


## RESEARCH ARTICLE

# Modeling post-translational modifications and cancer-associated mutations that impact the heterochromatin protein 1 $\alpha$ -importin $\alpha$ heterodimers

Michael T. Zimmermann<sup>1,2</sup> | Monique M. Williams<sup>3,4</sup> | Eric W. Klee<sup>3,4</sup> |  
Gwen A. Lomber<sup>5,6,7</sup> | Raul Urrutia<sup>5,7,8</sup> 

<sup>1</sup>Bioinformatics Research and Development Laboratory, and Precision Medicine Simulation Unit, Genomic Science and Precision Medicine Center (GSPMC), Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>2</sup>Clinical and Translational Sciences Institute, Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>3</sup>Department of Biochemistry, Mayo Clinic, Rochester, Minnesota

<sup>4</sup>Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota

<sup>5</sup>Division of Research, Department of Surgery, Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>6</sup>Department of Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>7</sup>Genomic Science and Precision Medicine Center (GSPMC), Medical College of Wisconsin, Milwaukee, Wisconsin

<sup>8</sup>Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin

## Correspondence

Gwen A. Lomber and Raul Urrutia, Division of Research, Department of Surgery, Medical College of Wisconsin, Milwaukee, WI.  
Email: glomberk@mcw.edu (G. A. L.) and Email: rurrutia@mcw.edu (R. U.)

## Funding information

National Cancer Institute, Grant/Award Number: R01 CA178627; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Number: R01 DK52913; National Institutes of Health

## Abstract

Heterochromatin protein 1 $\alpha$  (HP1 $\alpha$ ) is a protein that mediates cancer-associated processes in the cell nucleus. Proteomic experiments, reported here, demonstrate that HP1 $\alpha$  complexes with importin  $\alpha$  (IMP $\alpha$ ), a protein necessary for its nuclear transport. This data is congruent with Simple Linear Motif (SLiM) analyses that identify an IMP $\alpha$ -binding motif within the linker that joins the two globular domains of this protein. Using molecular modeling and dynamics simulations, we develop a model of the IMP $\alpha$ -HP1 $\alpha$  complex and investigate the impact of phosphorylation and genomic variants on their interaction. We demonstrate that phosphorylation of the HP1 $\alpha$  linker likely regulates its association with IMP $\alpha$ , which has implications for HP1 $\alpha$  access to the nucleus, where it functions. Cancer-associated genomic variants do not abolish the interaction of HP1 $\alpha$  but instead lead to rearrangements where the variant proteins maintain interaction with IMP $\alpha$ , but with less specificity. Combined, this new mechanistic insight bears biochemical, cell biological, and biomedical relevance.

## KEYWORDS

cancer, CBX5, functional genomics, HP1 $\alpha$ , molecular modeling

## 1 | INTRODUCTION

Three known HP1 isoforms, HP1 $\alpha$  (CBX5), HP1 $\beta$  (CBX1), and HP1 $\gamma$  (CBX3), are critical for regulating gene expression networks that are crucial for normal embryonic development<sup>1</sup> and cancer associated

processes, including differentiation,<sup>2,3</sup> cell proliferation,<sup>4,5</sup> cell cycle control,<sup>4,6-9</sup> apoptosis,<sup>8,10</sup> and DNA damage response.<sup>11-18</sup> In fact, a conservation in sequence, structure, and function among these proteins in organisms ranging from *Drosophila melanogaster* to human reflects the biological importance and biomedical implications of their

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Proteins: Structure, Function, and Bioinformatics* published by Wiley Periodicals, Inc.

alterations. Alterations in HP1 proteins range from overexpression, somatic mutations, CNVs, and abnormalities in the function of their upstream regulators. Early studies performed by our group and others have demonstrated that signaling-mediated mechanisms, yet to be fully understood, lead to the extensive post-translational modifications of HP1 proteins. Some of these modifications fall in the globular domains of these proteins; the globular domains read histone marks and are responsible for homodimerization and heterodimerization. These domains are joined by an intrinsically disordered region known as the linker, which serves as an integrator of phosphorylation-mediated signaling events. However, some of the detailed structural-functional relationships of the linker domain are by far less understood than those of the globular domains. Thus, the current study uses the HP1 $\alpha$  linker domain, as a model for extending our understanding of how these proteins can be regulated via the linker domain through heterodimerization events that ultimately lead to the translocation of this protein to the nucleus, the cellular region where it is functionally needed. Moreover, we model how post-translational modifications, deposited in the linker region in response to upstream regulators such as cancer-associated mutations, alter linker structure and thereby bonding patterns during interphase, likely impacting import and the downstream pathway.

Our findings indicate that post-translational modifications of the HP1 $\alpha$  linker are a critical factor in altering interaction with IMP $\alpha$ , and that genomic variants alter the pattern of interactions making them less specific. We used computational molecular modeling and dynamics simulation, and structural bioinformatics tools to significantly expand our knowledge of the intramolecular and intermolecular behavior of HP1 $\alpha$  in direct relation to biological processes that are crucial for the maintenance of genomic integrity. This work significantly extends our understanding of the dynamic behavior of HP1 $\alpha$  and how it achieves intermolecular interactions outside of the chromoshadow domain. Additionally, our approach to use multiple computational measures that help to inform one another emphasizes that each metric should not be interpreted in isolation; they should be jointly considered to better understand what they indicate about the molecule. Most importantly, we have gained insight for how HP1 $\alpha$  alterations may disrupt signals relayed through mitogenic signaling pathways and heterochromatin regulation during each cell cycle. Thus, these results must be taken into consideration as mechanisms that are likely to influence the function of HP1 proteins during development, homeostasis control, and disease.

## 2 | MATERIALS AND METHODS

### 2.1 | Multiple sequence alignment

Protein sequences for the human and mouse HP1 family members, HP1 $\alpha$  (CBX5), HP1 $\gamma$  (CBX3), and HP1 $\beta$  (CBX1) were downloaded from UniProt<sup>19</sup> and aligned to each other using standard parameters of Clustal-Omega<sup>20</sup> at the European Bioinformatics Institute. Amino acid equivalences between isoforms used in molecular modeling were taken from the multiple sequence alignment (MSA). The sequence identity of the proteins was assessed by pairwise alignment via the

EMBOSS-Needle open-source software platform.<sup>21</sup> HP1 $\alpha$  and HP1 $\beta$  demonstrated 76.9% residue similarity, while HP1 $\alpha$  and HP1 $\gamma$  share 66.7% residue similarity. The MSA is available in Table S1 and was used in the construction of the protein models.

### 2.2 | Linear motif analysis for nuclear localization signal and phosphorylation site prediction

The canonical binding mode of the nuclear importin receptor protein, IMP $\alpha$ , and nuclear localization signal (NLS) containing proteins was explored in the PDB and the literature.<sup>22-30</sup> The sequence of HP1 $\alpha$  was examined for NLS motifs using the prediction algorithms of PSORTII<sup>31</sup> and cNLS Mapper.<sup>32,33</sup> Both algorithms identified a bipartite NLS in the HP1 $\alpha$  linker, as previously predicted.<sup>34</sup> Potential phosphorylation sites in the HP1 $\alpha$  linker and the kinases that catalyze these modifications were identified using several in silico predictive servers including: NetPhos 3.1,<sup>35,36</sup> Kinase Phos 2.0,<sup>37</sup> DISPHOS 1.3,<sup>38</sup> GPS 3.0,<sup>39</sup> and PhosphoSVM.<sup>40</sup> The predictions were cross-referenced, and experimentally validated sites were assessed via PhosphositePlus<sup>41</sup> and PHOSIDA.<sup>42,43</sup> We also annotated the protein sequence using standard metrics and sequence-based linear motifs from multiple tools: InterProScan,<sup>44</sup> MESSA,<sup>45</sup> VADAR,<sup>46</sup> Molprobity,<sup>47-49</sup> DisEMBL,<sup>50</sup> IUPred2A,<sup>51</sup> and OnD-CRF.<sup>52</sup>

### 2.3 | Modeling of the HP1 $\alpha$ -IMP $\alpha$ complex

The crystal structure of the linker peptide which connects the chromodomain and chromoshadow domain has not been solved, likely due to its flexibility. Therefore, we built the model of the peptide using the sequence of residues 87-109 and the Builder Function in Discovery Studio<sup>53</sup> linking the chromodomain and chromoshadow domain. The sequence ID and 3D structures utilized to generate our models can be found in Table S2. We used PyMol<sup>54</sup> for molecular visualization.

The binding mode of a bipartite NLS to IMP $\alpha$  was previously demonstrated for the histone-binding protein N1N2 in *Xenopus laevis*<sup>55</sup> and now serves as a standard prototype of this intermolecular association. The experimental structure for N1N2 NLS was downloaded (PDB: 1PJN<sup>55</sup>) and used in threading to produce a model of the HP1 $\alpha$  linker bound to IMP $\alpha$ . We compared our threaded model to the results of full-length HP1 $\alpha$  docking to IMP $\alpha$  via ClusPro.<sup>56-60</sup>

### 2.4 | Immunoprecipitation and mass spectrometry

HP1 $\alpha$  immunoprecipitation was performed from HeLa cells. An antibody to HP1 $\alpha$  (Abcam, ab77256) or control IgG was conjugated to Protein A/G Magnetic Beads (ThermoFisher Scientific) through disuccinimidyl suberate (DSS) crosslinking. Cells were lysed with immunoprecipitation Lysis/Wash Buffer (Thermo Scientific-Pierce), and lysates were incubated with the conjugated antibodies overnight at 4°C. Immunoprecipitated complexes were washed, eluted and run on a 4%-15% Criterion Tris-HCl Protein Gel (Bio-Rad), which was subsequently visualized with BioSafe Coomassie Stain (Bio-Rad). Each gel

lane was divided into eight sections, destained, dehydrated, dried, and subjected to trypsin digestion. Liquid chromatography (LC)-ESI-MS/MS analysis was then performed on a Thermo Scientific LTQ Orbitrap mass spectrometer at the Mayo Clinic Proteomics Core.

## 2.5 | Modeled variants

We modeled experimentally relevant phosphorylated, nonphosphorylatable mutant, or phosphomimetic mutant forms of S92, S95, S97, and S103 within the HP1 $\alpha$  linker separately or in combination. As controls for binding, HP1 $\alpha$  NLS binding mutants were also generated by mutating each cluster of basic residues within the bipartite sequence to acidic residues. Additionally, rare single nucleotide variants in HP1 $\alpha$  that are rare in the currently healthy adult population (GnomAD minor allele frequency, MAF,  $\leq 1 \times 10^{-4}$ ) were mined from The Cancer Genome Atlas (TCGA),<sup>61</sup> the Catalogue Of Somatic Mutations In Cancer (COSMIC),<sup>62</sup> and the Exome Aggregation Consortium (ExAC) or GnomAD<sup>63</sup> (see Table S3). The cancer-related variants were then assessed for degree of pathogenicity using multiple sequence-based algorithms: Mutation Taster2,<sup>64</sup> Polyphen-2,<sup>65</sup> MutPred2,<sup>66</sup> and SIFT.<sup>67</sup> These algorithms did not provide equivalent annotations for each variant, (Table S4) and without any previous functional validation, variants with conflicting annotations were considered variants of uncertain significance (VUS). Computational mutagenesis was performed in Discovery Studio.<sup>53</sup> Variants are listed in Table S5.

## 2.6 | Molecular dynamics simulations

Molecular dynamics (MD) simulations were executed as in our previous studies.<sup>68</sup> Briefly, simulations for each condition were run in NAMD using the Charmm36 all-atom force-field<sup>69</sup> and a 1 fs integration time-step. The backbone C $\alpha$  atoms of IMP $\alpha$  were constrained using harmonic restraints. Models were simulated in the Generalized Born Molecular Volume (GBMV) implicit solvent model in Discovery Studio<sup>53</sup> at a dielectric constant of 80 and pH 7.4. Energy minimization commenced for 2000 steps with steepest descent. Hydrogen bonds were constrained using the SETTLE algorithm. Independent triplicate systems were initialized and heated to 300 K over 600 ps to mimic physiological conditions, equilibrated for 400 ps with subsequent production run for 10 ns (30 ns total for each condition).

## 2.7 | Statistical analysis

Prior to calculations, all trajectories were superimposed using C $\alpha$  atoms of IMP $\alpha$ . We calculated root-mean-square deviation (RMSD) and root-mean-square-fluctuation (RMSF) values to infer the flexibility and mobility of the peptides throughout the simulation and using C $\alpha$  atoms. The surface area of the IMP $\alpha$  receptor alone and the peptide alone were measured to uncover any differences in conformation or folding, and then buried solvent accessibility surface area (SASA) of the peptide in each condition was measured and plotted. Hydrogen bonds and salt bridges between the linker peptide and IMP $\alpha$  were calculated across the conditions and visualized as heat maps. We

analyzed hydrogen bonds that had a frequency  $\geq 0.5$  in at least three different conditions. The energy of interaction, encompassing van der Waals and electrostatic forces, was calculated to survey the interaction strength of each complex. The C $\alpha$  atoms of residue pairs involved in hydrogen bonding or salt bridge interactions between the peptide and IMP $\alpha$  were selected for distance metrics as simplified measures of interaction. Contact maps defining the intermolecular contacts between the HP1 $\alpha$  peptide across the trajectory were generated. Atoms were considered contacting if they were within 3.9 Å and the fraction of simulation time spent in contact was recorded. Principle component analysis (PCA) of the C $\alpha$  atoms was calculated in Cartesian space. All distributions were compared using permutation and *t* tests, as previously described.<sup>68</sup>

## 2.8 | Software

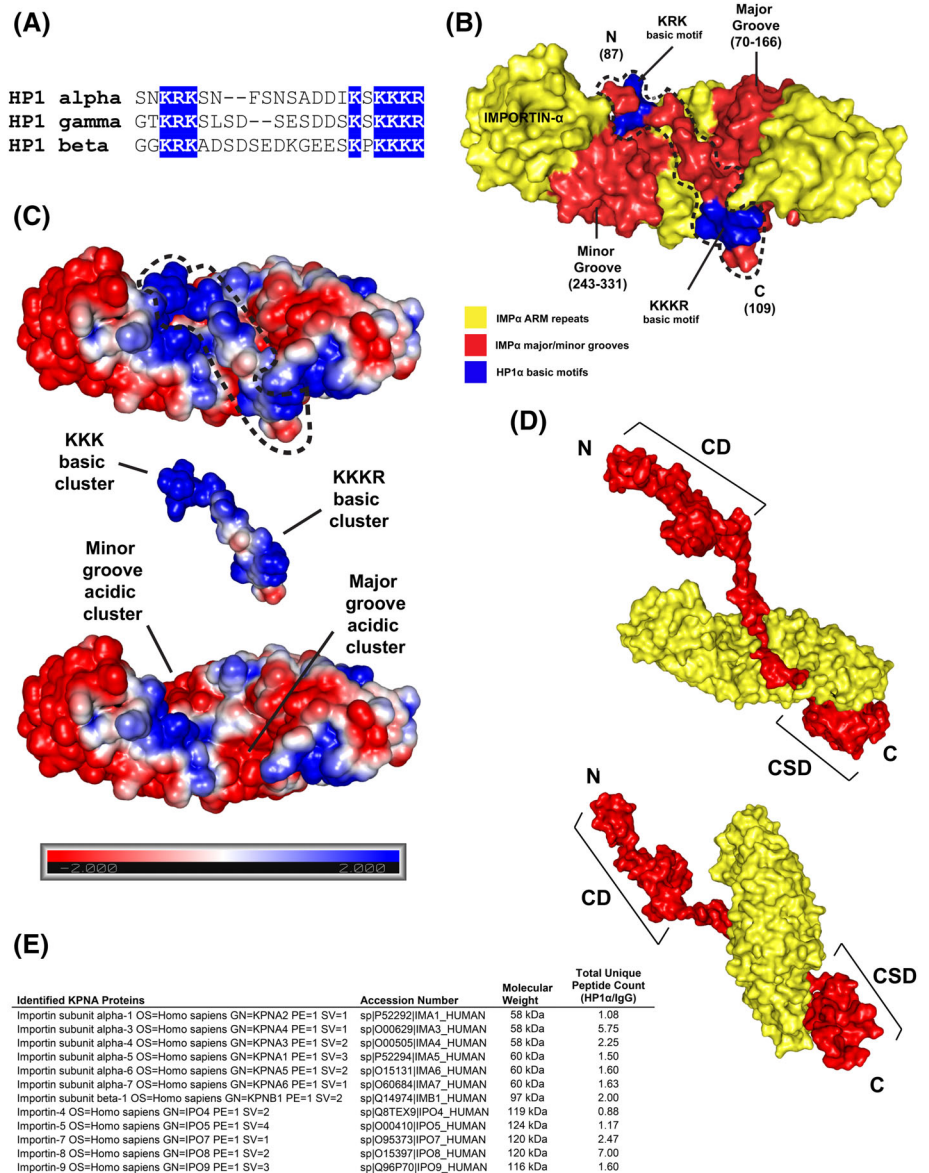
Initial protein models were constructed in Discovery Studio v2017<sup>53</sup> and simulations completed directly through NAMD<sup>70</sup> version 2.12. The bio3d<sup>71</sup> R package version 2.2.4 was used for analysis of the simulations. Electrostatic potential of the HP1 $\alpha$  peptide and IMP $\alpha$  alone or the complex was calculated via the APBS server.<sup>72</sup> Molecular visualizations were generated in VMD<sup>73</sup> version 1.9.3. and PyMol<sup>54</sup> version 1.8.7.

## 3 | RESULTS

### 3.1 | Building a three-dimensional model for better understanding structural and dynamic properties of the HP1 $\alpha$ -IMP $\alpha$ complex

The translocation of HP1 $\alpha$  to the nucleus is a critical step for realizing its functional role in epigenetic inheritance of cancer; however, mechanisms that can regulate this critical process remained poorly studied. Notably, SLiM analysis identified a bipartite NLS within the HP1 $\alpha$  linker (IDR2) (Figure 1A).<sup>34</sup> As rules governing the binding mode of IMP $\alpha$  with NLS motif-containing proteins has been previously shown,<sup>55</sup> this data allowed us to develop a knowledge-based approach to acquiring a model for the complex of HP1 $\alpha$  and IMP $\alpha$ . We took advantage of threading the HP1 $\alpha$  linker sequence onto the structure of NIN2 complexed with IMP $\alpha$ . In this conformation, the N-terminal basic cluster (KRR) of the HP1 $\alpha$  NLS associated with the minor groove of IMP $\alpha$ , while the C-terminal cluster (KKKR) interacted with the major groove (Figure 1B). The electrostatic map of the complex illustrates the charge distribution of both molecules that allows HP1 $\alpha$  to sit in the two binding pockets formed by the grooves in IMP $\alpha$  (Figure 1C). The complex of the full length HP1 $\alpha$  with IMP $\alpha$  was also constructed through docking with ClusPro.<sup>56,57,59,60</sup> This second approach placed the NLS sequences at the same position as our threading model (Figure 1D). The interaction of HP1 $\alpha$  and IMP $\alpha$  was provided by immunoaffinity purification of HP1 $\alpha$ -interacting proteins from HeLa cells and their mass spectrometry-based identification. We found that HP1 $\alpha$  copurified with several IMP $\alpha$  subunits and with IMP $\beta$  (Figure 1E). Thus, experimental evidence supports an interaction between these proteins, and our

**FIGURE 1** HP1 $\alpha$  forms a complex with the nuclear receptor IMP $\alpha$  via the canonical binding mode. The complex of HP1 $\alpha$  and IMP $\alpha$  was generated via threading the sequence of IDR2 onto the structure of an established IMP $\alpha$  interactor, N1N2, in Discovery Studio v17. A, Linear motif analysis of HP1 $\alpha$  via the PsortII algorithm predicted a bipartite NLS within IDR2 that conforms to the K/RX10-12KRRK consensus sequence. This NLS is shared among the three HP1 proteins. B, 3D space-filling model of the HP1 $\alpha$  IDR2 peptide bound to IMP $\alpha$ . The N-terminal cluster of basic residues (KRK, shown in blue) associates with the minor groove of IMP $\alpha$ , while the C-terminal cluster (KKKR, shown in blue) binds to the major groove. C, Electrostatic potential of the complex, the HP1 $\alpha$  peptide, and the IMP $\alpha$  receptor illustrates the charge distribution that facilitates the interaction of the HP1 $\alpha$  IDR2 and IMP $\alpha$ . D, 3D space-filling model of full-length HP1 $\alpha$  docked to IMP $\alpha$  via ClusPro. The interaction is identical to that of the IDR2 peptide with the globular CD and CSD pointing outward and downward. E, HP1 $\alpha$  copurifies with several IMP $\alpha$  subunits and with IMP $\beta$ . IMP $\alpha$ , importin  $\alpha$ ; NLS, nuclear localization signal

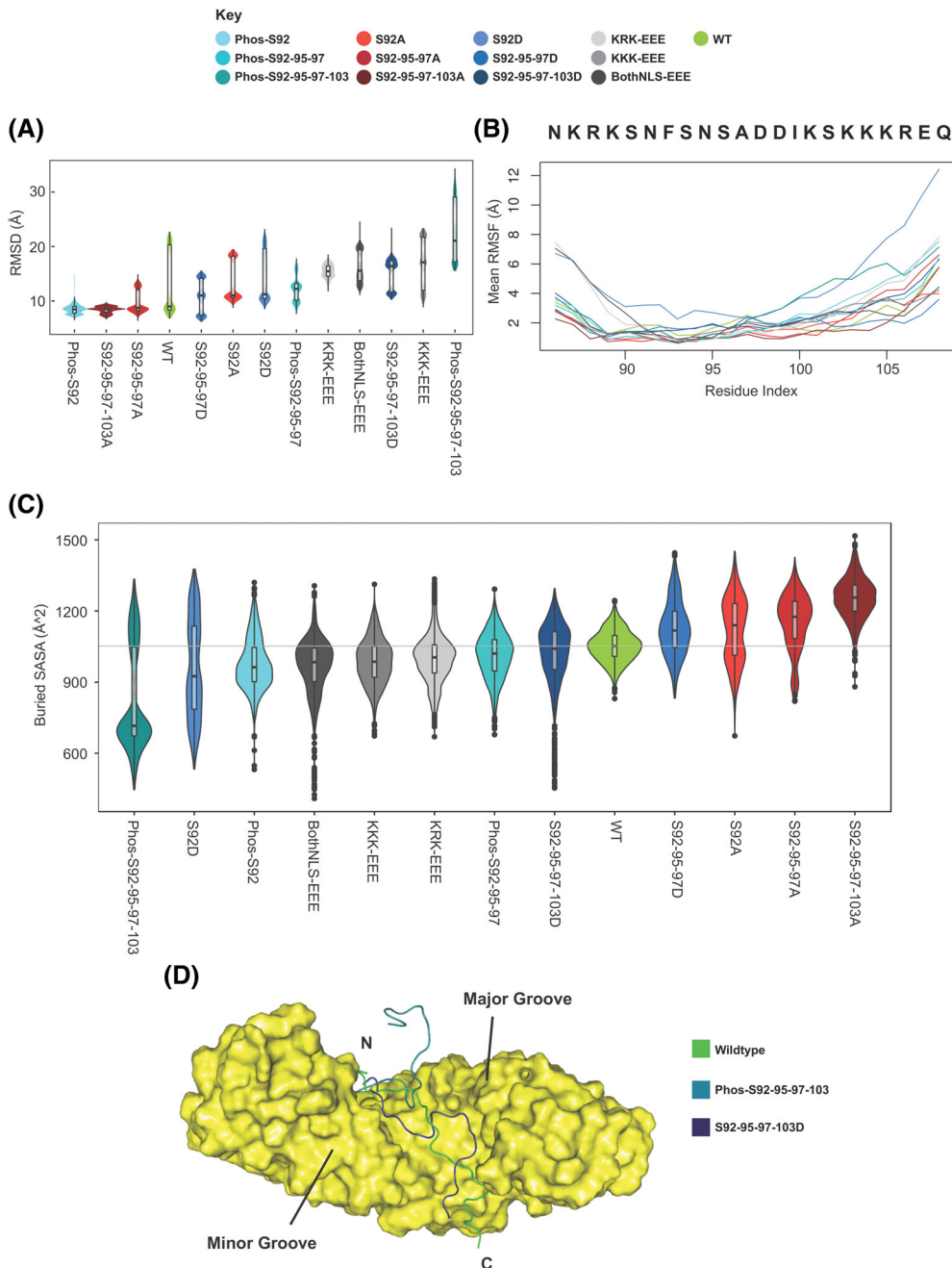


molecular modeling provides molecular details underlying this interaction. Last, because of the high levels of homology considered in building our models, the conclusion derived from studying their behavior has a considerable level of reliability.

### 3.2 | Modeling the impact of phosphorylation on the regulation of the HP1 $\alpha$ -IMP $\alpha$ complex

We and others have previously demonstrated that the linker domain of HP1 $\alpha$  is amenable to extensive post-translational modifications.<sup>74-77</sup> However, how these phosphorylation events affect critical functions of this protein remains to be fully understood. Thus, we considered information from SLiM analyses and data from experimentally derived mass spectrometry-based datasets to identify phosphorylation events that most likely regulate the function of this protein. For this purpose, we generated several models of the HP1 $\alpha$  linker including partial or fully phosphorylated linkers, as well as

nonphosphorylatable, or phosphomimetic mutants in a complex with IMP $\alpha$  and performed MD simulations. Subsequently, we evaluated MD simulations using multiple metrics. RMSD and RMSF of each HP1 $\alpha$  peptide quantified the overall deviation from the canonical interaction pose and which residues experienced the greatest deviations, respectively. RMSD values were increased above the range observed in the wild-type for mutations of the NLS motifs and by phosphorylation or phosphomimetic mutation of all four serine residues in the HP1 $\alpha$  linker peptide. Thus, these modifications likely induce changes in the linker peptide-protein interaction (Figure 2A). RMSF (Figure 2B) showed the greatest changes in the first seven residues—specifically for S92D, N-terminal NLS variants, and the joint NLS variants. From the middle of the peptide to the C-terminal end, RMSF values were higher for S92D and the phosphorylation of S92-95-97-103. We also assessed SASA of the HP1 $\alpha$  peptide in each condition. The total area of the peptide was significantly reduced only by mutation of the NLS motifs, suggesting that the peptide undergoes

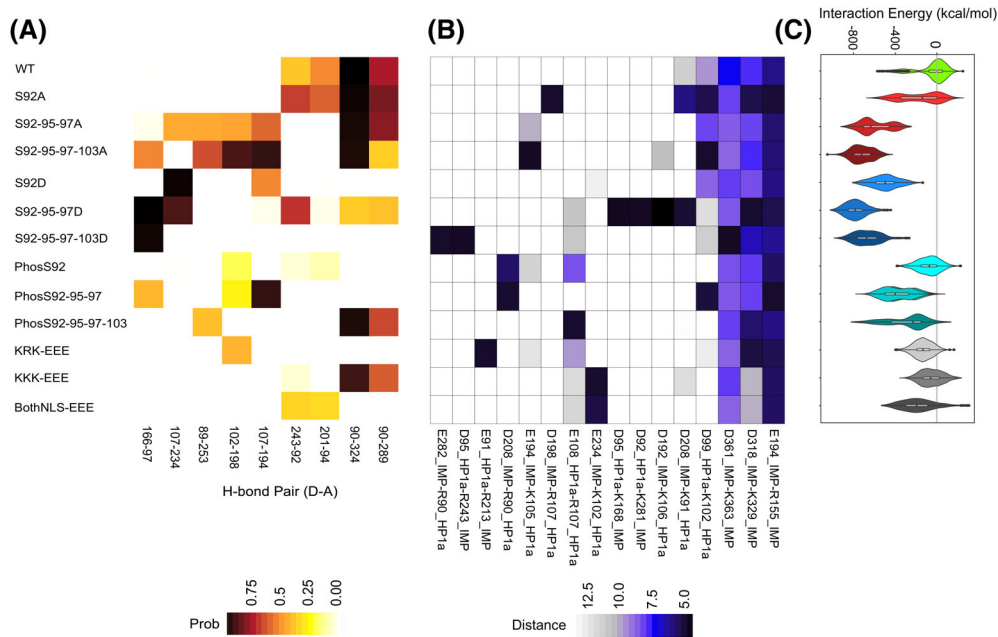


**FIGURE 2** Phosphorylation of the HP1 $\alpha$  linker inhibits binding to IMP $\alpha$  by affecting conformational changes in IDR2. MD simulation of the HP1 $\alpha$ -IMP $\alpha$  complex was employed to interrogate the impact of phosphorylation of the serine residues located in the region of the HP1 $\alpha$  linker that comprises the bipartite NLS. A, B, The flexibility and mobility of the peptides in all the experimental conditions were reflected in calculations of RMSD and RMSF. The clusters of basic residues in the NLS are indicated above the RMSD plot in blue. C, Buried SASA of the HP1 $\alpha$  peptide in all conditions. Decreased values indicate increased access to solvent. D, Visualization of the last frame of the wild-type, Phos-S92-95-97-103, and S92-95-97-103D phosphomimetic mutant trajectories was congruent with the results of RMSD and RMSF calculation and demonstrates the conformational changes sustained by the peptides. IMP $\alpha$ , importin  $\alpha$ ; MD, molecular dynamics; NLS, nuclear localization signal; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SASA, solvent accessibility surface area

a conformational change that resembles folding on itself (Figure S1). Generally, buried SASA of the HP1 $\alpha$  linker peptide was reduced with phosphorylation, phosphomimetic mutations, and NLS motif mutation. Few of the conditions were associated with significant changes in total buried SASA, but there was high variability indicating the potential for rearrangements in the interaction (Figure 2C). Visualization of the last frame of the trajectories (Figure 2D and Figure S3), which represents the final conformation of the complex, summarized these overall changes, as the loss of interaction between the NLS motif residues and IMP $\alpha$  basic patches is visually evident. Thus, together, these global geometric measures demonstrate that the phosphomimetic mutation and phosphorylation of the serine residues within the HP1 $\alpha$  linker alter peptide conformation.

### 3.3 | Energetic and molecular mechanic calculations further assess the conformation of The HP1 $\alpha$ -IMP $\alpha$ complex

We assessed changes in noncovalent bonding to infer the strength of the molecular association of the HP1 $\alpha$  linker peptide and IMP $\alpha$ . The hydrogen bond matrix in Figure 3A showed that the engineered NLS binding controls, in which the basic residues in each cluster of the bipartite NLS (KRK, KKK) were mutated to acidic residues (KRK-EEE, KKK-EEE), produced the expected loss of hydrogen bonding and facilitated interpretation of the results of the experimental conditions. Residues R90, S92, and F94 consistently formed hydrogen bonds with residues within IMP $\alpha$  (eg, W201, R243, N289, and E324). Nonphosphorylatable alanine mutation of one or all the serine residues in the linker (S92,



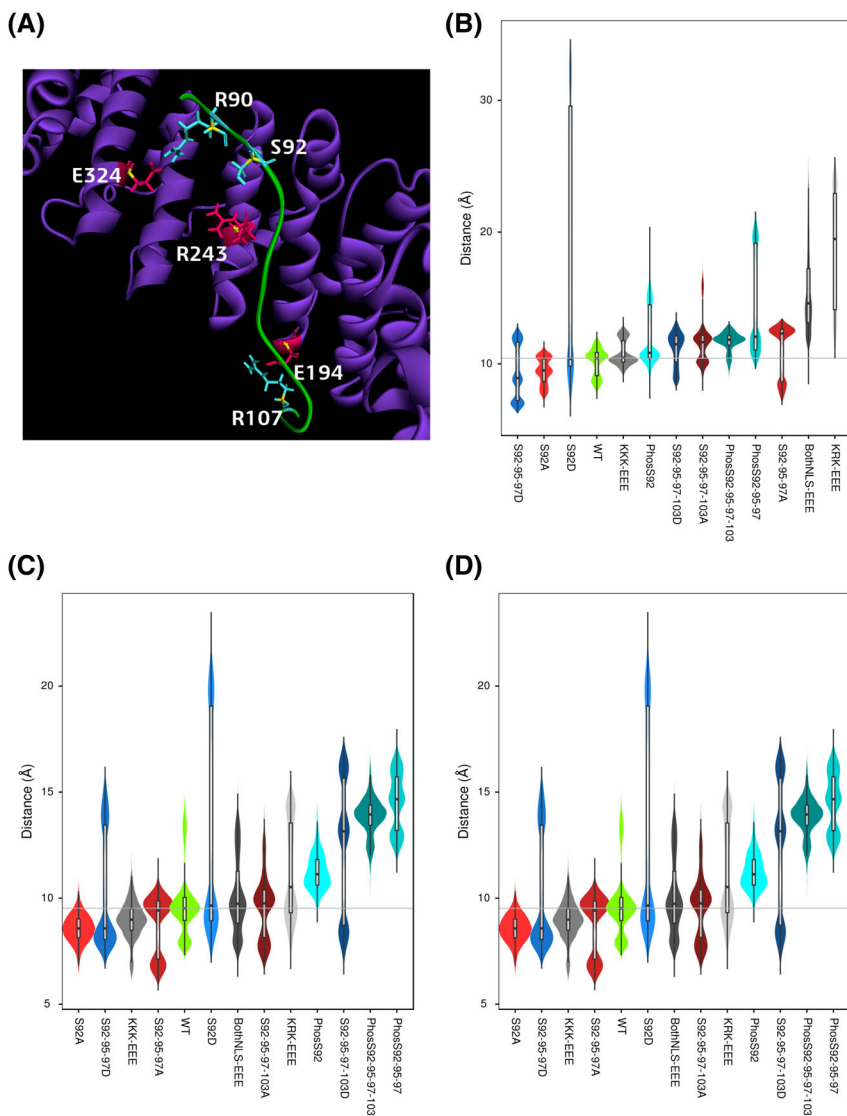
**FIGURE 3** Phosphorylation of the HP1 $\alpha$  linker prevents association with IMP $\alpha$  by inhibiting molecular contacts necessary for the interaction. Analysis of changes in molecular contacts throughout the MD simulations was accomplished through monitoring of noncovalent interactions and interaction energy between the two proteins. A, The hydrogen bond matrix visualizes the frequency with which each residue pair interacts, across conditions. B, The salt-bridge matrix visualizes the median distance between donor and acceptor residue pairs. C, Interaction energy of the HP1 $\alpha$ -IMP $\alpha$  complex in each condition. The gray line represents the median value of wild-type for comparison across conditions. IMP $\alpha$ , importin  $\alpha$ ; MD, molecular dynamics

95, 97, 103) permitted the retention of some of the same hydrogen bonds that form at the N-terminal part of the NLS, in addition to formation of new hydrogen bonds that were not present in the wild-type. These new interactions were between the C-terminal basic cluster of residues (K102, R107) and residues within IMP $\alpha$  that are between the major and minor groves (eg, E294, D198, and E235). Thus, non-phosphorylatable variants are associated with loss of specific interactions within the major and minor groves, perhaps leading to less specific interactions.

Subsequently, we modeled the effect of serine phosphorylation of the HP1 $\alpha$  linker domain. Phosphorylation of S92 alone, for instance, was sufficient to reduce the frequency of hydrogen bonds between HP1 $\alpha$  and IMP $\alpha$ . Additional phosphorylation of S95 and S97 resulted in the formation of new hydrogen bonds, while phosphorylation of all four serine residues, S92, S95, S97, and S103 demonstrated that the peptide maintains contact with IMP $\alpha$  only through the hydrogen bonds formed between two basic residues in the N-terminal portion of the NLS (K89, R90) and adjacent residues in IMP $\alpha$ . Monitoring salt-bridge formation revealed that many interprotein interactions are common to all conditions, but some are variant specific (Figure 3B). Variants within the C-terminal cluster of basic residues (KKK-EEE) disrupted an intrachain salt-bridge (between D318 and K320) to the minor groove of IMP $\alpha$  (Figure S2); specifically, K102 in HP1 $\alpha$  to D99 in IMP $\alpha$ . One intrachain salt-bridge within HP1 $\alpha$  was compromised by the phosphorylation of S92 only, phosphorylation of S92-95-97-103, mutation of one or both NLS clusters of basic residues (KKK-EEE), and phosphomimetic mutation of S92-95-97 or S92-95-97-103. This result indicates that the interaction between the IMP $\alpha$  minor groove

and NLS motif residues in HP1 $\alpha$  is crucial for the maintenance of the specific conformation of the peptide that is compatible with IMP $\alpha$  binding. While known biochemically, this is an important feature recapitulated in our simulations, increasing our confidence in the other results returned by our simulations. The NLS binding mutants (eg, KRK-EEE) maintained non-WT contact with IMP $\alpha$  by forming one or more salt-bridges between one of the mutated residues and a residue outside of the IMP $\alpha$  major or minor groves (eg, K91E in HP1 $\alpha$  to R213 in IMP $\alpha$ ).

The nonphosphorylatable variants (S92-95-97-103 to alanine) formed new salt bridges through K105 and K107 that contributed to the stabilization of the interaction with IMP $\alpha$ . Phosphorylation at S92 only or S92-95-97 resulted in the formation of a salt-bridge between R90 (N-terminal) in HP1 $\alpha$  and D208 (major groove) in IMP $\alpha$ , which is congruent with the occurrence of a conformational change that causes the N-terminal part of the HP1 $\alpha$  NLS (KKR) to lose contact with the minor groove and make new contacts with the major groove to preserve the interaction. Only one new salt-bridge was formed upon phosphorylation of S92-95-97-103, and this intrachain interaction occurred at the end of the peptide between E108 and R107. The phosphomimetic mutants again displayed different patterns of interactions from that of the phosphorylated protein. New salt-bridges were formed between the mutated residues (S92, S95) and nearby residues in IMP $\alpha$  suggesting a closer affinity of the two proteins. This observation is important as experimentalists typically consider phosphomimetic mutants to be similar to a phosphorylation event, but our data revealed differences.



**FIGURE 4** Phosphorylation of the HP1 $\alpha$  linker increases distance between the peptide and IMP $\alpha$ . A, Schematic of the C $\alpha$  atom pairs selected from the hydrogen bond matrix for receptor-ligand distance metrics. IMP $\alpha$  is shown in purple with the selected residues represented as stick structures colored pink. The HP1 $\alpha$  peptide is shown in green with the selected residues illustrated as stick structures colored cyan. All C $\alpha$  atoms are highlighted in yellow. Residue labels are shown in white. Residue pairs for distance measurements are written with the IMP $\alpha$  residue appearing first. B, Distance metrics for E324:R90. C, Distance metrics for R243:S92. D, Distance metrics for E194:R107. IMP $\alpha$ , importin  $\alpha$

Next, we computed the total interaction energy between each HP1 $\alpha$  peptide and IMP $\alpha$ . The interaction energy is influenced by all other interactions quantified above, and thus the combination of metrics is more interpretable than each alone. The wild-type and nonphosphorylatable variants showed lower interaction energies (on average,  $\sim$ –100 kcal/mol), compared to phosphorylated HP1 $\alpha$  peptides and phosphomimetic mutants (on average,  $\sim$ –600 kcal/mol) (Figure 3C). The similarities among, for example, phosphorylations and phosphomimetic mutants make our previous observations of rearranged interprotein interaction critical for interpretation. Phosphorylation of the HP1 $\alpha$  peptide at any site, but primarily S92, increased interaction energy similar to the NLS binding mutants, indicating that the complex is destabilized and supporting conformational changes observed in these conditions (Figure S2). Therefore, phosphorylation of the HP1 $\alpha$  linker changes the binding mechanics with IMP $\alpha$  by modifying the noncovalent hydrogen bonds, salt-bridge connections, and energy of interaction to inhibit the full interaction of the peptide with IMP $\alpha$ . Genomic variants lead to rearrangements of intermolecular contacts; many interactions seen in the wild type are lost for each variant and new contacts not observed in the wild type are

gained (Figure S3). Because genomic variants lead to loss of interaction at the IMP $\alpha$  minor groove and new interaction between the IMP $\alpha$  grooves, our data indicate that genomic variants likely alter the interaction to have less specificity.

We also monitored the distance between residues participating in hydrogen bonds as a simplified geometric measure to facilitate interpretation (Figure 4A). The first residue pair analyzed was R90 (HP1 $\alpha$ ) and E324 (IMP $\alpha$ ) which forms a bond in the wild-type condition, between the N-terminal cluster of basic residues in the HP1 $\alpha$  NLS (KRR) and the minor groove of IMP $\alpha$ . The violin plot in Figure 4B shows the median value of the distance between these two residues in the wild-type peptide, represented by a gray line. The S92D phosphomimetic mutant displayed much variability, which is attributable to the peptide shifting down and slightly out of the binding pocket created by the minor groove of IMP $\alpha$  (Figure S2). All phosphomimetic and nonphosphorylatable mutations impacted the conformation of the nearest NLS motif residues, often leading to the peptide leaving the major or minor groove binding site and visually evident in Figure S2 but were also associated with similar alteration at the more distant motif. The greatest separation between IMP $\alpha$  and HP1 $\alpha$  was observed

in the mutation of the N-terminal cluster of basic residues to acidic ones (KRK-EEE), because the peptide loses contact with IMP $\alpha$  on that end. Phosphorylation of one residue in the linker, phosphorylation of all serine residues, and for phosphomimetic substitution considerably increased the distance between S92 and R243, indicating that the post-translational modification of the serine residues alters interactions with IMP $\alpha$ . The third residue pair, R107 (HP1 $\alpha$ ) and E194 (IMP $\alpha$ ), which connect the C-terminal part of the HP1 $\alpha$  NLS (KKK) to the major groove of IMP $\alpha$ , were not found to be constantly connected by a hydrogen bond in the wild-type simulations but were consistently present in the nonphosphorylatable mutants (Figure 3A). As an approximate quantification of how much HP1 $\alpha$  had left the IMP $\alpha$  groove, we used distance monitors between specific pairs of residues that interact in the wild-type (Figure 4C,D and S4). Distance monitors demonstrate that the peptide is dynamic within the IMP $\alpha$  binding site, but that consistent changes in the conformation are seen between conditions. Thus, our data indicate that phosphorylation or phosphomimetic genomic variants prevent the HP1 $\alpha$  linker to bind IMP $\alpha$ .

### 3.4 | Modeling the impact of phosphorylation-specific mutations on the regulation of the HP1 $\alpha$ -IMP $\alpha$ complex

We considered the specific patterns of residue-level rearrangements across conditions. First, we visualized the final conformation between the two members of the complex, as a visual guide to how the interactions changed during simulation (Figure S2). We quantified rearrangements by plotting the minimum time-averaged distance between each residue of the two proteins (Figure 5). Compared to WT, nonphosphorylatable linker variants retained some previously established contacts, but also introduced new ones, primarily with the N-terminus. Phosphomimetic and nonphosphorylatable mutants lost more native contacts with the peptide C-terminus and linker and gained even more interactions across the peptide N-terminus. Interestingly, both phosphorylation and the nonphosphorylatable mutation of S92 (S92A) closely resembled WT in its bonding pattern, while its phosphomimetic mutation (S92D) induced loss of interaction between the N-terminal portion of the HP1 $\alpha$  peptide (residues 87-95) and the minor groove (residues 315-403) of IMP $\alpha$  (Figure 5). This finding suggests that the single phosphomimetic mutant affects changes in noncovalent interactions that secure the association of the N-terminal end of HP1 $\alpha$  linker to IMP $\alpha$ , allowing it to shift out of the binding pocket. Phosphorylation of the same residues resulted in complete loss of those C-terminal contacts that is reflective in the last frame of the simulation (Figure S2). These data indicate a progression, where partially modified HP1 $\alpha$  linkers retain a subset of their specific native contacts, while full phosphorylation further perturbs native contacts, resulting in the spatial separation of the molecules.

Subsequently, we analyzed conformational dynamics of the bound linker peptide using PCA. Comparing the mean values for each trajectory in PC1-PC2 space shows variation among technical replicates (Figure S5), but with consistent trends among conditions. Phosphorylation of S92-95-97-103 resulted in the most deviation from the wild-type, while

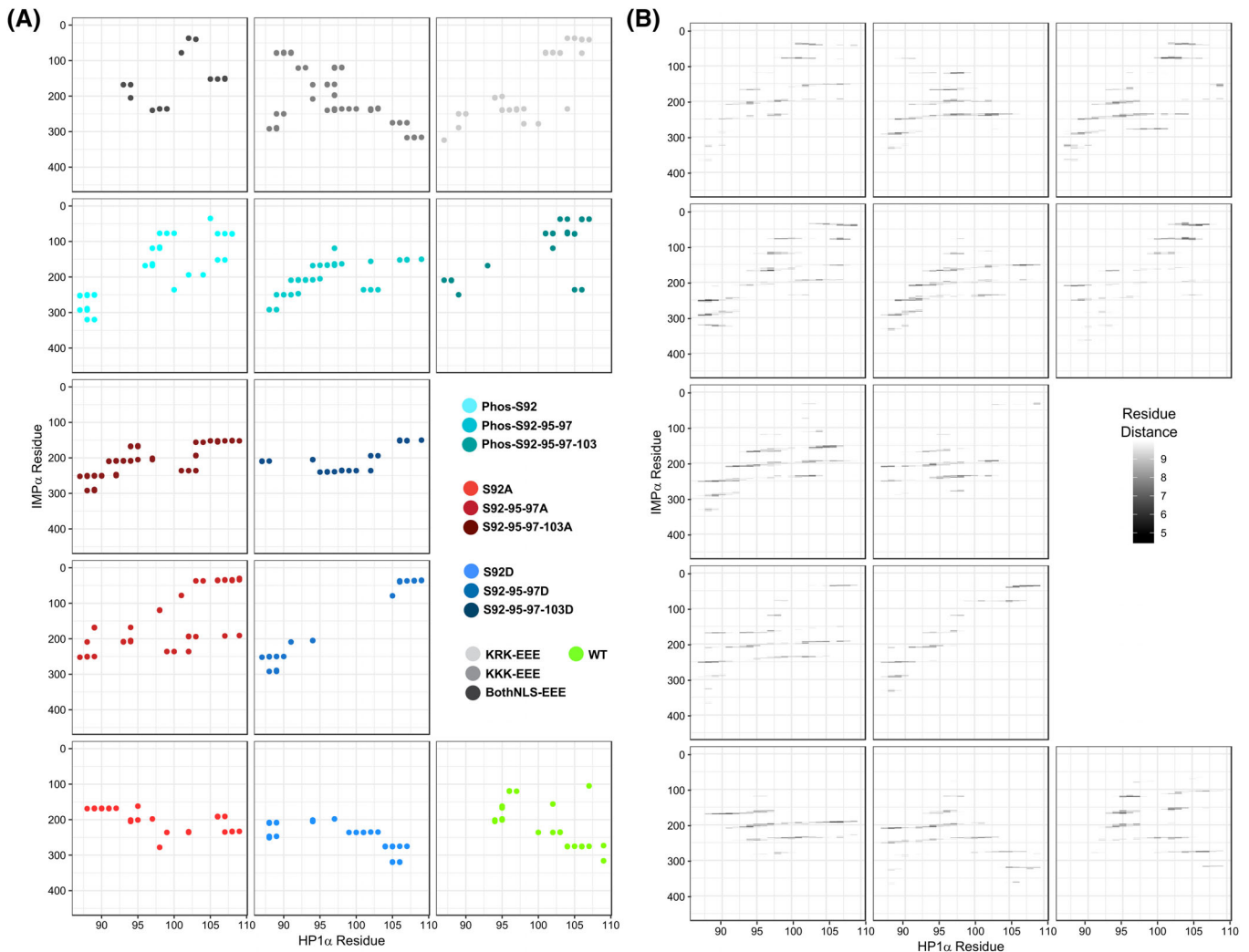
conditions with fewer phosphorylated residues and nonphosphorylatable mutants clustered together and with the wild-type. Similar to the fully phosphorylated linker, the NLS binding mutants are separated from wild-type in PC space, indicating their consistent global differences from the wild-type. PC3 demonstrated relatively little consistent difference among phosphorylation or phosphomimetic mutation and wild-type (Figure S5). However, PC2 and PC3 together showed phosphorylation of S92-95-97-103 diverged from all the conditions except the NLS binding mutants (Figure S5). We also assessed changes in the charge distribution of the HP1 $\alpha$  linker peptide upon phosphorylation, nonphosphorylatable or phosphomimetic mutation. The complex of the wild-type HP1 $\alpha$  peptide and IMP $\alpha$  demonstrated the normal conformation of the two proteins where the major and minor groove surfaces of IMP $\alpha$  are negatively charged to complement the positively charged peptide (Figure 6). Regions of IMP $\alpha$  that flank the major and minor grooves are positively charged, likely helping to orient the peptide into the two grooves. Therefore, we conclude that full phosphorylation of the HP1 $\alpha$  linker is most likely to be incompatible with its transport through the nuclear membrane. This finding is congruent with the fact that the highest level of linker phosphorylation occurs during mitosis at a moment in which the nuclear membrane is not present, a phenomenon that would leave most of this form of HP1 $\alpha$  in the cytoplasm.

### 3.5 | Modeling the impact of cancer-associated mutations on the regulation of the HP1 $\alpha$ -IMP $\alpha$ complex

To gain more insights into how the regulation of HP1 $\alpha$ , through its linker domain, may be altered by cancer-associated mutations, we evaluated six additional genomic variants, applying the same approaches used above to assess phosphorylation. Peptide RMSD measures global changes from the initial contacts; S92Y and Q109H were associated with smaller deviations, while F94L and D99A were associated with larger departures (Figure S6A). RMSF values indicated that the C-terminal side is more mobile than the N-terminal side for WT and the majority of variants. This mobility increased for F94L and D99A compared to all other conditions (Figure S6B). Buried SASA was increased for S92Y and S95L but decreased for F94L (Figure S6C). The total SASA of the peptide was increased for S92Y and decreased for F94L (Figure S6D), indicating rearrangements to intrapeptide interactions. Visualization of the final conformation for each simulation demonstrated differences in positioning of the center of the peptide, which was no longer between the IMP $\alpha$  grooves for many variants. Lastly, the C-terminus often departed from the major groove acidic cluster (Figure S7).

We next analyzed interacting residues at the binding interface between both proteins. Certain hydrogen bonds between the peptide and IMP $\alpha$  were stable across all variants. For example, the hydrogen bond between R90 in HP1 $\alpha$  and E324 in IMP $\alpha$  was stable in all but D99A. However, all the variants formed new bonds with IMP $\alpha$  (Figure S4A). Interestingly, the D99A and I101M variants, which occurred relatively close to the C-terminal end of the HP1 $\alpha$  linker peptide, formed additional hydrogen bonds through residues K91 and





**FIGURE 5** Intermolecular contacts between HP1 $\alpha$  and IMP $\alpha$  in all conditions—HP1 $\alpha$ -IMP $\alpha$  intermolecular contacts are disrupted by full phosphorylation of the linker. A, Residue-residue contacts between HP1 $\alpha$  and IMP $\alpha$  were calculated and plotted in a contact map matrix where points indicate residues within 6 Å for at least 10% of the trajectory. Each condition is indicated by color, as in previous figures. B, We also plot the minimum average distance between residues on a grayscale where residues further than 10 Å apart are white and those closer than 5 Å are black. IMP $\alpha$ , importin  $\alpha$

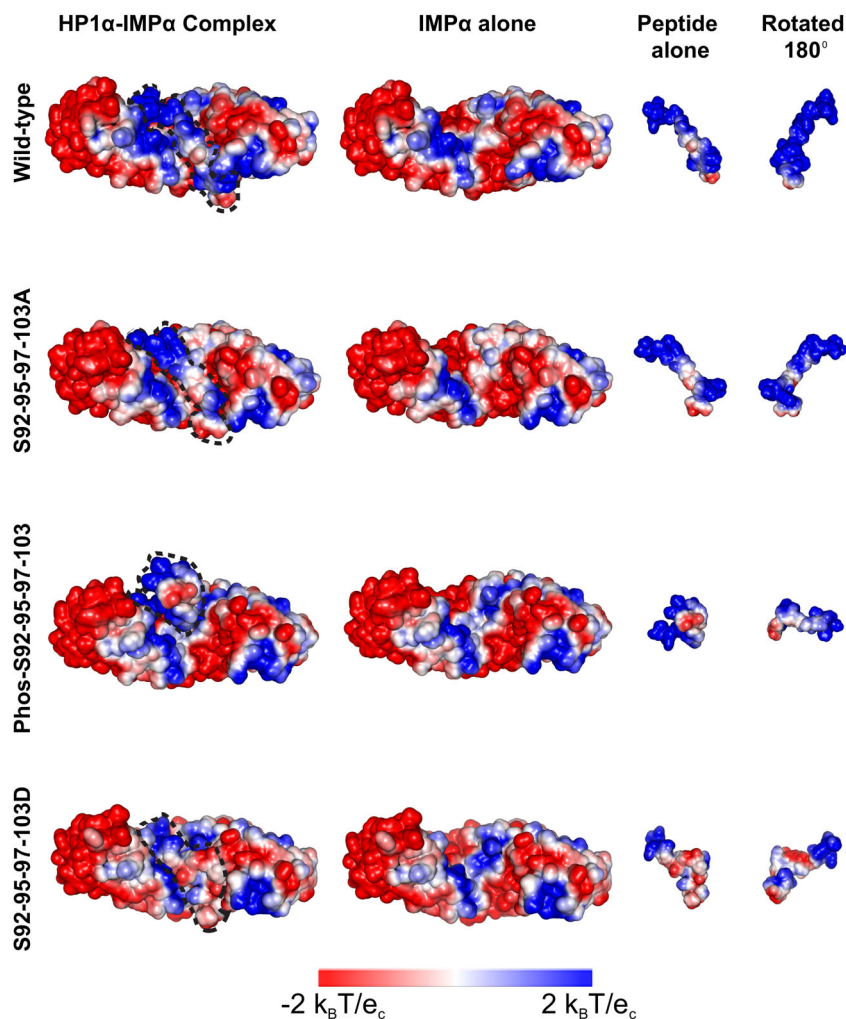
S92 at the N-terminal end of the peptide. Similar to wild-type, I101M did not establish any hydrogen bond contacts with the C-terminal end of the peptide. Analysis of salt-bridge interactions showed that one HP1 $\alpha$  intrachain salt-bridge (D99 and K102) was abolished by the D99A and S95L variants; with D99A also forming a new interchain salt-bridge with R166 in IMP $\alpha$  (Figure S4B). The F94L variant formed two new salt-bridges (D100 in HP1 $\alpha$  and R155 in IMP $\alpha$ ; K102 in HP1 $\alpha$  and D198 in IMP $\alpha$ ) that could further stabilize binding of the C-terminal end of the peptide to the major groove of IMP $\alpha$ . The total interaction energy was lower for F94L and S95L and higher for the S92Y variant (Figure S4C), indicating tighter and weaker interaction, respectively. We assessed residue-residue proximity for all pairs, which revealed an increase in the number of intermolecular contacts formed between the HP1 $\alpha$  variants and IMP $\alpha$ , which clearly indicated the extent of rearranged interactions for the majority of variants (Figure S5). Normally, the HP1 $\alpha$  linker peptide makes specific contacts with the IMP $\alpha$  major and minor grooves. Thus, altogether, the

interaction metrics between HP1 $\alpha$  and IMP $\alpha$  suggest that certain genomic variants rearrange intermolecular contacts in a manner to render HP1 $\alpha$  less specific for the IMP $\alpha$  major and minor grooves, likely leading to impaired function.

## 4 | DISCUSSION

Our laboratory has studied HP1 proteins for several years in relationship to their function as epigenomic effectors of growth signaling pathways. Interestingly, our own work and that of others has been, throughout the years, biased to studying the structural-functional relationship of the globular domain of these proteins. Thus, the current study, which is focused on using molecular modeling and dynamic stimulations to interpret HP1 $\alpha$  linker phosphorylation and cancer-associated genomic variants bears novelty and biomedical relevance. In particular, we specifically investigate how these regulatory and

**FIGURE 6** Electrostatic potential maps of the HP1 $\alpha$ -IMP $\alpha$  complexes. The charge distribution of the IMP $\alpha$  receptor and HP1 $\alpha$  peptides under conditions of phosphorylation, nonphosphorylatable, and phosphomimetic mutation was evaluated and visually represented. The complex (HP1 $\alpha$  peptide outlined in dashed lines), IMP $\alpha$  alone, and the peptide alone are shown with the peptide optimally rotated (180 $^\circ$ ) to expose the binding surface. IMP $\alpha$ , importin  $\alpha$



pathological events alter interactions with the nuclear receptor, IMP $\alpha$ . Linear motif analyses identify, with a high level of confidence, an NLS consensus site within this intrinsically disordered region of the protein. This data is congruent with our immunopurification experiments that reveal the existence of a HP1 $\alpha$ -IMP $\alpha$  complex. Linear Motif analyses for post-translational modifications combined with data from large scale mass spectrometry experiments demonstrate that the linker is amenable to phosphorylation at different sites. Thus, the HP1 $\alpha$  linker appears to function as a binding interface for IMP $\alpha$  and in the integration for upstream regulatory cell signaling pathways. More importantly, this complex is predicted to be responsible of transporting HP1 $\alpha$  into the nucleus, where it becomes functional. We find that several cancer-associated mutations affect the sequence of this domain in a manner that we predicted may disrupt these functions. Indeed, molecular modeling and dynamic simulations, combined with an analytical approach that uses different metrics, provide evidence to support this prediction. Briefly, we find that the unmodified linker peptide binds to IMP $\alpha$  via hydrogen bonds and salt-bridges formed between the bipartite NLS of HP1 $\alpha$  and acidic patches within the major and minor grooves of IMP $\alpha$ . The selected variants studied here were rare missense mutations reported in various cancers but not in the general population. Our modeling experiments show that

the interaction between these proteins is inhibited by phosphorylation of the four serine residues within the linker, which leads to a more extended HP1 $\alpha$  conformation that is not compatible with the canonical binding mode for IMP $\alpha$ . Missense mutations in the HP1 $\alpha$  linker rearranged interactions with IMP $\alpha$ , likely leading to loss of binding or nonspecific binding. The changes in linker conformation were mediated by modification of intramolecular and intermolecular contacts that reflect the dynamic property of the linker peptide. Interestingly, phosphorylation of S92 might be sufficient to disrupt the association of the HP1 $\alpha$  linker with IMP $\alpha$ ; however, the phosphomimetic mutant of the same residue was more potent at shifting the peptide out of the IMP $\alpha$  binding pocket. Furthermore, the phosphorylation of S92 only or even phosphorylation of S92-95-S97 did not achieve the full effect in the form of reduced binding affinity for IMP $\alpha$ , as did phosphorylation of all four serine residues (S92, S95, S97, S103), suggesting that all four residues may affect the mechanism of IMP $\alpha$  interaction. These data support our own and previous findings that each serine residue within the NLS is phosphorylated during mitosis.<sup>74-76</sup> The S92Y and S95L variants could potentially disrupt the aforementioned phosphorylation-dephosphorylation cycle and present problems for accurately completing mitosis and nuclear translocation of HP1 $\alpha$ . The loss of interaction at the N-terminus potentially

modifies the flexibility of the peptide, allowing the new interactions to occur. Our results demonstrate the dynamics of hydrogen bond chemistry that govern the interaction between IMP $\alpha$  and the unmodified or phosphorylated HP1 $\alpha$  linker. Hence, this new knowledge extends our understanding of the functional impact that both, signaling and cancer-associated modifications may have on HP1 $\alpha$ -IMP $\alpha$  interactions. Thus, this information, generated at an atomic resolution level, should help to draw important inferences that explain critical functions of this key epigenetic regulator. In fact, one may consider that HP1 $\alpha$  binding to IMP $\alpha$  via its linker NLS is of paramount functional importance. The formation of an HP1 $\alpha$ -IMP $\alpha$  complex is a necessary step for subsequent translocation of the former through the nuclear membrane into the nucleoplasm; and ultimately to chromatin, where it exercises gene regulatory functions. The current study predicts that hyperphosphorylation of the linker domain of this protein likely prevents its association with IMP $\alpha$ . This information is congruent with studies that show that this phenomenon takes place during mitosis, the time in which the nuclear membrane has dissolved. However, insights into how these post-translational modifications achieve these effects, at an atomic resolution level, had remained unknown until this study. An additional contribution of this study is the modeling of cancer-associated HP1 $\alpha$  mutants within the linker domain of this protein, which also predict to disrupt interactions in tumor samples. Notably, this mechanism of dysfunction has never been explored before for HP1 $\alpha$ . We would also like to briefly consider that the current study can serve as a model for understanding the interaction of IMP $\alpha$  to other proteins, for which the NLS is similar in content and distribution of acceptor motifs for phosphorylation. Similarly, changes in the surface of the HP1 $\alpha$  linker under phosphorylated or dephosphorylated conditions may allow predictions for its interaction with other proteins or DNA. Consequently, these are few examples of how our results not only offer explanations of mechanisms underlying functional aspects of these proteins, but also have predictive value, such as for understanding how somatic mutations in the linker region may lead to functional inactivation in nonmitotic cells.

## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grants R01 CA178627 (to G.L.) and R01 DK52913 (to R.U.), the Advancing a Healthier Wisconsin Endowment (to each G. Lomberk and R. Urrutia) and an endowment from the Linda T. and John A. Mellowes Foundation (to R. Urrutia).

## CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

## ORCID

Raul Urrutia  <https://orcid.org/0000-0002-1640-6780>

## REFERENCES

- Nestorov P, Tardat M, Peters AHFM. Chapter eight - H3K9/HP1 and Polycomb: two key epigenetic silencing pathways for gene regulation and embryo development. In: Heard E, ed. *Current Topics in Developmental Biology*. San Diego, CA: Academic Press; 2013:243-291.
- Panteleeva I, Boutillier S, See V, et al. HP1 $\alpha$  guides neuronal fate by timing E2F-targeted genes silencing during terminal differentiation. *EMBO J*. 2007;26(15):3616-3628.
- Takanashi M, Oikawa K, Fujita K, Kudo M, Kinoshita M, Kuroda M. Heterochromatin protein 1 $\gamma$  epigenetically regulates cell differentiation and exhibits potential as a therapeutic target for various types of cancers. *Am J Pathol*. 2009;174(1):309-316.
- Grzenda A, Leonard P, Seo S, et al. Functional impact of Aurora A-mediated phosphorylation of HP1 $\gamma$  at serine 83 during cell cycle progression. *Epigenetics Chromatin*. 2013;6(1):21.
- De Koning L, Savignoni A, Boumendil C, et al. Heterochromatin protein 1 $\alpha$ : a hallmark of cell proliferation relevant to clinical oncology. *EMBO Mol Med*. 2009;1(3):178-191.
- Minc E, Allory Y, Worman HJ, Courvalin JC, Buendia B. Localization and phosphorylation of HP1 proteins during the cell cycle in mammalian cells. *Chromosoma*. 1999;108(4):220-234.
- Hayakawa T, Haraguchi T, Masumoto H, Hiraoka Y. Cell cycle behavior of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase. *J Cell Sci*. 2003;116(16):3327-3338.
- De Lucia F, Ni JQ, Vaillant C, Sun FL. HP1 modulates the transcription of cell-cycle regulators in *Drosophila melanogaster*. *Nucleic Acids Res*. 2005;33(9):2852-2858.
- Auth T, Kunkel E, Grummt F. Interaction between HP1 $\alpha$  and replication proteins in mammalian cells. *Exp Cell Res*. 2006;312(17):3349-3359.
- Lomberk G, Mathison AJ, Grzenda A, et al. Sequence-specific recruitment of heterochromatin protein 1 via interaction with Krüppel-like factor 11, a human transcription factor involved in tumor suppression and metabolic diseases. *J Biol Chem*. 2012;287(16):13026-13039.
- Ayoub N, Jeyasekharan AD, Bernal JA, Venkitaraman AR. HP1-beta mobilization promotes chromatin changes that initiate the DNA damage response. *Nature*. 2008;453(7195):682-686.
- So Hee K, Jerry LW. The heterochromatin protein 1 (HP1) family: Put away a bias toward HP1. *Mol Cell*. 2008;26(3):217-227.
- Ball AR, Yokomori K. Revisiting the role of heterochromatin protein 1 in DNA repair. *J Cell Biol*. 2009;185(4):573-575.
- Dinant C, Luijsterburg MS. The emerging role of HP1 in the DNA damage response. *Mol Cell Biol*. 2009;29(24):6335-6340.
- Luijsterburg MS, Dinant C, Lans H, et al. Heterochromatin protein 1 is recruited to various types of DNA damage. *J Cell Biol*. 2009;185(4):577-586.
- Baldehyron C, Soria G, Roche D, Cook AJL, Almouzni G. HP1 $\alpha$  recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair. *J Cell Biol*. 2011;193(1):81-95.
- Cann KL, Dellaire G. Heterochromatin and the DNA damage response: the need to relax. *Biochem Cell Biol*. 2011;89(1):45-60.
- Lee YH, Liu X, Qiu F, O'Connor TR, Yen Y, Ann DK. HP1beta is a biomarker for breast cancer prognosis and PARP inhibitor therapy. *PLoS One*. 2015;10(3):e0121207.
- The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158-D169.
- Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol*. 2011;7:539.
- Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 2000;16(6):276-277.

22. Nakada R, Hirano H, Matsuura Y. Structure of importin- $\alpha$  bound to a non-classical nuclear localization signal of the influenza A virus nucleoprotein. *Sci Rep*. 2015;5:15055.
23. Miyamoto Y, Yamada K, Yoneda Y. Importin  $\alpha$ : a key molecule in nuclear transport and non-transport functions. *J Biochem*. 2016;160(2):69-75.
24. Roman N, Christie M, Swarbrick CMD, Kobe B, Forwood JK. Structural characterisation of the nuclear import receptor importin alpha in complex with the bipartite NLS of Prp20. *PLOS One*. 2013;8(12):e82038.
25. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA. Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol*. 2004;14(9):505-514.
26. Harel A, Forbes DJ. Importin beta: conducting a much larger cellular symphony. *Mol Cell*. 2004;16(3):319-330.
27. Gorlich D, Henklein P, Laskey RA, Hartmann E. A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus. *EMBO J*. 1996;15(8):1810-1817.
28. Conti E, Uy M, Leighton L, Blobel G, Kuriyan J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell*. 1998;94(2):193-204.
29. Kobe B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat Struct Biol*. 1999;6(4):388-397.
30. Fontes MR, Teh T, Kobe B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol*. 2000;297(5):1183-1194.
31. Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci*. 1999;24(1):34-36.
32. Kosugi S, Hasebe M, Entani T, Takayama S, Tomita M, Yanagawa H. Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling. *Chem Biol*. 2008;15(9):940-949.
33. Kosugi S, Hasebe M, Matsumura N, et al. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem*. 2009;284(1):478-485.
34. Velez G, Lin M, Christensen T, Faubion WA, Lomberk G, Urrutia R. Evidence supporting a critical contribution of intrinsically disordered regions to the biochemical behavior of full-length human HP1gamma. *J Mol Model*. 2016;22(1):12.
35. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;294(5):1351-1362.
36. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4(6):1633-1649.
37. Wong YH, Lee TY, Liang HK, et al. KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res*. 2007;35Web Server Issue:W588-W594.
38. Iakoucheva LM, Radivojac P, Brown CJ, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*. 2004;32(3):1037-1049.
39. Xue Y, Ren J, Gao X, Jin C, Wen, L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics*. 2008;7(9):1598-1608.
40. Dou Y, Yao B, Zhang C. PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino Acids*. 2014;46(6):1459-1469.
41. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res*. 2015;43(Database Issue):D512-D520.
42. Gnad F, Ren S, Cox J, et al. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol*. 2007;8(11):R250.
43. Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res*. 2011;39(suppl 1):D253-D260.
44. Mitchell A, Chang HY, Daugherty L, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 2015;43(Database Issue):D213-D221.
45. Cong Q, Grishin NV. MESSA: MEta-Server for protein sequence analysis. *BMC Biol*. 2012;10:82.
46. Willard L, Ranjan A, Zhang H, et al. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*. 2003;31(13):3316-3319.
47. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007;35Web Server Issue:W375-W383.
48. Chen VB, Arendall WB 3rd, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 1):12-21.
49. Williams CJ, Headd JJ, Moriarty NW, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27(1):293-315.
50. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure*. 2003;11(11):1453-1459.
51. Meszaros B, Erdos G, Dosztanyi Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res*. 2018;46(W1):W329-w337.
52. Wang L, Sauer UH. OnD-CRF: predicting order and disorder in proteins using [corrected] conditional random fields. *Bioinformatics*. 2008;24(11):1401-1402.
53. Biovia DS. Discovery Studio. San Diego: Dassault Systèmes ; 2017.
54. Schrödinger L. The PyMOL Molecular Graphics System; 2017.
55. Fontes MR, Teh T, Jans D, Brinkworth RI, Kobe B. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. *J Biol Chem*. 2003;278(30):27981-27987.
56. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res*. 2004;32Web Server Issue:W96-W99.
57. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 2004;20(1):45-50.
58. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins*. 2006;65(2):392-406.
59. Kozakov D, Beglov D, Bohnuud T, et al. How good is automated protein docking? *Proteins*. 2013;81(12):2159-2166.
60. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc*. 2017;12(2):255-278.
61. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*. 2018;173(2):291-304.e6.
62. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777-D783.
63. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
64. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11:361-362.
65. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.

66. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*; 2017.
67. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLOS One*. 2012;7(10):e46688.
68. Zimmermann MT, Urrutia R, Cousin MA, Oliver GR, Klee EW. Assessing human genetic variations in glucose transporter SLC2A10 and their role in altering structural and functional properties. *Front Genet*. 2018;9(9):276.
69. Cornell WD, Cieplak P, Bayly CI, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 1995;117(19):5179-5197.
70. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-1802.
71. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. 2006;22(21):2695-2696.
72. Jurrus E, Engel D, Star K, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci*. 2017;27(1):112-128.
73. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38.
74. Chakraborty A, Prasanth KV, Prasanth SG. Dynamic phosphorylation of HP1 $\alpha$  regulates mitotic progression in human cells. *Nat Commun*. 2014;5:3445.
75. Chakraborty A, Prasanth SG. Phosphorylation–dephosphorylation cycle of HP1 $\alpha$  governs accurate mitotic progression. *Cell Cycle*. 2014;13(11):1663-1670.
76. LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, Garcia BA. Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. *Mol Cell Proteomics*. 2009;8(11):2432-2442.
77. Lomber G, Bensi D, Fernandez-Zapico ME, Urrutia R. Evidence for the existence of an HP1-mediated subcode within the histone code. *Nat Cell Biol*. 2006;8:407-415.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zimmermann MT, Williams MM, Klee EW, Lomber GA, Urrutia R. Modeling post-translational modifications and cancer-associated mutations that impact the heterochromatin protein 1 $\alpha$ -importin  $\alpha$  heterodimers. *Proteins*. 2019;87:904–916. <https://doi.org/10.1002/prot.25752>