Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Computer-aided molecular modeling and structural analysis of the human centromere protein-HIKM complex

Olanrewaju Durojaye[1]

[1]University of Science and Technology of China

July 22, 2021

Abstract

Protein-peptide and protein-protein interactions play an essential role in different functional and structural cellular organizational aspects. While X-ray crystallography generates the most complete structural characterization, most biological interactions exist in biomolecular complexes that are neither compliant nor responsive to direct experimental analysis. The development of computational docking approaches is therefore necessary. This starts from component protein structures to the prediction of their complexes, preferentially with precision close to complex structures generated by X-ray crystallography. To guarantee faithful chromosomal segregation, there must be a proper assembling of the kinetochore (a protein complex with multiple subunits) at the centromere during the process of cell division. As an important member of the inner kinetochore, defects in any of the subunits making up the CENP-HIKM complex leads to kinetochore dysfunction and an eventual chromosomal mis-segregation and cell death. Previous studies in an attempt to understand the assembly and mechanism devised by the CENP-HIKM in promoting functionality of the kinetochore, have reconstituted the protein complex from different organisms including fungi and yeast. Here, we present a detailed computational model of the physical interactions that exist between each component of the human CENP-HIKM, while validating each modeled structure using orthologs with existing crystal structures from the protein data bank. Results from this study substantiates the existing hypothesis that the human CENP-HIK complex share a similar architecture with its fungal and yeast orthologs, and likewise validates the binding mode of CENP-M to the C-terminus of the human CENP-I based on existing experimental reports.

Introduction

Although, molecular and cell biology have made huge advancement towards the delivery of powerful methodologies for the discovery and identification of protein-protein interactions, likewise their sub-cellular localization, yet structural biology alone is able to give definite answers regarding interaction mechanisms through the uncovering of atomistic and high-resolution structures of the underlying complexes [1]. Determination of the structure of such biomolecular interactions however can be a costly, laborious and time-consuming endeavor [2]. The gap increment between the universe of determined 3D structures and that of known sequences is a proof that high-throughput structural biology remains a fantasy [3], as the gap increases the more with the consideration of available number of biomolecular complex structures [4]. By contrast, computational structural biology has the potential to generate protein–protein interaction models of high resolution [5].

Timely and accurate segregation of chromosomes in meiosis and mitosis is crucial for organismal and cellular viability. Sister chromatids produced through DNA replication during mitosis maintain strong cohesion till a bioriented arrangement is formed on the mitotic spindle. The loss of sister chromatid cohesion during the transition from metaphase to anaphase allows for a successful separation of the sister chromatids into daughter cells with genetic identity [6]. The sister chromatid attachment to microtubules is mediated by the kinetochores. Kinetochores become established on a part of the centromere (a specialized chromatin), with the presence of CENP-A (a variant of histone H3) as a major hallmark [7, 8]. The kinetochores at low resolution assume a laminar structure appearance, with the ends of each microtubule connected to its

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

outer plate and a dense centromeric chromatin adjacent to its inner plate [9]. The outer kinetochore plate serves as a host for the KMN network (Knl1, Mis12 and Ndc80 complexes); an assembly consisting of ten protein subunits that act as a microtubule receptor [10, 11]. The inner kinetochore on the other hand serve as a host for the CCAN (constitutive centromere–associated network), a complex consisting of sixteen different centromeric proteins (CENPs), most of which were identified originally in the vertebrates' CENP-A interactome [12, 13].

The sixteen CCAN proteins of vertebrates are grouped into different sub-complexes including, CENP-LN, CENP-C, CENP-OPQUR, CENP-HIKM and CENP-TWSX [14, 15]. Orthologs of most of the listed sub-complexes have been recognized in species like fungi and yeast [16, 17]. As a nucleosomal canonical H3 substitute, the CENP-A accumulates at the nucleosome of centromeres [18] for the initiation of the CCAN assembly through the binding to CENP-C and CENP-LN [19, 20]. Several studies have also established the crucial role of the CCAN in mediating the outer kinetochore assembly [21, 22]. CENP-T and CENP-C function as the outer kinetochore structural platform through a direct interaction with the NDC80 and MIS12 complexes [23, 24].

Many CCAN components are held in place by a cumbersome protein-protein interaction network [25, 26]. However, the exact way in which the CCAN complex is assembled by these interactions is yet to be completely understood. As a core CCAN subunit, CENP-H (Mcm16/Fta3), CENP-I (Ctf3/Mis6) and CENP-K (Mcm22/Sim4) assemble into a ternary complex and are likewise crucial for the kinetochore integrity. Chromosomal congression is compromised upon the loss of any of these proteins [27] while their localization to the centromere has also been revealed to be dependent on each other [28]. CENP-M (another subunit of the CCAN) through *in vitro* reconstitution has been shown to form a stable complex with the CENP-HIK via an interaction with the CENP-I C-terminus. This interaction is essential for chromosomal alignment and also for the localization of the CENP-IM to the centromere [29]. Although low-resolution electron microscopy analyses have shown the overall CENP-HIKM organization, the specific molecular basis for the complex assembly remains predominantly uncharacterized [29]. With reference to the existing complex structure of the CENP-HIK from yeast and fungi, we have predicted in this study the organizational model of the human CENP-HIKM complex, using extensive computational approaches. Our result also shows great consistency with experimental inter-model interaction studies from several published literature.

Methods

Reference sequence and structure retrieval

For structural validation purpose, the modeled protein complex (*hs* CENP-HIKM) was compared with structural orthologs with known 3-dimensional structures. The reference sequences and structures were retrieved from the NCBI (National Center for Biotechnology Information) database [30], and the PDB (Protein Data Bank) [31]. 5Z08 and 6YPC which represent the PDB codes for the crystal structures of the fungal (*Thielavia terrestris*) kinetochore CENP-HIK triple complex subunits and the yeast (*Saccharomyces cerevisiae*) kinetochore CENP-HIKTW subunits respectively, were used for the retrieval of the corresponding structures from the protein data bank. The crystal structure of the human CENP-M was also retrieved with the PDB code 4P0T. The PDB codes for each structure were submitted to the NCBI database to obtain their corresponding amino acid sequences while the full length sequence for each subunit of the human CENP-HIK were retrieved using their respective accession numbers; Q9H3R5, Q92674 and Q9BS16.

3D structural modeling of the human CENP-HIK

High-quality 3D structural models of the *hs* CENP-H, -I, and -K were individually predicted using the RaptorX Contact tool [32]. RaptorX Contact predicts contacts through the integration of both sequence conservation and evolutionary coupling information by using an ultra-deep neural network formed by two residual neural networks. Different forms of 1-dimensional sequential feature transformation is conducted by the first residual network while the second conducts different types of 2-dimensional pairwise information transformation which include, pairwise potential, first residual network output, and evolutionary coupling information. Through the use of these very deep residual networks, RaptorX Contact accurately models

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

patterns of contact occurrence and complex sequence-structure relationship [32].

Structural refinement and model quality evaluation

Following a successful modeling of each protein subunit (*hs* CENP-H, -I, and -K), a structural refinement protocol was conducted using the GalaxyRefine [33] which is based on a method of refinement that has successfully undergone trials in CASP10. In this method, side chains are first rebuilt and through molecular dynamics simulation, an overall structure relaxation is performed. This approach according to the assessment by CASP10 displayed the best potential in local structure quality improvement. Both local and global structure qualities were improved upon the refinement of the RaptorX Contact-generated models using this method. The quality of each refined structural model was assessed using the ProSA-web program [34] which implements the z-scoring function for structural analysis. Additional model quality assessment was conducted using the PROCHECK suite [35]. The PROCHECK suite delivers a comprehensive stereochemistry check on protein structures. The generated output is made up of several PostScript format plots and a detailed residue-by-residue listing. This highlights regions of the protein structure that might require additional investigation and also gives an evaluation of the overall structural quality in comparison to well refined structures [35].

Structural alignment and visualization

In order to evaluate the degree of structural similarity between each human model and their corresponding orthologs, we carried out a structural alignment protocol using the Alignment/Superposition function of the Pymol molecular visualizer plugin [36]. PyMOL is a cross-platform tool for molecular graphics and it has been popularly used for the 3D visualization of trajectories, surfaces, electron densities, small molecules, nucleic acids and proteins. The tool is also used for movie making, molecule editing and ray tracing. PyMOL being a Python-based software has been designed along with many plugin tools to facilitate its usage for the 3D visualization of macromolecules as performed in this study [36].

Validation of residue conservation

Specific residues of CENP-H and CENP-M have been reported in different studies as essential in facilitating intermolecular interaction with residues of other subunits in the complex [14, 37]. We validated the evolutionary conservation of these residues with the use of ConSurf [38]. Consurf is popularly used for the detection of macromolecules' functional regions, through the analysis of the evolutionary dynamics of nucleic acid and amino acid substitutions in homologous sequences. The tool evaluates the nucleic acid and amino acid evolutionary rates by mapping them onto the structure or sequence of the query macromolecule. Slowly evolving regions on the surface of the query macromolecule are known to be essential for functionality and thus, the analysis of ConSurf can highlight very important regions within the query macromolecule [38].

Protein-protein docking

The molecular docking protocol for the purpose of predicting the binding modes and pattern of organization of each member of the CENP-HIKM complex was conducted using ClusPro [39], which is a popularly used tool for the docking of different proteins. ClusPro provides multiple computational steps: rigid docking sampling of billions of conformations, RMSD (root-mean-square deviation)-based clustering of structures with the lowest energy (which are generated to detect the largest clusters that will represent the complex's closest models), and energy minimization refinement of selected structures. ClusPro employs PIPER, a docking algorithm that is anchored on the Fast Fourier Transform (FFT) correlation technique, to dock the rigid body. The FFT technique has made significant progress in rigid body protein-protein docking [39]. The method involves placing a protein (the receptor) at the coordinate system origin on a fixed grid and another protein (the ligand) on a moveable grid, with the energy of interaction represented as a correlation function. The numerical efficiency is reinforced by the fact that such energy functions can be generated quickly, allowing for the sampling of various conformations of protein-protein interactions as well as the evaluation of grid point energies. As a result, a FFT-based approach allows for protein docking without prior knowledge of their structures [39].

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Normal mode analysis and changes in vibrational entropy

The normal mode dynamics of the hypothetical *hs* CENP-HIKM complex was assessed using the iMOD [40] and DynaMut [41] tools. This analysis was directed at determining the stability of the docked complex and also for the exploration of the protein-protein interaction dynamics. iMOD analyses conformational flexibility of nucleic acid and protein structures by utilizing the normal mode analysis in internal coordinates. Considering the dihedral angles as variables, lowers the non-physical distortions and cost of computation of classical Cartesian normal mode analysis approaches. Operation of the framework is at various coarse-grained levels and delivers an active framework for the conduction of normal mode analysis-based conformational studies which include pathway exploration, vibrational analysis or Monte-Carlo simulations [40]. The iMOD normal mode analysis also function as a rational option for atomistic simulation. The stiffness of motion is presented by a given value while covariance matrix, eigenvalue, deformability and elastic network model are also calculated [40]. DynaMut on the other hand implements normal mode analysis using two different methods, ENCoM and Bio3D, delivering simplified and rapid access to insightful and efficient protein motion analysis [41]

In-silico mutagenesis

To assess the consistency of the predicted CENP-HIKM organizational pattern with experimental reports from previous studies [14, 37],*in-silico* mutants of the CENP-H and CENP-M were designed using the Chimera-curated backbone-dependent Dunbrack rotamer library [42]. The backbone-dependent rotamer library is composed of variances, rotamer frequencies, and mean dihedral angles as a function of the backbone dihedral angles. The prediction of structures and methods of design that utilizes backbone flexibility benefit strongly from smoothly varying angles and probabilities. A new backbone-dependent rotamer library version was developed to use adaptive kernel regression for variances and mean dihedral angle calculations and also the adaptive kernel density calculations for the frequency of rotamers. This design creates an avenue for the estimation of variances, probabilities of rotamers, and mean angles as a continuous and smooth function of psi and phi [42]. The estimation of continuous probability density for the nonrotameric degrees of freedom of aromatic side chains, amides, and carboxylates was modeled as a function of the rotamers and backbone dihedrals of the residual degrees of freedom.

Binding free energy prediction

The binding free energy of the wild-type protein complex and the change in binding free energy upon mutation was predicted using different predictive tools, such as BeAtMuSiC [43], mCSM-PPI2 [44], mmCSM-PPI [45], MutaBind2 [46] and HawkDock [47]. BeAtMuSiC is a coarse-grained prediction tool for the binding free energy changes as a result of point mutations. The algorithm depends on a set of statistical potentials extracted from proteins with known structures and combines the mutation effect on the overall complex stability and on the strength of the protein-protein interactions at the interface [43]. mCSM-PPI2 is a novel machine learning tool developed for the precise prediction of missense mutation effects on the binding affinity of protein-protein interactions. The tool utilizes graph-based structural signatures for the modeling of variation effects on energetic terms, complex network metrics, evolutionary information and inter-residue interaction network for the generation of an optimized prediction tool [44]. mmCSM-PPI is an effective and scalable machine learning tool for the accurate assessment of protein-protein interaction binding affinity changes resulting from multiple and single missense mutations. The tool utilizes a well-established graph-based signature in capturing geometrical and physiochemical properties of various wild-type residues and integrates them with both normal mode analysis dynamics terms and substitution scores [45]. MutaBind2 estimates protein-protein interaction binding affinity changes as a result of single- and multiple-site mutations in corresponding sequences. The tool makes predictions based on the protein-protein complex structure. MutaBind2 uses rapid side chain optimization algorithms built through random forest method, mechanics force fields and statistical potentials. The training set used for the development of multiple and single models of mutation consist of 1707 multiple mutations from 120 protein complexes and 4191 single mutations from 265 protein complexes respectively [46]. The development of HawkDock was targeted at the prediction and analysis of protein-protein interactions through the integration of the MM/GBSA free energy decomposition

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

analysis, ATTRACT docking algorithm and the HawkRank scoring function. The integration of MM/GBSA into HawkDock is to serve the purpose of analyzing important residues in the binding interface of protein-protein interactions and also for the purpose of model re-ranking [47].

Interatomic interaction analysis

The existing non-covalent interactions between subunits of the CENP-HIKM complex were analyzed using Arpeggio [48]. The program is implemented in Python and it calculates interactions between and within proteins and small-molecule ligands, protein or DNA. Analyzed interactions in this study include van der Waals', hydrogen bonds and hydrophobic interactions.

Results

Modeling of the human CENP-H, CENP-I and CENP-K

Structural models were generated using the amino acid sequence of each protein as described in the methods segment, as input. The 3D structure prediction method employed by RaptorX contact is unique in that it makes a simultaneous prediction of all protein contacts, which allows for an easy modeling of high order residue correlation. The output provides 5 different models that are ranked by estimated root mean square deviation (RMSD). The estimated RMSD is a calculated average deviation distance in Å of a 3D model from its experimental structure. The smaller the estimated RMSD value is, the higher the likelihood of the 3D model to a good quality. The estimated RMSD values of the top rank models for CENP-H, -I and -K (Figure 1) are 5.7546 Å, 13.445 Å and 5.7311 Å respectively. All generated models share high similarity with the PDB structures of their respective orthologs (Supplementary Figures S1 and S2) and as such were selected for structural refinement.

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

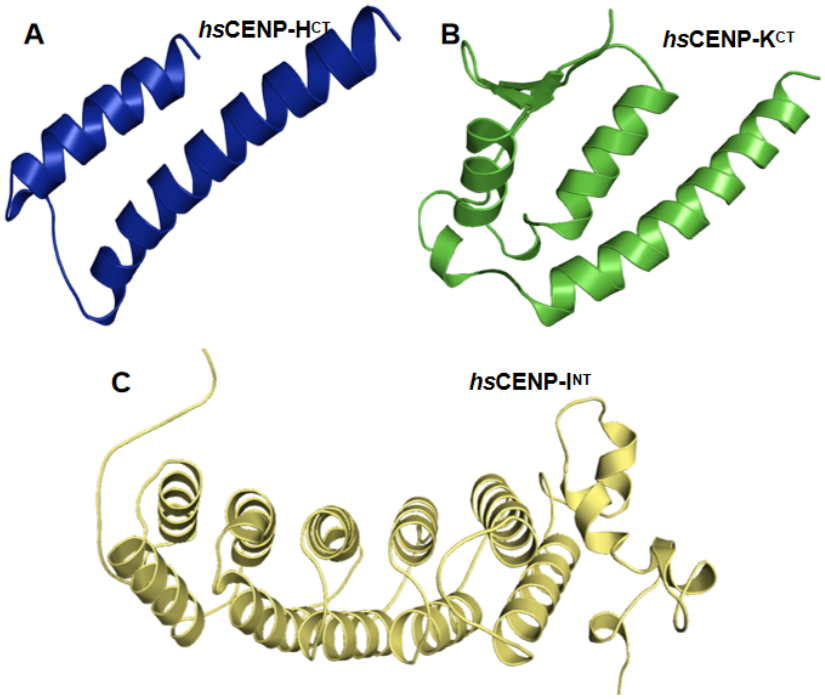

Figure 1. Predicted 3D models of the human CENP-H (blue), CENP-I (pale yellow) and CENP-K (green) showing close similarity with X-ray crystal structures of fungal (PDB 5Z08) and yeast (PDB 6YPC) orthologs.

Model quality evaluation

Following the structural refinement of the three models with GalaxyRefine, we proceeded with the protocol for the quality evaluation of each model. Using the ProSA-web, the z-score for each model was obtained (Figure 2). The z -score is an indication of the overall model quality. *hs* CENP-H, *hs* CENP-K and *hs* CENP-I produced individual z-scores of -3.17, -4.66 and -7.38 respectively, indicating that all the three models fall within the quality range of the nuclear magnetic resonance (NMR) as shown in Figure 2.

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

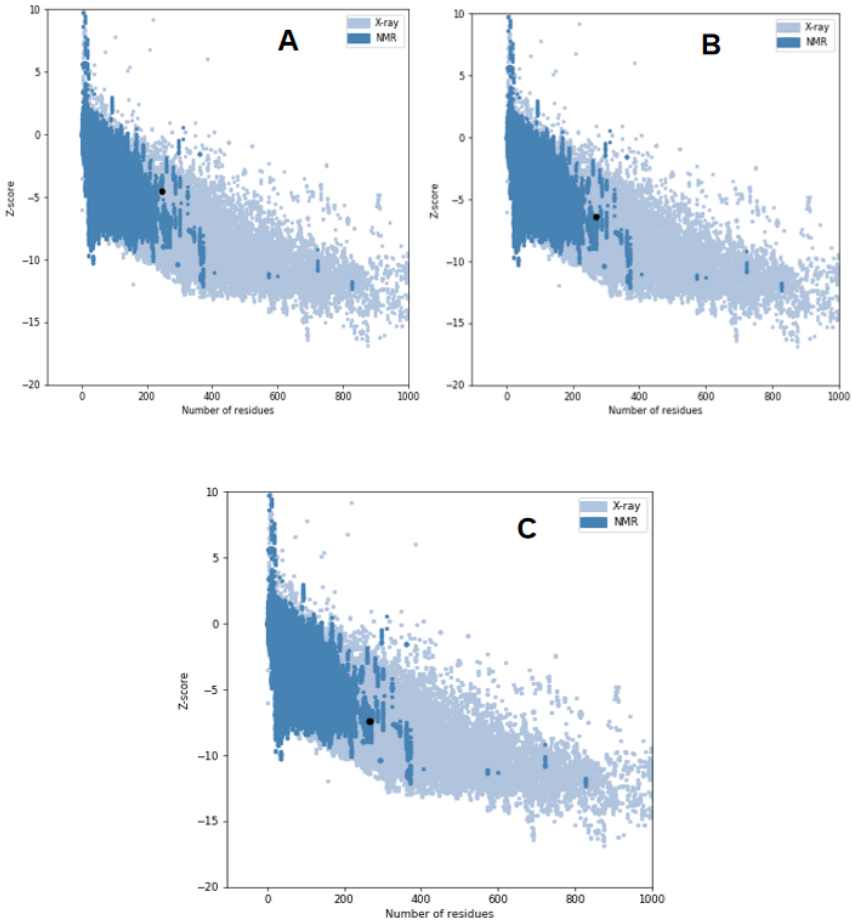

Figure 2. Z-score plots for (A) CENP-H, (B) CENP-K and (C) CENP-I. The plots contain the z -scores of all protein chains in the current PDB that has been experimentally determined. In this plot, structural groups from different sources such as, X-ray and NMR ,are differentiated by different colors.

The PROCHECK suite was used for the calculation of the stereochemical quality of the models through the analysis of the overall structural and individual residue geometry. The Ramachandran plot for each modeled protein showed that over 92% of the residues were located in the most favourable region, with an average of 5.4% of the residues located in the allowed region while less than 0.4% were in the disallowed region (Supplementary Figure S3). Based on the analysis of 118 structures with a minimum of 2.0 Å and a maximum R-factor value of 20%, it is expected that a good quality model will have more than 90% of its resides in the most favored region.

Structural alignment

Previous studies have reported a high degree of sequence conservation between the various subunits of the

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

CENP-HIK complex across different organisms. It is therefore expected that the human model of each subunit display a high level of structural similarity with the reference structures from fungi and yeast (Figure 3) to further validate the reliability and quality of the models.

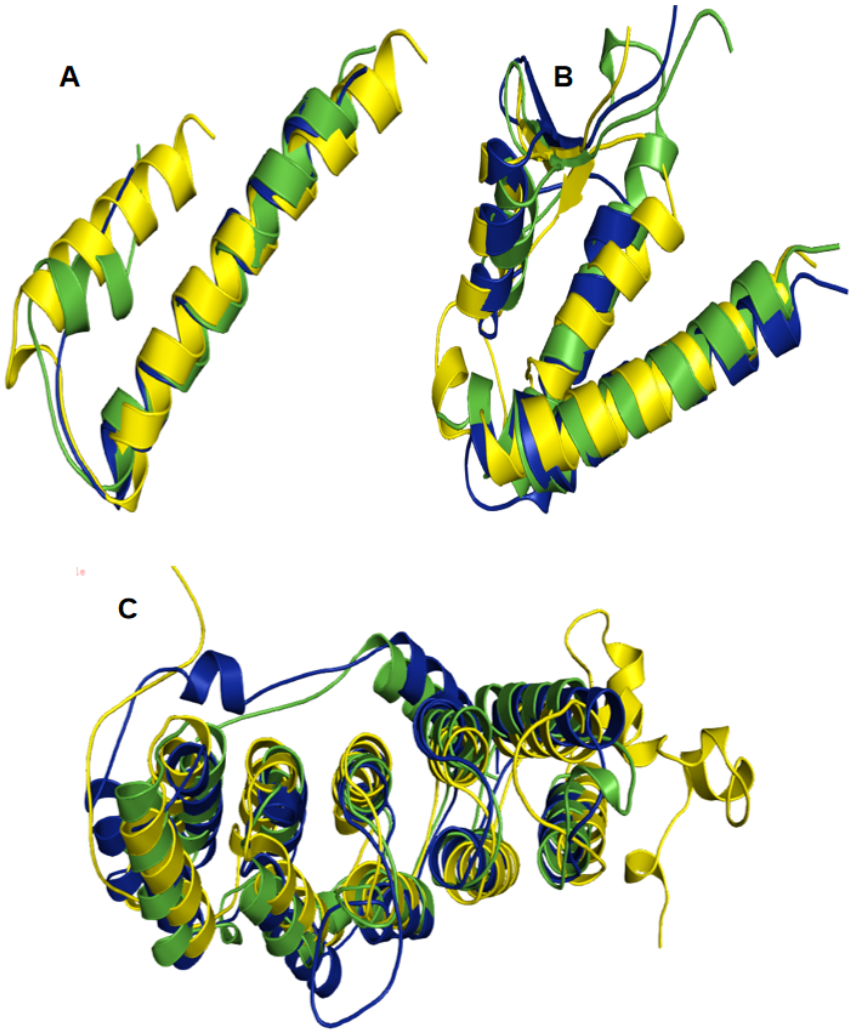

Figure 3. Alignment of the CENP-HIK subunits from fungi, yeast and the human models. (A) structural alignment of the*hs* CENP-H^{CT}, *th* CENP-H^{CT}and *sc* CENP-H^{CT}, coloured in yellow, green and blue respectively. (B) and (C) also shows the structural alignment of CENP-K^{CT} and CENP-I from the three organisms, colored in yellow, green and blue respectively. The alignment shows a high degree of structural similarity between the structures.

Computational validation of residue conservation

Details of the CENP-HK binding interface at the C-terminal was revealed in the crystal structure of the fungal HIK complex (5Z08). The side chain of ILE-205, ILE-211 and LEU-219 from *th* CENP-H were shown

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

to insert into the hydrophobic pocket of *th* CENP-K which is surrounded by several residues including, LEU-177, TRP-179, PHE-180, HIS-184, ILE-270 and PHE-300. On the CENP-HI interface, *th* CENP-H uses its contacting helix (HH2) in interacting with the*ct* CENP-I^{NT} HEAT repeat. A salt bridge was reported to be formed between the ARG-220 of *th* CENP-H and the GLU-86 of *ct* CENP-I^{NT}, while the LEU-224 of was reported to insert into the *ct* CENP-I^{NT}hydrophobic pocket (surrounded by LEU-89, VAL-126 and VAL-130) [14]. The alignment of amino acid sequences from different orthologs of CENP-H (*T. terrestris, G. gallus, O. aries, R. norvegicus, M. musculus and H. sapiens*) revealed a high degree of conservation in favour of the CENP-K and -I-binding residues of the protein, which in human correspond to LEU-219, VAL-225, LEU-233, LYS-234 and LEU-238 [14]. Using ConSurf, we validated the degree of conservation of the reported residues in the *hs* CENP-H model. The output depicted that all the five reported residues (LEU-219, VAL-225, LEU-233, LYS-234 and LEU-238) are conserved with varying degrees of conservation (Figure 4).

Posted on Authorea 22 Jul 2021 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.162693577.78775249/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Figure 4. Residue conservation output for the (A) *hs* CENP-H model and the (B) *hs* CENP-M (PDB 4P0T). The color gradient shows the different scales of conservation. Highly conserved residues are colored in deep red while white coloration indicate average conservation. Non-conserved residues are colored in blue. Based on the neural network algorithm, "e" and "b" indicate exposed and buried residues respectively while "f" and "s" respectively denote functional and structural residues.

In a similar study involving the human CENP-M (PDB 4P0T), conserved surface residues were also identified to be involved in the interaction with the C-terminal of the *hs* CENP-I; an interaction which leads to the stabilization of the *hs* CENP-I and likewise required for an unabated kinetochore localization [37]. Using the same computational approach, the reported conserved surface residues of the *hs* CENP-M were also shown to be conserved and each exhibit varying degrees of conservation (Figure 4B), hence validating experimental reports from the previous studies.

Protein-protein docking study

With the availability of the *hs* CENP-M crystal structure (PDB 4P0T) and having successfully generated high quality models for each component of the *hs* CENP-HIK complex, we proceeded with the docking of the subunits. According to the Hu *et al* . [14] model, biochemical analysis and structures revealed that the*th* CENP-K and *th* CENP-H form a heterodimer via interactions at both N-terminal and C-terminal. The integration of*ct* CENP-I^{NT} into the complex is through its interaction with the *th* CENP-H C-terminal, resulting in the formation of a ternary complex where *th* CENP-H is sandwiched between *ct* CENP-I^{NT} and *th* CENP-K [14]. The study also reported the conservation of this architecture in the human HIK complex. Upon the stepwise docking of each generated model of the *hs* CENP-H, -I, and -K, the resulting output showed a similar architecture with the experimental reports from literature, suggesting a structural conservation across the species (Figure 5).

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

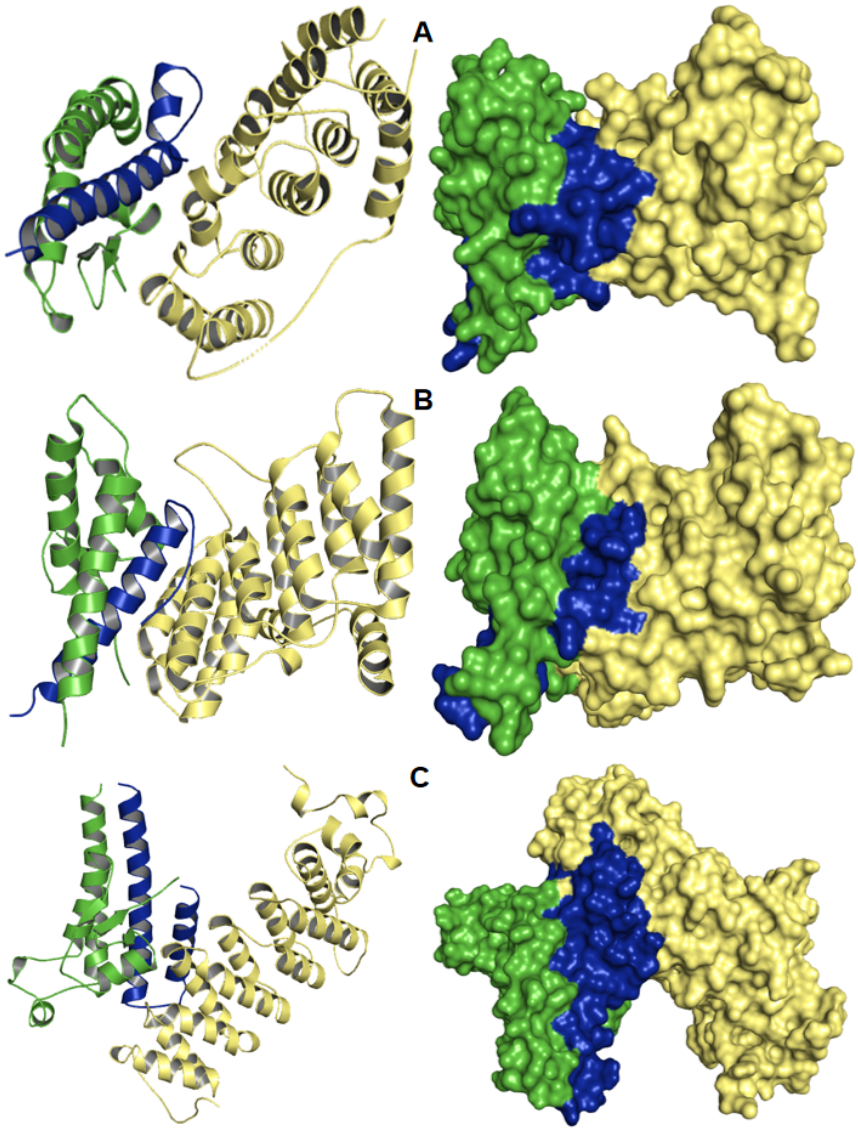

Figure 5. Cartoon and surface representation of the (A) fungal (PDB 5Z08), (B) yeast (PDB 6YPC), and (C) human model of the CENP-HIK structural architecture. All CENP-H (*th* CENP-H$^{184\text{-}227}$,*sc* CENP-H$^{147\text{-}181}$, and*hs* CENP-H$^{194\text{-}247}$) are sandwiched between CENP-I (*ct* CENP-I$^{8\text{-}229}$,*sc* CENP-I$^{2\text{-}241}$, and*hs* CENP-I$^{1\text{-}265}$) and CENP-K (*th* CENP-K$^{161\text{-}328}$,*sc* CENP-K$^{136\text{-}237}$, and*hs* CENP-K$^{161\text{-}269}$). The blue, green and pale yellow colors denote CENP-H, CENP-K and CENP-I respectively for all species.

In a similar study, Basilico *et al.* [37] reported the structural organization of the *hs* CENP-HIKM complex, using a computational model to represent the full length *hs* CENP-I as there existed no full length ortholog of the protein. Consistent with existing literature reports, the molecular docking output also showed that the *hs* CENP-M binds to the C-terminal of the full length*hs* CENP-I model (Figure 5) in an appearance that resembles the importin-β/Ran complex as reported by Basilico *et al* . [37]. The α-solenoid fold of importin-β is consistently reported to be a high-confidence *hs* CENP-I structural modeling template [37].

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

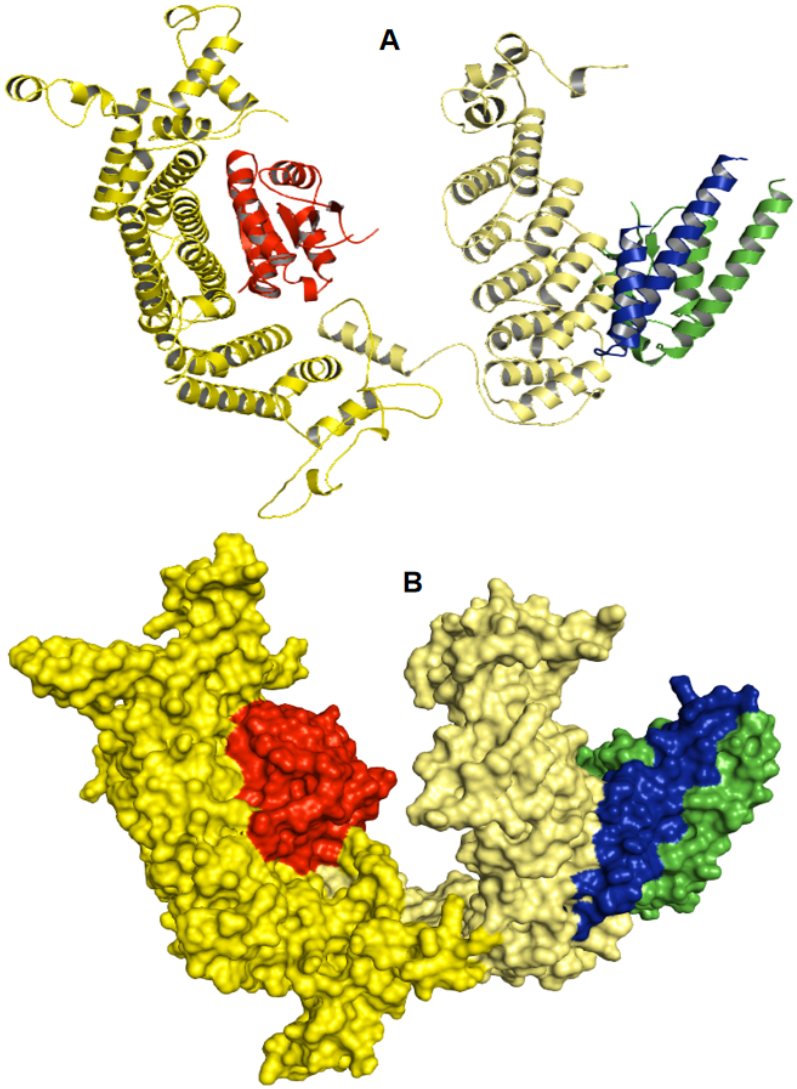

Figure 6. (A) Cartoon and (B) surface representation of the docked *hs* CENP-HIKM complex. The *hs* CENP-I is colored in two shades of yellow in order to distinguish the N-terminal (pale yellow) from the C-terminal. The binding of the *hs* CENP-M to the C-terminal of *hs* CENP-I shows great resemblance with the previously described binding of Ran to importin-β. The *hs* CENP-H,*hs* CENP-K and *hs* CENP-M are colored in blue, green and red respectively.

Normal mode analysis

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

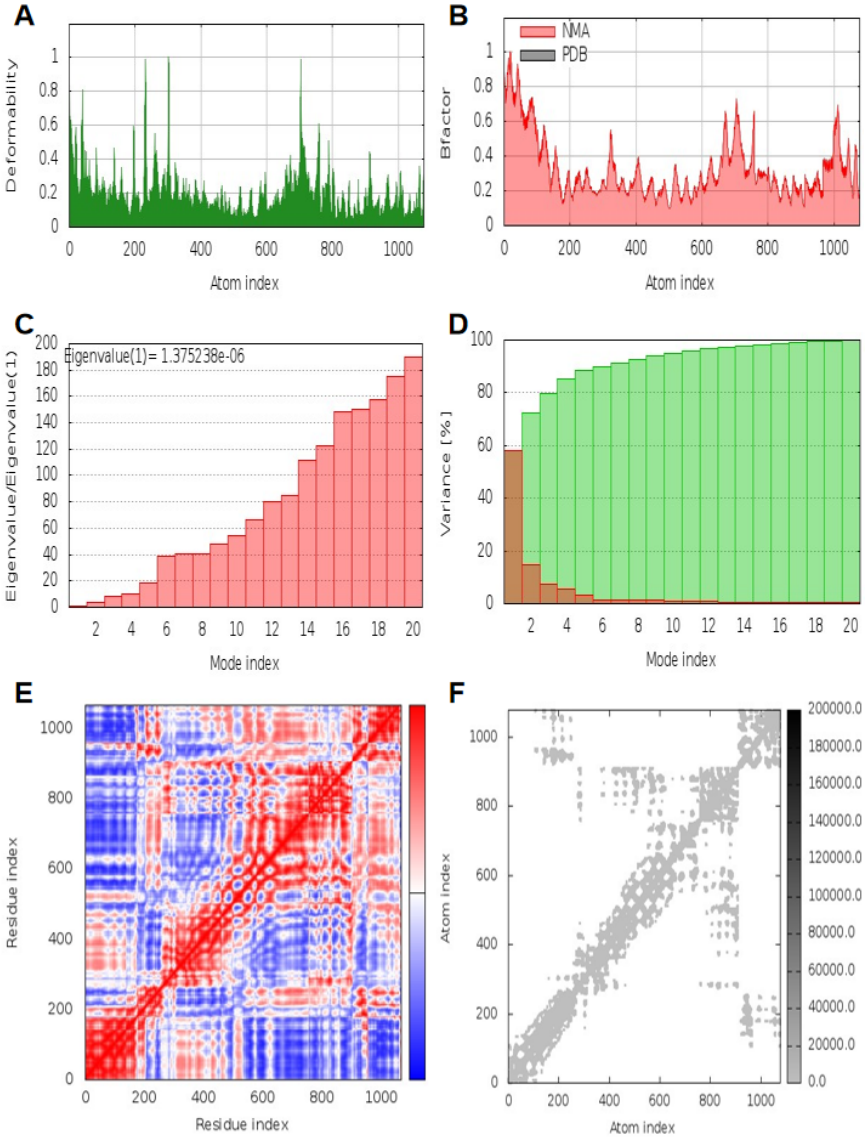

Figure 7. Normal mode analysis output for the*hs* CENP-HIKM complex model, showing the (A) main-chain deformability, (B) Bfactor, (C) Eigenvalues, (D) Variance, (E) Covariance map and the (F) Elastic network map.

The quality and stability of the hypothetical *hs* CENP-HIKM model were evaluated through the iMod-estimated elastic network map, deformability, covariance map, eigenvalue and the B-factor (Figure 7). The deformability of the main chain is an estimation of the deformation capability of a molecule at each of its residues. The Bfactor (a crystallographic atomic displacement parameter) is reported for most X-ray crystal structure of proteins and it is directly related to the fluctuations due to static disorder or motion in structures. The Bfactor also provides a measure of an averaged root mean square (RMS). Motion stiffness is represented by the eigenvalue that is associated to each normal mode. Its value is related directly to the required energy for structural deformation. The green and red colored bars show the cummulative and individual variances respectively, while the covariance matrix denotes residue pair coupling, i.e. whether the paired residues experience anti-correlated, uncorrelated or correlated motions (colored in blue, white and

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

green respectively). Atom pairs that are connected by springs are defined by the elastic network model. Each graphical dot represent a spring between the corresponding atom pairs. The dots are colored based on their stiffness, which means the darker gray colors denote stiffer springs and vice versa. Figure 7 shows an average root mean square in the Bfactor and an insignificant hinge. The high eigenvalue (1.375238e-06) is an indication of a low deformation chance, while the elasticity and correlation also demonstrated the high quality of the hypothetical protein complex model (Figure 7).

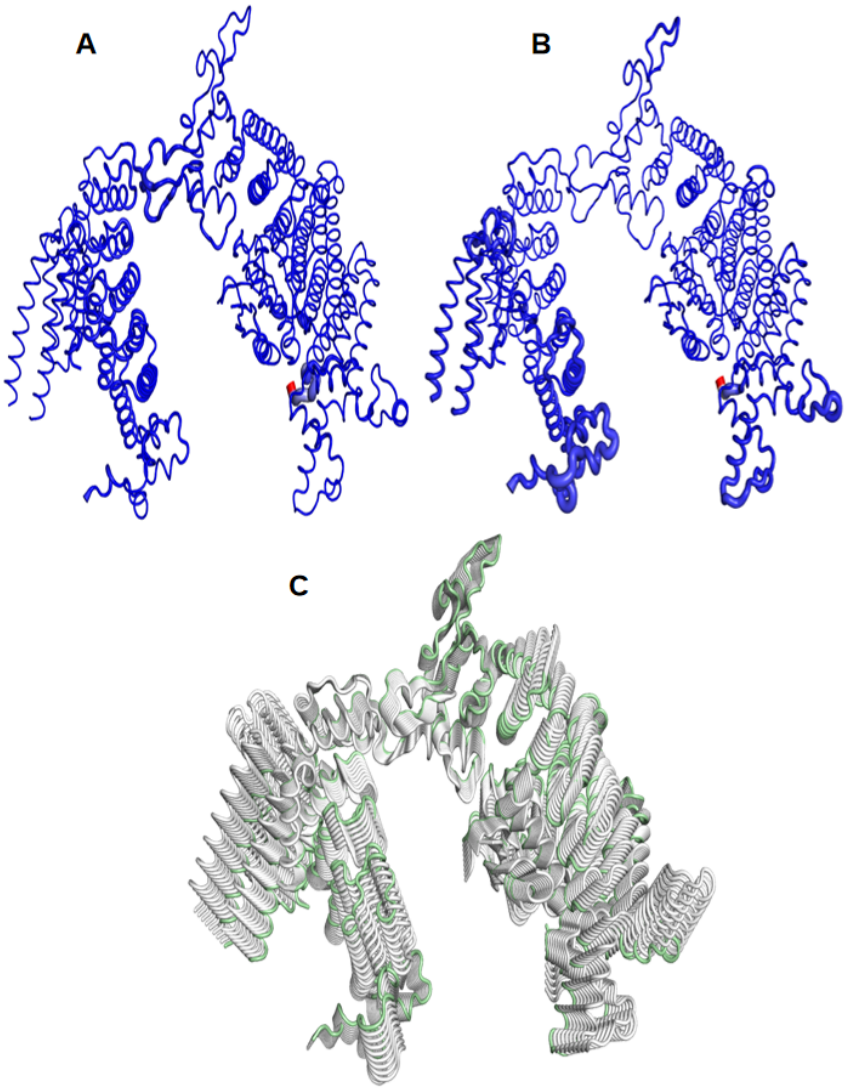

Figure 8. Visual analysis of the hypothetical*hs* CENP-HIKM complex (A) deformation energy and (B) atomic fluctuation. The deformation and fluctuation magnitude is represented by the differential coloration of the 3D structure. Low, moderate and high deformation and fluctuation magnitudes are represented by the blue, white and red colors respectively. (C) NMA-based representation of the trajectory for the first non-trivial mode of the hypothetical*hs* CENP-HIKM complex. The figure shows the superimposition of each mode in the trajectory.

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

A similar analysis was conducted using DynaMut. The DynaMut normal mode analysis protocol is based on a bio3D package that utilizes a default C-alpha force field. The DynaMut-calculated deformation energy gives an estimation of protein complex local flexibility while the atomic fluctuation shows the amplitude for the absolute atomic motion. The predominant blue coloration of the 3D protein complex structure as depicted in Figure 8A and B, denotes a high level of structural stability. All calculations were performed over the first ten non-trivial modes of the protein complex. Included in the DynaMut output also is the flexibility trajectory of the protein complex based on normal mode analysis (Figure 8C), and the correlation map which reveals the anti-correlated and correlated regions in the protein complex. Both regions (anti-correlated and correlated) on the map are colored in blue and red respectively (Supplementary Figure S4). A 3D animation was also generated to simulate the motion of the protein complex (Supplementary Figure S5).

In-silico mutagenesis and binding free energy prediction

Following the experimental mutational analysis from previous studies [14, 37], we designed *in-silico* mutants of the*hs* CENP-H and *hs* CENP-M in an attempt to validate the predicted interactions between each subunit of the hypothetical*hs* CENP-HIKM complex (Supplementary Figures S6 and S7). In order to validate predicted interface interactions between the C-terminal of the *hs* CENP-H and other subunits (C-terminal of the*hs* CENP-K and the N-terminal of the *hs* CENP-I), Hu *et al* . [14] constructed several mutants of the protein (L219A, V225A, L233A, K234A and L238A) based on residue conservation. The mutated residues correspond to ILE-205, ILE-211, LEU-219, ARG-220 and LEU-224 respectively in the *th* CENP-H. A dramatic reduction in binding affinity was recorded upon the mutation of each residue to alanine, indicating that the residues are essential for the protein-protein interaction of the complex. In a similar study by Basilico *et al* . [37], mutants of the *hs* CENP-M (L94A and L163E) were also designed based on residue conservation analysis and the mutation of both residues to alanine and glutamate respectively affected the interaction of the protein with the C-terminal of the *hs* CENP-I.

Having successfully designed the *in-silico* mutants of these proteins in line with reports from existing literature, we predicted the binding free energy changes using several predictive tools as reported in the materials and methods section. The reduction in binding free energy as a result of these mutations shows the consistency of our computational model with experimental reports (Table 1-4).

Table 1. The BeAtMuSiC-predicted change in binding affinity as a result of mutation. With the main input being the protein–protein complex structures, the output reports the mutated chains, the specific mutations, change in binding free energy (Kcal/mol) and solvent accessibility, both in partner and in complex. Solvent accessibility depicts the solvent-accessible surface ratio in the structure, based on DSSP computation.

Chain(s)	Mutation(s)	$\Delta\Delta\Gamma_{\beta\iota\nu\delta}$ (κςαλ /μολ)	SA (in partners)	SA (in complex)
H	L219A	2.91	51.27%	0.52%
H	V225A	0.42	42.55%	38.91%
H	L233A	0.89	13.46%	13.46%
H	K234A	1.70	44.29%	8.41%
H	L238A	1.84	49.72%	9.84%
M	L94A	1.85	36.25%	2.07%
M	L163E	1.23	45.05%	20.71%

Table 2. mCSM-PPI2 binding affinity change prediction upon residue mutation. The wild-type structure is required as input while the output displays the mutated chains, position of each mutated residue, the wild-type residues, mutants, binding free energy change in Kcal/mol and the effect of each mutation on the binding affinity of the protein to its interacting partner

Chain(s)	Position	Wild-type residues	Mutant residues	$\Delta\Delta\Gamma^{\mathrm{A}\varphi\varphi\iota\nu\iota\tau\psi}$ (Κςαλ/μολ)	Effect
H	219	LEU	ALA	-1.387	Decreasing affinity
H	225	VAL	ALA	-0.22	Decreasing affinity

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

H	233	LEU	ALA	-0.427	Decreasing affinity
H	234	LYS	ALA	-2.149	Decreasing affinity
H	238	LEU	ALA	-1.091	Decreasing affinity
M	94	LEU	ALA	-1.615	Decreasing affinity
M	163	LEU	GLU	-1.19	Decreasing affinity

Table 3. mmCSM-PPI-predicted binding affinity change upon residue mutation. Using the 3D structure of the wild-type complex as input file, the tool display details about the mutated chains, mutated residues, average distance of the mutated residue from its closest interacting partner, individual binding free energy change in Kcal/mol, average (predicted) binding free energy change in Kcal/mol, and the impact of each mutation on the binding affinity of the complex

Chain(s)	**Mutation(s)**	**Average distance (Å)**	**Ινδιιδυαλ $\Delta\Delta\Gamma^{\text{Βινδινγ}}$ (Κςαλ/μολ)**	**Πρεδιςτεδ $\Delta\Delta\Gamma^{\text{Βινδινγ}}$ (Κςαλ/μολ)**	**Effect**
H	L219A V225A L233A	9.75	-1.39 -0.22 -0.43	-1.15	Decreasing affinity
H	K234A L238A	6.16	-2.15 -1.09	-2.88	Decreasing affinity
M	L94A L163E	4.84	-1.61 -1.19	-2.09	Decreasing affinity

Table 4. The MutaBind2 mutation-induced change in binding affinity prediction. The input requires the PDB structure of the protein complex with a minimum of 2 distinct chains. A negative and positive $\Delta\Delta G_{bind}$ (kcal/mol) value correspond to stabilizing and destabilizing mutations, predicted to increase and decrease the binding affinity respectively. An all-positive value for the MutaBind2 output denote a destabilizing effect on the protein-protein interaction.

Chain(s)	**Mutation(s)**	**$\Delta\Delta\Gamma^{\text{Βινδινγ}}$ (Κςαλ/μολ)**
H	L219A	1.95
H	V225A	0.28
H	L233A	0.8
H	K234A	0.13
H	L238A	2.18
M	L94A	1.88
M	L163E	0.7

The binding free energy of the wild-type and mutant complexes were further estimated using the MM/GBSA approach which calculates $\Delta\Delta G_{bind}$ based on molecular dynamics simulation of the protein-protein complex. The prediction which was achieved using HawkDock is intermediate both in accuracy and computational effort between strict alchemical perturbation and empirical scoring methods. The output revealed the total binding energy scores on per-residue bases for both wild-type and mutant complexes (Tables 5 and 6). Detailed contribution of each residue in the complex can be accessed from the Supplementary Tables 1-6.

Table 5. HawkDock-MM/GBSA per-residue binding energy table for the wild-type protein-protein complexes. The output displays the mutated chains, residue positions, van der Waals potential, electrostatic potential, generalized born scores, solvent accessibility and the total binding energy score.

Chain(s)	**Position**	**Residue(s)**	**VDW**	**ELE**	**GB**	**SA**	**Total (Kcal/mol)**

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

H	119	LEU	-3.88	-2.80	2.74	-0.68	-4.62
H	225	VAL	-0.06	-0.36	0.31	0	-0.11
H	233	LEU	-0.04	0.17	-0.15	0	-0.02
H	234	LYS	-1.73	1.37	-0.27	-0.35	-0.98
H	238	LEU	-1.11	1.63	-1.02	-0.26	-0.75
M	94	LEU	-3.54	-1.07	1.98	-0.55	-3.18
M	163	LEU	-2.38	-0.58	0.7	-0.36	-2.62

Table 6. HawkDock-MM/GBSA per-residue binding energy table for the mutant protein-protein complexes. The output displays the mutated chains, residue positions, van der Waals potential, electrostatic potential, generalized born scores, solvent accessibility and the total binding energy score.

Chain(s)	**Position**	**Residue(s)**	**VDW**	**ELE**	**GB**	**SA**	**Total (Kcal/mol)**
H	119	ALA	-1.88	-1.59	1.70	-0.37	-2.14
H	225	ALA	-0.04	-0.38	0.43	0	0.01
H	233	ALA	-0.02	0.37	-0.35	0	0
H	234	ALA	-3.60	71.35	-66.21	-0.74	0.80
H	238	ALA	-1.32	1.66	-1.05	-0.36	-1.07
M	94	ALA	-1.07	-1.55	2.04	-0.16	-0.74
M	163	GLU	-2.82	-57.81	62.05	-0.40	1.01

Interatomic interaction analysis

Protein-protein interactions are essential for regular biological processes and for the regulation of cellular reactions that affect the function and expression of genes. Several studies [49] have elucidated the role of protein-protein complex interface residues in conferring specificity and stability. Interface residues of proteins are known to interact with main chain and side chain atoms of their interacting partners. However, the impact and relative contribution of inter-protein interactions involving interface residue as compared to intra-protein interactions in protein-protein complexes are unclear. In order to ensure that essential interactions involved in the binding affinity and stability of the hypothetical *hs* CENP-HIKM complex are not overlooked, we report the observed changes in interatomic interactions of the wild-type and mutant protein complex subunits (Table 7 and 8). A comparative study of the wild-type and mutant protein complexes showed that both inter- and intra-model interactions contributed to the stability of the complex (Supplementary Figures S8-S10). Upon the mutation of each residue, a dramatic loss of specific interatomic interactions (van der Waals interactions, hydrogen bond interactions and hydrophobic interactions) were observed, which speculatively led to the reported reduction in the experimental and predicted binding affinity of the mutants.

Table 7. Tabular representation of the interatomic interactions for the wild-type hypothetical *hs* CENP-HIKM complex. The first and second column shows the chains and residues of interest, while the remaining columns represent the observed interaction types (van der Waals, hydrogen bond and hydrophobic interactions respectively). Specific chains and residues that forms a interaction with the residues of interest (wild-type residues) are presented in corresponding columns for each interaction types.

Chain(s)	**Residue(s)**	**VDW**	**H-bond**	**Hydrophobic**
H	LEU-219	H: GLN-217 H: LEU-221 H: SER-223	H: GLN-217; K: HIS-187 H: GLY-222 H: SER-223	K: PHE-183 K: HIS-187 K: PHE-188
H	VAL-225	H: SER-223	H: ILE-220 H: SER-223 H: TRP-227	H: ILE-220 H: LEU-223

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

H	LEU-233	H: GLU-235	H: ASP-230 H: ILE-236 H: VAL-237	H: ILE-220 H: VAL-225 H: ASP-230 H: ILE-236
H	LYS-234	H: TRP-227; I: GLY-186 H: ASP-230; I: PHE-187 H: ALA-232; I: PHE-189	H: TRP-227; I: LEU-183 H: ASP-230 H: PRO-231 H: LEU-238	H: TRP-227; I: PHE-187 I: PHE-189
H	LEU-238	H: LEU-240; I: GLN-180 I: LEU-183	H: LYS-234; I: LEU-183 H: GLU-235	I: VAL-145 I: LEU-183 I: PHE-187
M	LEU-94	M: ALA-90 M: SER-91 M: PHE-92	M: ALA-90 M: SER-91 M: LYS-96 M: LEU-163	M: LEU-163; I: ARG-281 M: LEU-166; I: TRP-461
M	LEU-163	M: SER-161 M: LEU-165	M: LEU-94; I: TRP-461 M: SER-161 M: LEU-166	M: LEU-94; I: CYS-426

Table 8. Tabular representation of the interatomic interactions for the mutant hypothetical *hs* CENP-HIKM complex. The first and second column shows the mutated chains and residues of interest, while the remaining columns represent the observed interaction types (van der Waals, hydrogen bond and hydrophobic interactions respectively). Specific chains and residues that interact with the mutant residues are presented in corresponding columns for each interaction types.

Chain(s)	**Residue(s)**	**VDW**	**H-bond**	**Hydrophobic**
H	ALA-219	H: PHE-216 H: GLN-217 H: LEU-221 H: SER-223	H: VAL-215 H: PHE-216 H: SER-223	K: PHE-183 K: PHE-188
H	ALA-225	H: ILE-220	H: ILE-220	H: ILE-220
H	ALA-233	H: GLU-235	H: ASP-230 H: ILE-236 H: VAL-237	H: TRP-227 H: ASP-230
H	ALA-234	H: ASP-230; I: LEU-183 H: ALA-232 H: VAL-237 H: LEU-238	H: TRP-227; I: LEU-183 H: ASP-230 H: PRO-231 H: LEU-238	I: PHE-189
H	ALA-238	H: GLU-235 H: LYS-242	H: LYS-234 H: GLU-235 H: VAL-237	I: LEU-183 I: PHE-187
M	ALA-94	M: ALA-90 M: SER-91 M: PHE-92 M: LYS-96 M: LEU-166	M: ALA-90 M: SER-91 M: LYS-96 M: LEU-163	M: LEU-163; I: TRP-461 M: LEU-166

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

M	GLU-163	M: SER-161 M: LEU-165 M: LEU-166	M: LEU-94; I: TRP-461 M: SER-161 M: LEU-166	M: LEU-94

Discussion

Being a busy environment, thousands of molecules constantly interact in the cell and through information exchange define the cellular metabolic state. Among all cellular homeostasis contributors, proteins are both the most active and most abundant [50] therefore, understanding their interactions and delineating their information sharing mechanism is essential for a detailed comprehension of cellular functionality. This further provides the first approach towards rational therapeutic agent development against many incapacitating or deadly diseases [51]. Despite the advances in structure determination through experimental methods, most of the known protein-protein interactions still have no atomic structure. NMR spectroscopy and X-ray crystallography, both of which are high resolution techniques struggle with high-throughput demand, while low resolution methods like the small-angle X-ray scattering and cryo-electron microscopy provide excessively coarse data. The development of molecular docking or computational structure prediction was first aimed at complementing experimental results but has since developed into a lively and independent research field [52].

Elucidating the organization and structural architecture of the CCAN is crucial for the understanding of the functionality and assembly of the kinetochore. The CENP-H, -I, -K and -M, among other subunits of the CCAN have previously been reported to form a stable complex based on reconstitution experiments and proteomic analyses [37, 53, 54, 55, 56, 57]. Our study for the first time present a computationally modeled high quality structure of the human CENP-HIKM complex (Figure 6) alongside a detailed report of the inter- and intra-residue interactions. Previously reported computational model of the*hs* CENP-I suggests that it assumes a fold in form of an α-solenoid which shares resemblance with the folding of β-importin [37, 58, 59]. The *hs* CENP-I N-terminal domain (composed of residues 57-281) was also reported to be sufficient enough for the binding of the*hs* CENP-H and *hs* CENP-K while the *hs* CENP-M sufficiently binds to the C-terminal domain. Contiguity between CENP-H, -I, and -K was hypothesized on the basis of proteomic analysis involving precipitates from phenotypic similarities as a result of individual subunit depletion, from 2-hybrid interaction data and from cell lysates [13, 60]. Additional analyses suggest that the revealed complex interaction is a representation of the evolutionarily conserved assembling mechanism of the CENP-HIK complex [14].

Structures of biologically essential proteins are consistently on a high demand, especially the large proteins and those that are members of complex systems. It is however not always feasible, for numerous reasons, to experimentally generate high resolution structures using the NMR, cryo-electron microscopy or X-ray crystallography. Among the numerous challenges are the poor diffraction of crystals, high aggregation and low stability of proteins [61]. *In silico*molecular modeling in this situation can provide a high quality alternative for experimental research. One of the most challenging computational biology problems has been shown to be the De novo structure prediction of proteins only from amino acid sequences [32]. Recent advances in the field has revealed that some accurately-predicted long range contacts may permit correct topology-level structural modeling [62] and that the DCA (direct evolutionary coupling analysis) for most multiple sequence alignments may generate appreciable amount of long range native contacts for protein-protein interactions and proteins with a large number of homologous sequences [63, 64]. We have therefore employed the contact-assisted folding of proteins and contact prediction in the modeling of each subunit of the *hs* CENP-HIK 3D structure (Figure 1, Supplementary Figures S1 and S2).

Significant improvement has been made towards the generation of potential protein-protein interaction networks with the use of mass spectrometry, yeast two-hybrid assays [65] and high-throughput proteomics studies [66, 67]. X-ra crystallography-obtained atomic-level details are frequently required for the mechanistic interpretation of observed interactions. However, the occurrence of most biologically relevant interactions

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

are in transient protein complexes, which makes the experimental determination of their structures largely difficult, even when structures of the interacting partners are known. Computational docking approaches have therefore been designed for the structural prediction of protein complexes with an accuracy similar to that provided by X-ray crystallography [68, 69]. A substantial amount of models with well defined atomic positions are usually generated after protein-protein docking protocols, but the currently available scoring functions possess low predictive accuracy for a reliable discrimination of models, and most often, models closest to the native structure are not easily detected solely through computational tools [69]. However, our near-native model selection in this study was guided by the architectural similarity of each generated model with the fungal and yeast orthologs of the protein complex, previously reported to be evolutionarily conserved (Figure 5).

The main cellular functions such as DNA replication, transcription, translation, protein folding and turnover, are directed by large macromolecular complexes such as proteasomes, chaperonins, ribosomes and polymerases. The mechanism of action of these macromolecules are often dynamical and require collective and large conformational changes [70]. Normal mode analysis is an approach that can be used for the description of the accessible flexible states of a protein around an equilibrium position based on small oscillation physics. When a macromolecule in a minimum energy conformation is perturbed slightly, a force is activated to restore the system back to its state of equilibrium [71]. There is always an equal division of vibrational energy in the system so that all vibrational modes have equal energy and the average amplitude of oscillation for any given mode scales as the inverse of its frequency. Thus, higher frequency modes with energetically greater displacement typically describe fast but small local amplitude movement relatively involving fewer atoms, while lower frequency modes describe slow displacements and changes in conformation on a large scale with the involvement of larger number of atoms [72]. Coarse-grained models merged with normal mode analysis has proven to be a popular and powerful substitute for the collective motion simulation of macromolecular complexes at extended timescales. In addition to the conformational sampling and motion dynamics visualization (Supplementary Figure S4 and S5), the normal mode analysis result also suggest that the hypothetical protein model assumes a stable conformation (Figure 7).

An essential prerequisite for regular biological function is the ability of a protein to establish inordinately selective interactions with its macromoleclar partner. Sequence mutations that changes protein interactions may lead to a complete functional abolishment or result into a significant perturbation [73]. A feasible method to evaluate mutational effect on the binding affinity of proteins is to experimentally quantify it. However, while site-directed mutagenesis methodologies are fast and inexpensive, FRET, isothermal titration calorimetry, surface plasmon resonance and other methods used for binding affinity measurements can be costly and time-consuming [74]. We have therefore directed computational approaches towards the prediction of binding affinity changes upon mutation (Tables 1-6, Supplementary Tables 1-6), which has shown great consistency with results from earlier reported experimental mutagenesis studies. Our interatomic interaction visualization study also provided insights into the molecular nature of the studied interactions and likewise the comprehension of the functional and structural impact of each mutation (Tables 7 and 8, Supplementary Figures S8-S10).

Conclusion

With the aid of extensive computational approaches and following experimentally validated site-directed mutagenesis from literature reports, we have designed an hypothetical model of the*hs* CENP-HIKM complex. Structurally refined models of each subunit were individually docked to generate an hypothetical complex which was subjected to several *in silico* protocols such as, the normal mode analysis, *in silico* mutagenesis, binding free energy prediction upon mutation, and analysis of the non-covalent interactions, in an attempt to validate the model reliability. Knowledge of the*hs* CENP-HIKM architecture and the surface residues at the interaction site as presented in this study may provide more insight into the mechanisms of abnormal interactions in disease states, through the comprehension of simple molecular recognition mechanisms. Such information may present future therapeutic potentials for rational development of drugs that regulate or mimic the effects of protein-protein interactions.

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Declarations

Ethics approval and consent to participate Not applicable**Consent for publication** Not applicable**Availability of data and material** Not applicable**Competing Interest** Authors declare no competing interest.**Funding** Authors received no funding for this project from any organization.

Authors Contribution

OAD: Literature review and manuscript writing

All authors have read and approved the manuscript

Acknowledgement

Not applicable

References

1. Vidal M, Cusick ME & Barabási A-L (2011) Interactome networks and human disease. *Cell* **144** , 986– 998
2. Von Eichborn J, Gunther S & Preissner R (2010) Structural features and evolution of protein–protein interactions. *Genome Inform* **22** , 1– 10.
3. Kolodny R, Pereyaslavets L, Samson AO & Levitt M (2013) On the universe of protein folds. *Annu Rev Biophys* **42** , 559– 582.
4. Schlick T, Portillo-Ledesma S, Myers CG, Beljak L, Chen J, Dakhel S, Darling D, Ghosh S, Hall J, Jan M, Liang E, Saju S, Vohr M, Wu C, Xu Y, Xue E. Biomolecular Modeling and Simulation: A Prospering Multidisciplinary Field. Annu Rev Biophys. 2021 May 6;50:267-301. doi: 10.1146/annurev-biophys-091720-102019. Epub 2021 Feb 19. PMID: 33606945; PMCID: PMC8105287.
5. Rodrigues JP, Bonvin AM. Integrative computational modeling of protein interactions. FEBS J. 2014 Apr;281(8):1988-2003. doi: 10.1111/febs.12771. Epub 2014 Mar 26. PMID: 24588898.
6. Klare K, Weir JR, Basilico F, Zimniak T, Massimiliano L, Ludwigs N, Herzog F, Musacchio A. CENP-C is a blueprint for constitutive centromere-associated network assembly within human kinetochores. J Cell Biol. 2015 Jul 6;210(1):11-22. doi: 10.1083/jcb.201412028. Epub 2015 Jun 29. PMID: 26124289; PMCID: PMC4494010.
7. Cheerambathur D.K., and Desai A.. 2014. Linked in: formation and regulation of microtubule attachments during chromosome segregation. *Curr. Opin. Cell Biol.* 26:113–122. 10.1016/j.ceb.2013.12.005.
8. Fukagawa T., and Earnshaw W.C.. 2014. The centromere: chromatin foundation for the kinetochore machinery. *Dev. Cell.* 30:496–508. 10.1016/j.devcel.2014.08.016.
9. Cheeseman I.M. 2014. The kinetochore. *Cold Spring Harb. Perspect. Biol.* 6:a015826 10.1101/cshperspect.a015826.
10. Cheeseman I.M., Chappie J.S., Wilson-Kubalek E.M., and Desai A.. 2006. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell.* 127:983–997. 10.1016/j.cell.2006.09.039.
11. DeLuca J.G., Gall W.E., Ciferri C., Cimini D., Musacchio A., and Salmon E.D.. 2006. Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. *Cell.* 127:969–982. 10.1016/j.cell.2006.09.047.
12. Izuta H., Ikeno M., Suzuki N., Tomonaga T., Nozaki N., Obuse C., Kisu Y., Goshima N., Nomura F., Nomura N., and Yoda K.. 2006. Comprehensive analysis of the ICEN (Interphase Centromere Complex) components enriched in the CENP-A chromatin of human cells. *Genes Cells.* 11:673–684. 10.1111/j.1365-2443.2006.00969.x
13. Okada M., Cheeseman I.M., Hori T., Okawa K., McLeod I.X., Yates J.R. III, Desai A., and Fukagawa T.. 2006. The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. *Nat. Cell Biol.* 8:446–457. 10.1038/ncb1396.
14. Hu, L., Huang, H., Hei, M., Yang, Y., Li, S., Liu, Y., Dou, Z., Wu, M., Li, J., Wang, G. Z., Yao, X., Liu, H., He, X., & Tian, W. (2019). Structural analysis of fungal CENP-H/I/K homologs reveals a

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

conserved assembly mechanism underlying proper chromosome alignment. *Nucleic acids research* , *47* (1), 468–479. https://doi.org/10.1093/nar/gky1108.

15. Foltz D.R., Jansen L.E.T., Black B.E., Bailey A.O., Yates J.R.r., Cleveland D.W.. The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* 2006; **8** :458–469.
16. Westermann S., Schleiffer A.. Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol.* 2013; **23** :260–269.
17. Amano M., Suzuki A., Hori T., Backer C., Okawa K., Cheeseman I.M., Fukagawa T.. The CENP-S complex is essential for the stable assembly of outer kinetochore structure. *J. Cell Biol.* 2009; **186** :173–182.
18. Black B.E., Brock M.A., Bedard S., Woods V.L., Cleveland D.W.. An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. *Proc. Natl. Acad. Sci. U.S.A.* 2007; **104** :5008–5013.
19. Chittori S., Hong J., Saunders H., Feng H., Ghirlando R., Kelly A.E., Bai Y., Subramaniam S.. Structural mechanisms of centromeric nucleosome recognition by the kinetochore protein CENP-N. *Science* . 2018; **359** :339–343.
20. Falk S.J., Guo L.Y., Sekulic N., Smoak E.M., Mani T., Logsdon G.A., Gupta K., Jansen L.E.T., Van Duyne G.D., Vinogradov S.A. et al. . Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* . 2015; **348** :699–703.
21. Musacchio A., Desai A.. A molecular view of kinetochore assembly and function. *Biology (Basel)* . 2017; **6** :5.
22. McKinley K.L., Cheeseman I.M.. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.* 2016; **17** :16–29.
23. Huis In 't Veld P.J., Jeganathan S., Petrovic A., Singh P., John J., Krenn V., Weissmann F., Bange T., Musacchio A.. Molecular basis of outer kinetochore assembly on CENP-T. *Elife* . 2016; **5** :e21007.
24. Petrovic A., Keller J., Liu Y.H., Overlack K., John J., Dimitrova Y.N., Jenni S., van Gerwen S., Stege P., Wohlgemuth S. et al. . Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. *Cell* . 2016; **167** :1028–1040.
25. Hara M., Fukagawa T.. Kinetochore assembly and disassembly during mitotic entry and exit. *Curr. Opin. Cell Biol.* 2018; **52** :73–81.
26. Tian T., Li X.R., Liu Y.Y., Wang C.L., Liu X., Bi G.Q., Zhang X., Yao X.B., Zhou Z.H., Zang J.Y.. Molecular basis for CENP-N recognition of CENP-A nucleosome on the human kinetochore. *Cell Res.* 2018; **28** :374–378.
27. Pentakota S., Zhou K., Smith C., Maffini S., Petrovic A., Morgan G.P., Weir J.R., Vetter I.R., Musacchio A., Luger K.. Decoding the centromeric nucleosome through CENP-N. *Elife* . 2017; **6** :e33442.
28. Kim S., Yu H.T.. Multiple assembly mechanisms anchor the KMN spindle checkpoint platform at human mitotic kinetochores. *J. Cell Biol.* 2015; **208** :181–196.
29. McKinley K.L., Sekulic N., Guo L.Y., Tsinman T., Black B.E., Cheeseman I.M.. The CENP-L-N complex forms a critical node in an integrated meshwork of interactions at the Centromere-Kinetochore interface. *Mol. Cell* . 2015; **60** :886–898.
30. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2021 Feb 07]. Available from: https://www.ncbi.nlm.nih.gov/
31. The Protein Data Bank H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne (2000) *Nucleic Acids Research* , **28** : 235-242. doi:10.1093/nar/28.1.235.
32. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. PLoS Comput Biol. 2017 Jan 5;13(1):e1005324. doi: 10.1371/journal.pcbi.1005324. PMID: 28056090; PMCID: PMC5249242.
33. Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic acids research* , *41* (Web Server issue), W384–W388. https://doi.org/10.1093/nar/gkt458
34. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

errors in three-dimensional structures of proteins. *Nucleic acids research* , *35* (Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290

35. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996 Dec;8(4):477-86. doi: 10.1007/BF00228148. PMID: 9008363.
36. Yuan, S., Chan, H. C. S., & Hu, Z. (2017). *Using PyMOL as a platform for computational drug design. Wiley Interdisciplinary Reviews: Computational Molecular Science, 7(2), e1298.* doi:10.1002/wcms.1298
37. Basilico, F., Maffini, S., Weir, J. R., Prumbaum, D., Rojas, A. M., Zimniak, T., De Antoni, A., Jeganathan, S., Voss, B., van Gerwen, S., Krenn, V., Massimiliano, L., Valencia, A., Vetter, I. R., Herzog, F., Raunser, S., Pasqualato, S., & Musacchio, A. (2014). The pseudo GTPase CENP-M drives human kinetochore assembly. *eLife* , *3* , e02978. https://doi.org/10.7554/eLife.02978
38. Ashkenazy H., Abadi S., Martz E., Chay O., Mayrose I., Pupko T., and Ben-Tal N. 2016 ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucl. Acids Res.* 2016; DOI: 10.1093/nar/gkw408; PMID: 27166375
39. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. *Nature Protocols.* 2017 Feb;12(2):255-278.
40. Lopez-Blanco JR, Garzon JI, Chacon P. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics. 2011 Oct 15;27(20):2843-50. doi: 10.1093/bioinformatics/btr497. Epub 2011 Aug 27. PMID: 21873636.
41. Carlos HM Rodrigues, Douglas EV Pires, David B Ascher, DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability, *Nucleic Acids Research* , Volume 46, Issue W1, 2 July 2018, Pages W350–W355, https://doi.org/10.1093/nar/gky300
42. Shapovalov MV, Dunbrack RL Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure. 2011 Jun 8;19(6):844-58. doi: 10.1016/j.str.2011.03.019. PMID: 21645855; PMCID: PMC3118414.
43. Dehouck, Y., Kwasigroch, J. M., Rooman, M., & Gilis, D. (2013). BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic acids research* , *41* (Web Server issue), W333–W339. https://doi.org/10.1093/nar/gkt450.
44. Carlos H M Rodrigues, Yoochan Myung, Douglas E V Pires, David B Ascher, mCSM-PPI2: predicting the effects of mutations on protein–protein interactions, *Nucleic Acids Research* , Volume 47, Issue W1, 02 July 2019, Pages W338–W344, https://doi.org/10.1093/nar/gkz383
45. Rodrigues CHM, Pires DEV, Ascher DB. mmCSM-PPI: predicting the effects of multiple point mutations on protein-protein interactions. Nucleic Acids Res. 2021 Apr 24:gkab273. doi: 10.1093/nar/gkab273. Epub ahead of print. PMID: 33893812.
46. Li, M., Simonetti, F.L., Goncearenco, A. and Panchenko, A.R. (2016) MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. Nucleic Acids Res, 44, W494-501.
47. Weng GQ, Wang EC, Wang Z, Liu H, Li D, Zhu F, Hou TJ. HawkDock: a web server to predict and analyze the structures of protein-protein complexes based on computational docking and MM/GBSA. *Nucleic Acids Research* , 2019, 47(W1): W322-W330.
48. Jubb, H. C., Higueruelo, A. P., Ochoa-Montano, B., Pitt, W. R., Ascher, D. B., & Blundell, T. L. (2017). Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *Journal of molecular biology* , *429* (3), 365–371. https://doi.org/10.1016/j.jmb.2016.12.004
49. Jayashree, S., Murugavel, P., Sowdhamini, R. *et al.* Interface residues of transient protein-protein complexes have extensive intra-protein interactions apart from inter-protein interactions. *Biol Direct* **14,** 1 (2019). https://doi.org/10.1186/s13062-019-0232-2
50. Mosca R, Ceol A & Aloy P (2013) Interactome3D: adding structural details to protein networks. *Nat Method* **10** , 47– 53.
51. Vangone A, Cavallo L & Oliva R (2013) Using a consensus approach based on the conservation of inter-residue contacts to rank CAPRI models. *Proteins* **81** , 2210– 2220.
52. Karaca E, Bonvin AM. Advances in integrative modeling of biomolecular complexes. Methods. 2013

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Mar;59(3):372-81. doi: 10.1016/j.ymeth.2012.12.004. Epub 2012 Dec 23. PMID: 23267861.

53. Hori T., Amano M., Suzuki A., Backer C.B., Welburn J.P., Dong Y., McEwen B.F., Shang W.H., Suzuki E., Okawa K. et al. . CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* . 2008; **135** :1039–1052.
54. Carroll C.W., Milks K.J., Straight A.F.. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol.* 2010; **189** :1143–1155.
55. Gascoigne K.E., Takeuchi K., Suzuki A., Hori T., Fukagawa T., Cheeseman I.M.. Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* . 2011; **145** :410–422.
56. Nishino T., Takeuchi K., Gascoigne K.E., Suzuki A., Hori T., Oyama T., Morikawa K., Cheeseman I.M., Fukagawa T.. CENP-T-W-S-X forms a unique centromeric chromatin structure with a Histone-like fold. *Cell* . 2012; **148** :487–501.
57. Petrovic A., Keller J., Liu Y.H., Overlack K., John J., Dimitrova Y.N., Jenni S., van Gerwen S., Stege P., Wohlgemuth S. et al. . Structure of the MIS12 complex and molecular basis of its interaction with CENP-C at human kinetochores. *Cell* . 2016; **167** :1028–1040.
58. Cingolani G, Petosa C, Weis K, Muller CW. 1999. Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* 399:221–229. doi: 10.1038/20367.
59. Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A. 1999. Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell* 97:635–646. doi: 10.1016/S0092-8674(00)80774-6.
60. Measday V, Hailey DW, Pot I, Givan SA, Hyland KM, Cagney G, Fields S, Davis TN, Hieter P. 2002. Ctf3p, the Mis6 budding yeast homolog, interacts with Mcm22p and Mcm16p at the yeast outer kinetochore. *Genes & Development* 16:101–113. doi: 10.1101/gad.949302.
61. McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallogr F Struct Biol Commun. 2014 Jan;70(Pt 1):2-20. doi: 10.1107/S2053230X13033141. Epub 2013 Dec 24. PMID: 24419610; PMCID: PMC3943105.
62. Kim DE, DiMaio F, Yu-Ruei Wang R, Song Y, Baker D. One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. Proteins: Structure, Function, and Bioinformatics. 2014;82(S2):208–18.
63. de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. Nature Reviews Genetics. 2013;14(4):249–61. pmid:23458856.
64. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. P Natl Acad Sci USA. 2009;106(1):67–72.
65. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A. 2001 Apr 10;98(8):4569-74. doi: 10.1073/pnas.061034498. Epub 2001 Mar 13. PMID: 11283351; PMCID: PMC31875.
66. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature. 2002 Jan 10;415(6868):180-3. doi: 10.1038/415180a. PMID: 11805837.
67. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 2007;3:89. doi: 10.1038/msb4100134. Epub 2007 Mar 13. PMID: 17353931; PMCID: PMC1847948.
68. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol. 2002 Feb;12(1):28-35. doi: 10.1016/s0959-440x(02)00285-3. PMID: 11839486.

Posted on Authorea 22 Jul 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.162693577.78775249/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

69. Ritchie DW. Recent progress and future directions in protein-protein docking. Curr Protein Pept Sci. 2008 Feb;9(1):1-15. doi: 10.2174/138920308783565741. PMID: 18336319.
70. Lopez-Blanco, J. R., Aliaga, J. I., Quintana-Orti, E. S., & Chacon, P. (2014). iMODS: internal coordinates normal mode analysis server. *Nucleic acids research* , *42* (Web Server issue), W271–W276. https://doi.org/10.1093/nar/gku339
71. Mahajan S, Sanejouand YH. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. Arch Biochem Biophys. 2015 Feb 1;567:59-65. doi: 10.1016/j.abb.2014.12.020. Epub 2015 Jan 3. PMID: 25562404.
72. Bauer JA, Pavlović J, Bauerová-Hlinková V. Normal Mode Analysis as a Routine Part of a Structural Investigation. Molecules. 2019 Sep 10;24(18):3293. doi: 10.3390/molecules24183293. PMID: 31510014; PMCID: PMC6767145.
73. Stefl S., Nishi H., Petukh M., Panchenko A.R., Alexov E. Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* 2013;**425** :3919–3936.
74. Wainreb G., Wolf L., Ashkenazy H., Dehouck Y., Ben-Tal N. Protein stability: a single recorded mutation aids in predicting the effects of other mutations in the same amino acid site. *Bioinformatics.* 2011;**27** :3286–3292.