.G., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., Birney, E. Oct. 2002. Genome Res. 12(10), 1611-1618.

[13] Herráez, A. Jul. 2006. Biochem Mol Biol Educ. 34(4), 255-261.

[14] Bajaj, C., Goswami, S., Zhang, Q.J. Feb. 2012. Struct Biol. 177(2), 367-381. Epub 2011 Dec 13.

[15] Kuttel, M., Gain, J., Burger, A., Eborn, I.J. Nov. 2006. Mol Graph Model 25(3), 380-388

[16] Morange, M.J. Sep. 2011. Biosci. 36(4), 571-574.

[17] Atkinson, A., Inada, S., Li, J., Tellez, J.O., Yanni, J., Sleiman, R., Allah, E.A., Anderson, R.H., Zhang, H., Boyett, M.R., Dobrzynski, H.J. Nov. 2011. Mol Cell Cardiol. 51(5), 689-701.

[18] Subramani, A., Floudas, C.A. Feb. 2012. J Phys Chem B., 18.

[19] Jitsuhara, Y., Toyoda, T., Itai, T., Yamaguchi, H.J. Nov. 2002. Biochem. 132(5), 803-811.

[20] Jenney, F.E. Jr, Adams, M.W. Jan. 2008. Extremophiles 12(1), 39-50.