BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 10(2)

Mapping the amino acid properties of constituent nucleoporins onto the yeast nuclear pore complex

Ashok Kunda[1,#], Narayanan H Krishnan[1,#] & VV Krishnan[2, 3 *]

[1]Davis Senior High School, 315 West 14th Street, Davis, CA 95616; [2]Department of Chemistry, California State University, Fresno CA 93740 and; [3]Department of Medical Pathology and Laboratory Medicine, University of California Davis School of Medicine, Davis, CA 95616; VV Krishnan – Email: krish@csufresno.edu; vvkrishnan@ucdavis.edu; *Corresponding author
#Contributed equally

Received November 26, 2013; Revised January 04, 2014; Accepted January 06, 2014; Published February 19, 2014

Abstract:
Visualization of molecular structures aids in the understanding of structural and functional roles of biological macromolecules. Macromolecular transport between the cell nucleus and cytoplasm is facilitated by the nuclear pore complex (NPC). The ring structure of the NPC is large and contains several distinct proteins (nucleoporins) which function as a selective gate for the passage of certain molecules into and out of the nucleus. In this note we demonstrate the utility of a python code that allows direct mapping of the physiochemical properties of the constituent nucleoporins on the scaffold of the yeast NPC's cytoplasmic view. We expect this tool to be useful for researchers to visualize the NPC based on their physiochemical properties and how it alters when specific mutations are introduced in one or more of the nucleoporins. The code developed using Python is available freely from the authors.

Key words: NPC (nuclear pore complex), IDP (intrinsically disordered protein), AA Index.

Background:
A unique feature of the eukaryotic cell compared to the prokaryotic cell is the compartmentalization of nucleic acid synthesis and processing within the nucleus, and their separation from protein synthesis in the cytoplasm. This separation is achieved by the double-membrane of the nuclear envelope and the selective transport channels of nuclear pore complexes (NPCs) embedded within this barrier **[1-3].** NPCs are the sole gateways that facilitate macromolecular exchange across the nuclear envelope between the nucleoplasm and cytoplasm. NPCs are composed of individual proteins called nucleoporins (nups) and are characterized by a central framework with 8-fold rotational symmetry with additional filamentous structures extending from the nuclear and cytoplasmic facades **[4-7].** Intrinsically disordered /unstructured proteins (IDPs) exist without a stable three-dimensional (3D) structure as highly flexible conformational ensembles **[8,9].** These proteins seem to be more abundant in both prokaryotic and eukaryotic proteomes than originally assumed **[10-12].** IDPs are multifunctional in nature and involved in various biological processes such as signaling and regulation **[9,13,14].** Nucleoporins, the protein components of the central conduit play a predominant role in nuclear transport. The permeability barrier in the yeast NPC is formed by a number of conserved nucleoporins that are known to be intrinsically disordered. Being one of the largest and most complex macromolecular assemblies known in cells, the NPC presents a formidable challenge for structural determination. NPC structure has been studied for over five decades by various techniques such as transmission electron microscopy **[15],** cryoelectron tomography of whole NPCs and X-ray crystallography of individual nucleoporins and their complexes **[16-18]**. Cytoplasmic cross-sectional view of the NPC is generally obtained by cryoelectron tomographic reconstruction **[17].** Other structural views are also generated by a combination of experimental and computational data that are invaluable in understanding the architecture, topology and organization of a NPC **[19]**. A molecule transported through

the NPC from the cytoplasmic side into the nucleus does not recognize a geometric representation of the NPC, but more from physiochemical representation of the proteins that constitute the NPC. In order to facilitate an improved visualization of the NPC in terms of the primary characteristics of the constituent nucleoporins, we have developed a visualization tool. A straightforward code written in Python using PyGame and BioPython libraries provides researchers a new tool to visualize the NPC, which is color-coded with respect to a specific property of choice (e.g., hydrophobicity or charge). This tool can also be used to see how mutations in one or more of the proteins can alter the overall organization

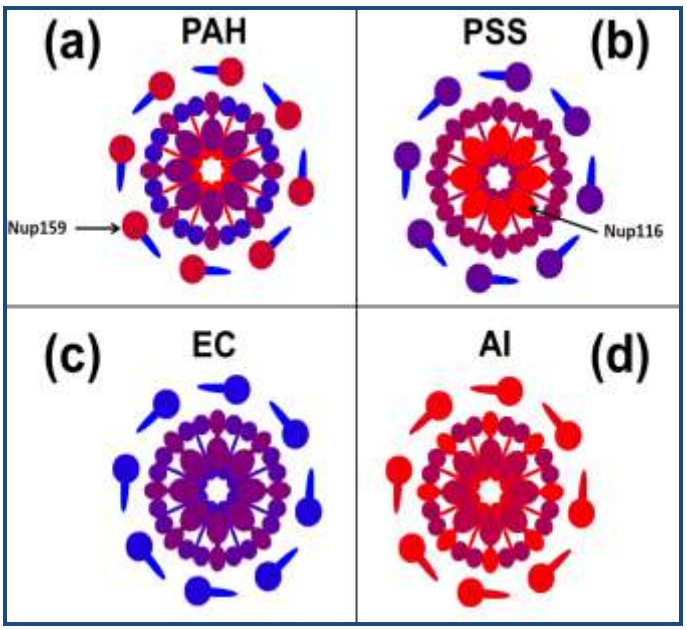

Figure 1: Direct mapping of the selected set of AA indices on the cytoplasmic view of yeast NPC. PAH: Polarity, Accessibility, Hydrophobicity; PSS: Propensity for Secondary Structure EC: Electrostatic Charge and AI: Amphiphilicity index. The index values are listed in **Table 1 (see supplementary material)**. The average values of each nucleoporin was calculated and transformed into a linear scale between red (high) and blue (low).

Methodology:

The geometric organization of the nucleoporins in the NPC from the cytoplasmic view is obtained from the structural model developed by Hoelz *et al* **[17]** and the forest model by Yamada *et a*l **[19].** Amino acid sequences of the constituent nucleoporins are obtained from (http://www.uniprot.org/). All the sequences are used in the standard FASTA format and are used without any additional changes. The size/shape of each protein is scaled proportionally to its expected hydrodynamic volume to fit the symmetry of the NPC. The primary sequence information of the amino acids was converted into physiochemical/biochemcial value or the linear combinations thereof to define a score for each of the 20 naturally occurring amino acids. To date, there are more than 500 different properties of amino acids, all of which have been investigated through a large number of experiments and theoretical studies (http://www.genome.jp/aaindex/) **[20].** This collection of AA indices has redundant information, such as the same physical/chemical variation might be represented by more than one AA index. To address this issue using multivariate statistics, Atchely *et a*l **[21]** defined a linear combination of the complete set, to a set of 5 unique parameters. Of these five factors, we have used the ones that have shown significant differences between the nucleoporins: Factor1 (PAH): Polarity, Accessibility, Hydrophobicity; Factor2 (PSS): Propensity for Secondary Structure; and Factor5 (EC): Electrostatic Charge. In addition, we have also included the amphiphilicity index from one of the AA indices (MITS020101) **[20]. Table 1 (see supplementary material)** lists these indices. The code was written in Python (version 2.71) with PyGame (version 1.9.1) and BioPython (version 1.61) libraries on a Microsoft Windows 7 desktop computer. A copy of the code and the instructions to run on a windows system are available from the authors.

Results & Discussion:

Figure 1 shows the mapping of the amino acid indices PAH (Polarity, Accessibility, Hydrophobicity), PSS (Propensity for Secondary Structure), EC (Electrostatic Charge) and AI (amphiphilicity index) in panels (a), (b), (c) and (d), respectively. PAH, PSS, and EC are cumulative indices derived from multiple physiochemical properties (Table 1). The color scales in **Figure 1** were determined linearly from red (high) to blue (low) and represented using the RGB (Red:Green:Blue) color model (with Green set to zero). The mapping of PAH on the NPC **(Figure 1a)** emphasizes the maximum differentiation in PAH by scaling the colors between the proteins accordingly. The outermost nucleoporin (Nup159) is differentiated from the folded (red) and the charged/unstructured (blue) regions. Similarly, the innermost part of the NPC is also highly differentiated with respect to the PAH scale (classified as a tree in the bimodal distribution model **[19]).** According to the current mapping of propensity to form secondary structures, (PSS, **Figure 1b**) most of the proteins show low propensities except the C-terminal part of Nup116 at the center of the NPC. Panels (c) and (d) demonstrate the variation in the charge factor and the amphiphilicity index across the NPC. Though panels (c) and (d) do not show much differentiation between the nucleoporins as in panels (a) and (b), several features are noted including that of Nup159 at the outermost ring of the NPC. The actual structure of the NPC is an evolutionarily conserved set of ~30 different proteins and is organized into several sub-complexes, each of which occur in multiple copies, resulting in ~500–1,000 protein molecules in the fully assembled NPC. Our results only use a representative set of these proteins from the cytoplasmic view and a first demonstration suggests that direct mapping of the physiochemical properties of amino acids could prove to be a useful tool in understanding the complex structures. The combination of PyGame and other high-level graphics programming can be utilized to build a complex visualization tool for the NPC using the principles demonstrated in this note.

The NPC is flexible and highly dynamic as one would expect during cargo translocation. Any dynamic information determined experimentally can easily be mapped onto the NPC by defining an amino acid/protein specific scale. The intrinsically disordered nature of the nucleoporins is an additional advantage of this model: it is not necessary to identify an orientation of a specific surface of a nucleoporin

with respect to the NPC due to its high flexibility. The permeability barrier of the NPC is formed by unstructured phenylalanine-glycine (FG)-repeat regions that also serve as docking sites for transport factor-cargo complexes. Therefore, mapping the average properties of the nucleoporins with respect to any amino acid scale may be considered justifiable. Several emerging techniques such as single molecule spectroscopic methods are now capable of studying a single NPC **[22, 23];** yet determining the complete 3D-structure with atomic resolution is difficult. While the precise nature of the permeability barrier remains to be an area of heavy debate, we hope that new tools such as the one presented here may prove useful in understanding the functional architecture of NPC.

Acknowledgement:
We thank Korey Reid for critical reading of the manuscript. VVK was in part supported by NIH grant P20 MD 002732 (Research Infrastructure for Minority Institution).

References:

[1] Gerace L & Burke B, *Annu Rev Cell Biol*. 1988 **4**: 335 [PMID: 2461721]
[2] Goldberg MW & Allen TD, *Curr Opin Cell Biol*. 1995 **7**: 301 [PMID: 7662358]
[3] Onischenko E & Weis K, *Curr Opin Cell Biol*. 2011 **23**: 293 [PMID: 21296566]
[4] Fahrenkrog B *et al*. *Trends Biochem Sci*. 2004 **29**: 175 [PMID: 15082311]
[5] Fernandez-Martinez J & Rout MP, *Curr Opin Cell Biol*. 2009 **21**: 603 [PMID: 19524430]
[6] Harel A & Gruenbaum Y, *Cell* 2011 **146**: 191 [PMID: 21784242]
[7] Vasu SK & Forbes DJ, *Curr Opin Cell Biol*. 2001 **13**: 363 [PMID: 11343909]
[8] He B *et al*. *Cell Res*. 2009 **19:** 929 [PMID: 19597536]
[9] Uversky VN *et al*. *Annu Rev Biophys*. 2008 **37**: 215 [PMID: 18573080]
[10] Csizmok V *et al*. *Curr Protein Pept Sci*. 2007 **8**: 173 [PMID: 17430198]
[11] Tompa P *et al*. *J Proteome Res. 2006* **5:** 1996 [PMID: 16889422]
[12] Galea CA *et al*. *J Proteome Res*. 2006. **5**: 2839
[13] Galea CA *et al*. *J Proteome Res*. 2006 **5:** 2839 [PMID: 17022655]
[14] Dunker AK *et al*. *BMC Genomics*. 2008 **9:** S1 [PMID: 18831774]
[15] Lim RY *et al*. *Histochem Cell Biol*. 2008 **129**: 105 [PMID: 18228033]
[16] Brohawn SG *et al*. *Structure*. 2009 **17:** 1156 [PMID: 19748337]
[17] Hoelz A *et al*. *Annu Rev Biochem*. 2011 **80**: 613 [PMID: 21495847]
[18] Kampmann M & Blobel G, *Nat Struct Mol Biol*. 2009 **16**: 782 [PMID: 2706296]
[19] Yamada J *et al*. *Mol Cell Proteomics*. 2010 **9:** 2205 [PMID: 2953916]
[20] Kawashima S *et al*. *Nucleic Acids Res*. 2008 **36**: D202 [PMID: 17998252]
[21] Atchley WR *et al*. *Proc Natl Acad Sci U S A*. 2005 **102**: 6395 [PMID: 15851683]
[22] Herrmann M *et al*. *Nano Lett*. 2009 **9**: 3330 [PMID: 19591452]
[23] Lim RY *et al*. *J Struct Biol*. 2007 **159**: 277 [PMID: 17446086]

Edited by P Kangueane
Citation: **Kunda *et al.*** Bioinformation 10(2): 094-097 (2014)

BIOINFORMATION

open access

Supplementary material:

Table 1: List AA indices mapped

Amino Acid[1]	PAH[2]	PSS[2]	EC[2]	AI[3]
A	-0.591	-1.302	-0.146	0
C	-1.343	0.465	-0.255	0
D	1.05	0.302	-3.242	0
E	1.357	-1.453	-0.837	1.27
F	-1.006	-0.59	0.412	0
G	-0.384	1.652	2.064	0
H	0.336	-0.417	-0.078	1.45
I	-1.239	-0.547	0.816	0
K	1.831	-0.561	1.648	3.67
L	-1.019	-0.987	-0.912	0
M	-0.663	-1.524	1.212	0
N	0.945	0.828	0.933	0
P	0.189	2.081	-1.392	0
Q	0.931	-0.179	-1.853	1.25
R	1.538	-0.055	2.897	2.45
S	-0.228	1.399	-2.647	0
T	-0.032	0.326	1.313	0
V	-1.337	-0.279	-1.262	0
W	-0.595	0.009	-0.184	6.93
Y	0.26	0.83	1.512	5.06

[1]Amino acids are listed in single letter codes; [2]Polarity, Accessibility, Hydrophobicity (PAH), Propensity for Secondary Structure (PSS) and Electrostatic Charge (EC) are derived by Atchley *et al.* [21]; [3]Amphiphilicity index is one of the AA indices (MITS020101)[20].