BioSystems 127 (2015) 47–59

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Structural analysis of *in silico* mutant experiments of human inner-kinetochore structure

Richard Henze[a,1], Jan Huwald[a,1], Nelly Mostajo[a], Peter Dittrich[a,*], Bashar Ibrahim[a,b,c,**]

[a] Bio Systems Analysis Group, Institute of Computer Science, Jena Centre for Bioinformatics and Friedrich Schiller University Jena, 07743 Jena, Germany
[b] Umm Al-Qura University, 1109 Makkah Al-Mukarramah, Saudi Arabia
[c] Al-Qunfudah Center for Scientific Research (QCSR), 21912 Al-Qunfudah, Saudi Arabia

ARTICLE INFO

Article history:
Received 26 March 2014
Received in revised form 20 August 2014
Accepted 4 November 2014
Available online 7 November 2014

Keywords:
Rule-based
Modeling
Simulation
Inner-kinetochore structure
Bond mutation
Structural analysis

ABSTRACT

Large multi-molecular complexes like the kinetochore are lacking of suitable methods to determine their spatial structure. Here, we use and evaluate a novel modeling approach that combines rule-bases reaction network models with spatial molecular geometries. In particular, we introduce a method that allows to study *in silico* the influence of single interactions (*e.g.* bonds) on the spatial organization of large multi-molecular complexes and apply this method to an extended model of the human inner-kinetochore. Our computational analysis method encompasses determination of bond frequency, geometrical distances, statistical moments, and inter-dependencies between bonds using mutual information. For the analysis we have extend our previously reported human inner-kinetochore model by adding 13 new protein interactions and three protein geometry details. The model is validated by comparing the results of *in silico* with reported *in vitro* single protein deletion experiments. Our studies revealed that most simulations mimic the *in vitro* behavior of the kinetochore complex as expected. To identify the most important bonds in this model, we have created 39 mutants *in silico* by selectively disabling single protein interactions. In a total of 11,800 simulation runs we have compared the resulting structures to the wild-type. In particular, this allowed us to identify the interaction Cenp-W-H3 and Cenp-S-Cenp-X as having the strongest influence on the inner-kinetochore's structure. We conclude that our approach can become a useful tool for the *in silico* dynamical study of large, multi-molecular complexes.

1. Introduction

During cell-cycle, accurate DNA segregation is mediated by the kinetochore, a multi-protein-complex that assembles at the centromere of each sister chromatid. Its dysfunction can cause aneuploidy and may lead to the development of cancer (Cimini and Degrassi, 2005; Suijkerbuijk and Kops, 2008). A single kinetochore complex contains over 100 proteins (Chan et al., 2005). These proteins can be grouped into two regions: the inner-kinetochore, which is tightly associated with the centromere DNA, and the outer-kinetochore, which interacts with microtubules. Kinetochore structure and function change during the cell cycle (Varma et al., 2013). Thus it is necessary to learn about the coherence of its structure and function. Furthermore, the structure itself and its stability are of special interest. The outer-kinetochore is thought to be structurally unstable and formed in early mitosis (Cheeseman and Desai, 2008; Maiato et al., 2004) while the inner-kinetochore is more stable and present during the entire cell cycle (Black and Cleveland, 2011; Dalal and Bui, 2010). The inner-kinetochore complex includes a centromeric H3, as well as its variant Cenp-A (Cse4 in budding yeast, Cnp1 in fission yeast, and CID/CenH3 in fruit flies), and 17 CCAN proteins (Cenp-B, Cenp-C, Cenp-H, Cenp-I, Cenp-K to Cenp-U, Cenp-W and Cenp-X) (Okada et al., 2006). These 19 proteins build a bridge between two histone octamers (*cf.* Fig. 1) (Tschernyschkow et al., 2013). The DNA wraps around the histone octamer forming a nucleosome.

In our study we are interested in the spatial behavior of the bridges that form. Studying the 3D structure of the inner-kinetochore is challenging: experimentally it is difficult because the average diameter of a CCAN protein is 40 Å and connections between them are not visible through a microscope. Also,

* Corresponding author. Tel.: +49 3641 9 46460; fax: +49 3641 9 46302.
** Corresponding author at: Bio Systems Analysis Group, Institute of Computer Science, Jena Center for Bioinformatics and Friedrich Schiller University Jena, 07743 Jena, Germany. Tel.: +49 3641 9 46463; fax: +49 3641 9 46302.
E-mail addresses: richard.henze@uni-jena.de (R. Henze), ja.hu.03@uni-jena.de (J. Huwald), nelly.mostajo@upch.pe (N. Mostajo), peter.dittrich@uni-jena.de (P. Dittrich), bashar.ibrahim@uni-jena.de (B. Ibrahim).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.biosystems.2014.11.004
0303-2647/

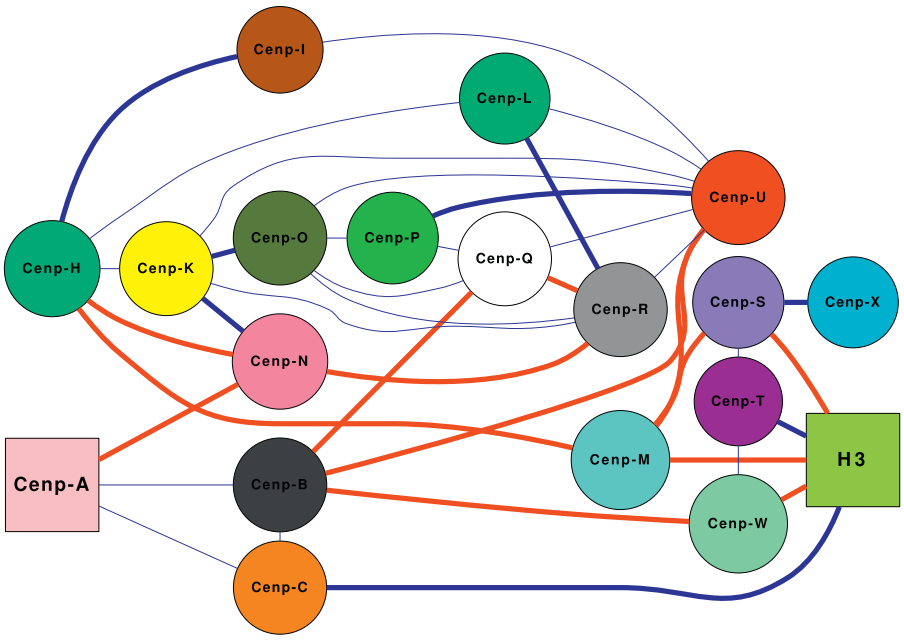

Fig. 1. Interaction graph of centromere proteins. The graph shows all centromere proteins (round vertices), the nucleosomes (square vertices) and all their possible interactions (edges) according to the presented model. The graph topology is derived from the literature (*cf.* Table 1). The edge annotations display the importance of each interaction as computed in this paper: Thin lines denote bonds considered insignificant: their deletion did not influence the assembly of the inner-kinetochore. Thick lines denote the most influencing interaction for every protein. Considering the graph with only thick lines, the blue lines denote terminal bonds while the red lines are the non-terminal bonds. This implies that the red lines are the most important ones for the formation of the inner-kinetochore. These results are discussed detailed in Section 2.3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

theoretical analysis methods are hindered by the combinatorial explosion of the amount of intermediate complexes and protein assembly states (Görlich et al., 2013; Tschernyschkow et al., 2013). Classical *in silico* modeling approaches based on explicit representations of all intermediate complexes, such as differential equations, cannot cope with that combinatorial explosion. A recent S-Phase inner-kinetochore model by Tschernyschkow et al. (2013) overcomes this hurdle using an implicit representation, combining a rule-based language with a particle based simulation approach (Gruenert et al., 2010). This model incorporates molecular geometric information, in contrast to classical models applied so far to cell-cycle mechanisms (*e.g.* Caydasi et al., 2012; Doncic et al., 2005; Ibrahim and Henze, 2014; Ibrahim et al., 2008b, 2009, 2008a, 2007; Kreyssig et al., 2012, 2014; Lohel et al., 2009; Rohn et al., 2008). The spatial rule-based approach with an application to mitotic kinetochore has been recently discussed in detail by Ibrahim et al. (2013). In this study, we extend the S-Phase inner-kinetochore model by Tschernyschkow et al. (2013). We include 13 additional interaction data from the literature and molecular geometries for Cenp-A, Cenp-C, and Cenp-T. While Tschernyschkows' model is built mainly on FRET (Fluorescence Resonance Energy Transfer) and F3H (Fluorescent Three Hybrid) measured interactions (*cf.* Eskat et al., 2012), our new literature data also contain exemplary NMRS (Nuclear Magnetic Resonance Spectroscopy, *cf.* Kato et al., 2013) measurements. All CCAN protein interactions or proximities, measured with any mentioned method (*e.g.* FRET, F3H, NMRS), are included in our simulation. This implies that any kind of source can be translated to rules for our simulation software, which results in a model defining the particles dynamics (*cf.* Section 4.1).

The conjectured kinetochore structure is obtained by simulating the aggregation of all 19 randomly distributed CCAN proteins for a sufficiently long time. With all the interactions included this results in our wild-type structure. Therefore, given different random initial conditions, the model generates not one but a set of structures. For analyzing these structures we build upon the techniques detailed in Tschernyschkow et al. (2013) and Ibrahim et al. (2013). To identify them as (dis)similar we used: bond frequency, relative particle positions, and moments of each particles position. In addition, mutual information to identify protein clusters, forming the same bonds in every simulation run. See Section 4 for details.

Assuming kinetochores loss of function occurs only if the structure changes significantly, we showed that all generated wild-types structures are highly similar. Also, for validating our model, we compared it with *in vitro* mutation experiments. Therefore, we generate mutant simulations for all single-protein deletions. For each mutation we check if it generates the same structures as the wild-type model. We show that the mutation-function mapping obtained by this procedure is largely in accord with *in vitro* experiments. Many inner-kinetochore proteins participate in a relatively large number of reactions: Cenp-U has 9 interaction partners and Cenp-R has 6 (*cf.* Table 1). We want to identify the relevance of individual bonds to gain a functional understanding of the kinetochore. Protein deletion in our model is simulated through the removal of its interactions. This reveals protein effect on the complex structure, but not which interaction affects it. Thus, we add novel mutation experiments in which we disable single interactions and compare the resulting structure with the wild-type.

In the following, we discuss the congruence between our *in silico* protein deletion experiments and the results from the literature. Subsequently the simulated bond deletion experiments are presented. Also shown are isolated interactions with the highest influence on the structure of the inner-kinetochore (*cf.* Fig. 1).

2. Results and discussion

The analysis performed was to study the inner-kinetochore during S-phase in order to accomplish: (1) wild-type behavior to determine a reliable structure and function while including the new data available; (2) protein deletion for comparing with *in vitro* results and validating the model; (3) bond deletion for