Vol 3 Issue 11 Dec 2013

ISSN No : 2230-7850

International Multidisciplinary Research Journal

Indian Streams Research Journal

Welcome to ISRJ

RNI MAHMUL/2011/38595 **ISSN No.2230-7850**

Indian Streams Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial board.Readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.isrj.net

Indian Streams Research Journal
Volume-3, Issue-11, Dec-2013
ISSN 2230-7850

Available online at www.isrj.net

Structure Prediction Of Neuronal Per- Arnt- Sim (b-hlh-pas) Protein- Npas4 Using Insightii And Modeller 9.11, A Homology Based Approach And A Comparative Study Of The Results.

Akhil Jobby , Satish Chandra Verma And Gess Thoms Xavier
R&D Rasa Life Science Informatics (Rasa Life Science Informatics),
Center for Bioinformatics (Saarland University),
Department of Bioinformatics (SRM University).

Abs tract:-*The neuronal pas domain protein 4 (npas4) is a neuronal activity dependant gene that has recently been identified as a transcription factor which regulates the transcription of genes that control inhibitory synapse development and synaptic plasticity. The role of npas4 is in learning and memory, however 3D structure of npas4 is unknown . The prediction of protein structure from purely sequenced data and improvement of homology modelling tools has benefited in modelling novel proteins. Hence, this approach was used to model the structure of unmodelled protein npas4. Here the protein structure was built manually, following each steps from alignment to loop modelling to refinement using insight II. Another homology modeling tool, Modeller 9.11 was used to compare the results from Insight II. It was concluded that model from Modeller 9.11 produced a better model when compared to Insight II due to the manual approach in Insight II when compared to automatic modelling in Modeller 9.11.*

Keyw ords:NPAS4, homology modelling, structurally Conserved Region (SCR), Insight II, Modeller, Most favourable region, PDBSum Generate.

REVIEW OF LITERATURE

Npas4 is a brain-specific transcription factor and may have a neuroprotective function (Ooe et al. 2009). It has been reported that Npas4 regulates the expression of drebrin, which engages in dendritic-cytoskeleton modulation at synapses in the hippocampus (Ooe et al. 2004). Recently, Npas4 has been shown to control GABAergic synapse development through the transcriptional regulation of BDNF (Lin et al. 2008). However, it remains to be determined whether Npas4 gene expression in the hippocampus is affected by deleterious stress. Npas4 may play a role in synaptic and structural plasticity as well as hippocampal function in CA3 and CA1 regions because this transcription factor has been shown to regulate GABAergic synapse development in an activity-dependent manner (Lin et al. 2008) and to possibly play a role in dendriticcytoskeleton modulation at synapses in the hippocampus (Ooe et al. 2004, 2009).

1. INTRODUCTION

1.1 NPAS4

Neuronal PAS Domain Protein 4 (NPAS4) falls under the class of transcription regulators that are involved in many physiological and developmental functions such as cardiac rhythm control, genetic response to some chemicals (xenobiotics), hypothalamus formation in the fetal brain etcetera and has a basic helix-loop-helix-PER-ARNT-SIM (bHLH-PAS) structure. It is also known as HLH-PAS Transcription Factor NXF or NXF and is very active in the hippocampus region of the brain. reported that the transcription factor NPAS4 plays a role in the development of inhibitory synapses by regulating the expression of activity-dependent genes, which in turn control the number of GABA-releasing synapses that form on excitatory neurons. Activity-dependent transcription factor Npas4 regulates a transcriptional program in the CA3 region of the hippocampus that is essential for contextual memory formation and selective deletion of Npas4 in CA3 both resulted in impaired contextual memory.

1.2 Insight II

Insight II is a graphical molecular modelling program that was developed by Accelrys Inc. (formerly a product of MSI and BIOSYM). Along with a molecular mechanics/dynamics program Discover, Insight II can be used to build and manipulate any class of molecule or molecular system using insilico methods. In conjunction with other Accelrys products, we can study molecular properties. Insight II version 2005, runs on both IRIX and Linux Platforms.

1.2.1 Insight II Modules

There are several modules such as the following; 1. Builder and Biopolymer (molecular building) 2. Delphi

(electrostatic calculation) 3. Solvation (solvent effect) 4. Search and Compare (conformation analysis) 5. Homology (sequence comparison)6. Decipher and Analysis (numerical analysis and Plotting tools) 7. Ludi and Affinity (drug design) 8. NMR refined (NMR noe studies) 9. Discover, Discover3, Charmm (MD simulation).

3 MODULE: HOMOLOGY

'Homology' is a module in Insight II that helps to predict a 3D structure of a protein when its amino acid sequence and the complete atomic structure of at least one reference protein are known. A protein that is structurally homologous to the model protein is used as the reference or the template protein and can be imported into Insight II in any format. Structural segments that are conserved regions in the homologous proteins, are taken directly from the reference protein. Homology can be used when more than one reference sequences is available where 3D structures of the protein can be directly imbibed into the model protein rather than going for sequence similarity as they are expected to have the same structural conformation when there is a conserved conformation in the reference protein. 'Homology' is a collection of tools that that can be used in the process of building a protein by homology based techniques.

1.3.1 Homology Model Building

The process of building a model for a protein using Homology, is divided into the following steps:
1. Determining which proteins are related to the model protein.
2. Determining the structurally conserved regions (SCRs).
3. Aligning the amino acid sequence of the unknown protein with those of the reference protein(s) within the SCRs.
4. Assigning coordinates in the conserved regions.
5. Predicting conformations for the rest of the peptide chain, including loops between the SCRs and possibly the N- and C-terminal.
6. Searching for the optimum side chain conformations for residues that differ from those in the reference proteins.
7. Using energy minimization and molecular dynamics to re?ne the molecular structure so that steric strain introduced during the model-building process can be relieved.

1.4 Modeller 9.11

Modeller is used for homology or comparative modeling of protein three-dimensional structures. the user provides an alignment of a sequence to be modeled with known related structures and modeller automatically calculates a model containing all non-hydrogen atoms. modeller implements comparative protein structure modeling by satisfaction of spatial restraints, and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.

2. METHODOLOGY

2.1 MODULE HOMOLOGY

The Homology Module is used to build a 3D model of a protein based on the 3D structure or sequences of one or more homologous proteins. The protein with the undetermined structure is called the "model", "unknown", or "sequence" protein. The protein with known 3D structures is referred to as the "reference" or "template" protein. The template 2k4m_A with a ~25% query coverage and ~35% identity was obtained after running a "Blast" on the model protein sequence.

Template protein sequence:

>gi|196049613|pdb|2K4M|A Chain A, Solution Nmr Structure Of M. Thermoautotrophicum Protein Mth_1000, Northeast Structural Genomics Consortium Target Tr8
MGSSHHHHHHSSGLVPRGSHMWNDLAVYIIRCSGP
GTRVVEVGAGRFLYVSDYIRKHSKVDLVLTDIKPSH
GGIVRDDITSPRMEIYRGAALIYSIRPPAEIHSSLMRV
ADAVGARLIIKPLTGEDIVTERKMKLVNYGRTYFYE
YIAEVRSR

Model protein sequence:

>gi|219520599|gb|AAI43631.1| NPAS4 protein [Homo sapiens]
MYRSTKGASKARRDQINAEIRNLKELLPLAEADKVR
LSYLHIMSLACIYTRKGVFFAGGTPLAGPTGLLSAQE
LEDIVAALPGFLLVFTAEGKLLYLSESVSEHLGHSMV
DLVAQGDSIYDIIDPADHLTVRQQLTLPSALDTDRLF
RCRFNTSKSLRRQSAGNKLVLIRGRFHAHPPGAYWA
GNPVFTAFCAPLEPRPRPGPGPGPGPASLFLAMFQSR
HAKDLALLDISESG

2.2 STEPS INVOLVED IN HOMOLOGY MODELING

2.2.1 Step 1: Invoking Homology Module from Insight II (Reading and Aligning Sequence)

The template protein, PDB file (2k4m) was imported and its sequence was extracted, then the model protein sequence was also imported. The sequences were then aligned separetely using pairwise alignment using the Clustal W to obtain a better alignment score than the default pairwise alignment tool in Insight II and to have a better understanding of SCRs. The sequences in the sequence window were manually aligned using the sequence mode.

2.2.2 Step 2: Creating Boxes for Conserved Regions

Boxes were created on the completely conserved regions in the sequences where two character were excluded from each side of the box before the gaps that were inserted while aligning, for assigning coordinates and building loops.

2.2.3 Step 3: Assigning Coordinates to the Unknown Protein

The boxes were drawn around the regions of two sequences as gaps cannot be included while freezing the box. Each box were assigned coordinates and bumps were fixed.

2.2.4 Step 4: Building a Loop

The start and the stop residues within the boxes were specified to generate loops. The best loop was then

selected from a list of loops with the lowest RMS alignment of loop-to-Start Residue and loop-to-Stop Residue interface and then coordinated were assigned to loops.

2.2.5 Step 5: Invoking Module Discover_3

The module "Discover_3" was invoked to perform energy minimization steps. This step was used to regularize the bonds and angle geometry.

2.2.6 Step 6: Invoking Module Builder

The structure was then optimized using the "optimize" menu in the module "Builder" to provide a quick route to minimizing the molecule within the module without moving to the "Discover_3" module and without setting up a new minimization stage when the previous minimization stage was complete. The obtained structure was then exported in PDB file format.

2.3 Modeller 9.11

The sequence file was first saved as ".ali" alignment file so that modeller can perform sequence alignment. The template that was selected from blast (2k4m_A) was used to align the model and the template protein sequence. Models were built using the alignment that was obtained. Out of the models generated, model with maximum GA and minimum Dope Score was selected as the result.

3. RESULTS AND DISCUSSION

3.1 Results

3D models obtained from the tools are displayed below. Insight II has the option to import the modelled protein in multiple coordinate file formats. In this study it was imported in PDB format to better understand the structural attributes and to have the comparison performed

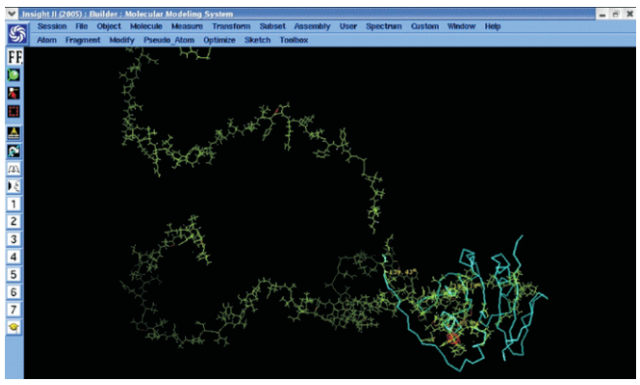


Fig 1: 3D model of npas4 modelled using Insight II (viewed using insight II)

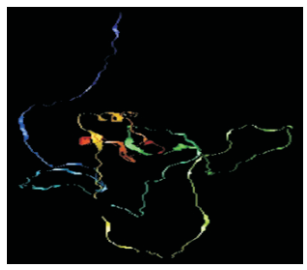


Fig 2: 3D model of npas4 modelled using Insight II (ribbon display)

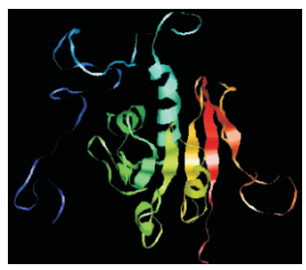


Fig 3: 3D model of npas4 modelled using Modeller 9.11 (ribbon display)

NPAS4	Structural Attributes				
Modelling Tool	Number of Residues	Number of Atoms	Number of Bonds	Helices	Sheets
Insight II	234	3614	3647	60	136
Modeller 9.11	234	1797	1839	227	238

Table 1: Structural attributes of Insight II and Modeller 9.11 for NPAS4

As seen in Fig 1 the structure has two tail like structures (regions that are not properly modelled) with a cluster formation at its juncture. The structures modelled by Insight II and Modeller showed considerable difference (visible in the model, Fig 2 and Fig 3). Fig1shows more loosely modelled structure but Fig2 shows the model a more precise form.

The number of residues in the model were 234 but the number of atoms modeller by Insight II and Modeller 9.11 varies, 3614 and 1797 repectively, whereas the number of bonds modelled also differed, 3647 and 1839 respectively. Helices modelled were 60 and 227 by Insight II and Modeller 9.11 respectively

3.1.1 Structure Validation

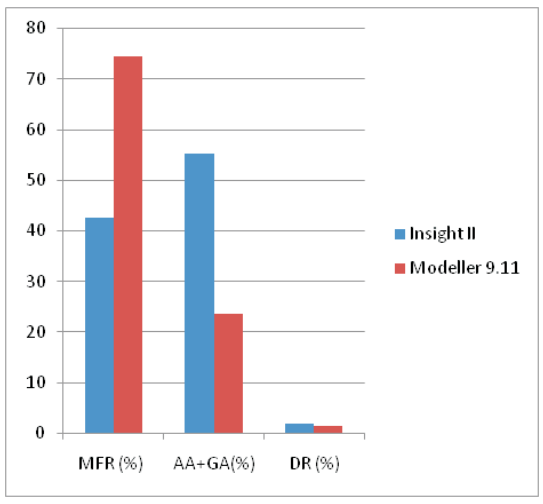
Models from Insight II and modeller were checked for structural validity with PDBSum Generate server. The following table displays the results obtained from PDBSum Generate.

NPAS4	PDBSum Generate - Procheck					
Modelling Tool	Most Favoured Region (MFR)	Most Favoured Region (%)- PDBSum Generate Plot (MFR %)	AA+GA	AA+GA(%)	Disallowed Region (DR)	Disallowed Region (%)- PDBSum Generate Plot (DR %)
Insight II	84	42.6	1.9	55.4	4	2
Modeller 9.11	147	74.6	47	23.8	3	1.5

Table 2: PDBSum Generate (Procheck) results of npas4 (AA: Additionally allowed region, GA: Generously allowed region)

3.2 Discussion

Table 1 and table 2 illustrates the attributes of the results from two different modelling tools used here. Table 1 displays structural features of the modelled npas4 structure and from it is evident that Modeller modelled three times more helices than Insight II whereas the number of atoms that were modelled remains maximum with Insight II. In the structure, helices have more importance as npas4 belongs to the basic helix-loop-helix class. Moreover, table 2 shows Ramchandran Plot details from the Procheck results of PDBSum Generate server for npas4 (modelled using Insight II and Modeller 9.11) and it is evident from the table that the plot for npas4 (Modeller) has maximum residues in MFR and minimum residues in DR which proves it to a better model than the model generated by Insight II. Graph1 plotted below explains the quality of the model. (74% in Most Favored Region- MFR and 3% DR in Modeller 9.11 whereas 42.6% in Most Favored Region- MFR and 4% in DR for Insight II)



Graph 1: Plot showing comparison between Procheck results of model for npas4 built by Insight II and Modeller 9.11.
MFR- Most favored region AA+GA- Additionally allowed region+ generously allowed.region DR- Disallowed region

4. CONCLUSION

The study conducted concludes that a better model was predicted using modeller 9.11 due to fact that better models can be generated from automatic modelling in Modeller 9.11 than from a manual approach in Insight. It is needless to mention that homology modelling using Insight II can yield impressive result if all the steps are followed minutely flawlessly with the usage of multiple templates. The limitations of this study is that a single template was used while modelling and in future multiple templates with more sequence coverage can be used to model better structures.

5. REFERENCE

1. Insight, I. I. (1993). User Guide, version 2.3. 0. San Diego: Biosym Technologies.
2. Insight, I. I. (2000). Accelrys Inc. San Diego, CA.
3. Ooe N., Saito K., Mikami N., Nakatuka I. and Kaneko H. (2004) Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive role in dendritic-cytoskeleton modulator Drebrin gene expression. Mol. Cell. Biol. 24(2), 608– 616.
4. Ooe N., Motonaga K., Kobayashi K., Saito K. and Kaneko H. (2009) Functional characterization of basic helix-loop-helix-PAS type transcription factor NXF in vivo: putative involvement in an "on demand" neuroprotection system. J. Biol. Chem. 284(2), 1057– 1063.
5. Yun, J., Koike, H., Ibi, D., Toth, E., Mizoguchi, H., Nitta, A., ... & Yamada, K. (2010). Chronic restraint stress impairs neurogenesis and hippocampus-dependent fear memory in mice: possible involvement of a brain-specific transcription factor Npas4. Journal of neurochemistry, 114(6), 1840-1851. 2.
6. Lin, Y., Bloodgood, B. L., Hauser, J. L., Lapan, A. D., Koon, A. C., Kim, T. K., ... & Greenberg, M. E. (2008). Activity-dependent regulation of inhibitory synapse development by Npas4. Nature, 455(7217), 1198-1204.
7. K. Ramamoorthi, R. Fropf, G. M. Belfort, H. L. Fitzmaurice, R. M. McKinney, R. L. Neve, T. Otto, Y. Lin. Npas4 Regulates a Transcriptional Program in CA3 Required for Contextual Memory Formation. Science, 2011; 334 (6063): 1669
8. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.
9. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
10. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

Akhil Jobb
R&D Rasa Life Science Informatics (Rasa Life Science Informatics),