Deciphering the Interplay among Multisite Phosphorylation, Interaction Dynamics, and Conformational Transitions in a Tripartite Protein System

Philip Lössl,[†,‡] Andrea M. Brunner,[†,‡] Fan Liu,[†,‡] Aneika C. Leney,[†,‡] Masami Yamashita,[§] Richard A. Scheltema[†,‡] and Albert J. R. Heck[†,‡]

[†]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584CH Utrecht, The Netherlands

[‡]Netherlands Proteomics Center, Padualaan 8, 3584CH Utrecht, The Netherlands

[§]Department of Structural Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

-- Supporting Information –

SI Table of Contents

Experimental Procedures

Materials

All reagents were purchased from Sigma-Aldrich, if not otherwise indicated.

Recombinant protein production

All proteins were expressed in *E.coli* BL21 Gold pLysS cells (Stratagene) with N-terminal His-tags cleavable by SUMO protease (for Plk1 and Aur-A) or by TEV protease (for $Bora^{NT}$).

The proteins were purified by cobalt-affinity chromatography using TALON resin (Clontech) as an initial step. Aur-A was additionally purified by size exclusion chromatography on a Superdex75 column (GE Healthcare). The N-terminal His-tags were cleaved using His-tagged SUMO protease (Plk1 and Aur-A) and TEV protease ($Bora^{NT}$) and cobalt-affinity chromatography was used to remove the proteases and remaining uncleaved species. $Bora^{NT}$ and Plk1 were further purified by ion exchange chromatography using a Mono Q column (GE Healthcare) and a Resource S column (GE Healthcare), respectively. Finally, the proteins were subjected to size exclusion chromatography using a Superdex 200 column (GE Healthcare) for Plk1 and a Superdex 75 column for Aur-A and $Bora^{NT}$. Plk1 was stored in 20 mM MOPS buffer pH 7.4, 100 mM NaCl, 1 mM dithiothreitol (DTT), 5 mM tris(2-carboxyethyl)phosphine (TCEP). Aur-A was stored in 20 mM HEPES pH 7.5, 200 mM NaCl, 1 mM DTT, 5 mM TCEP, 10% (*v/v*) glycerol. $Bora^{NT}$ was stored in 20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 5 mM TCEP.

Kinase Reactions

If not otherwise indicated, equimolar amounts of Aur-A, $Bora^{NT}$ and Plk1 were incubated at room temperature in 20 mM HEPES pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP. Reactions were quenched on ice by addition of 20 mM EDTA pH 8.0. If required, pprephosphorylation of $Bora^{NT}$ was achieved by overnight incubation with catalytic amounts of Aur-A (kinase:substrate ratio 1:100), after which Mg-ATP was removed. All experiments were performed in duplicate.

Native MS data acquisition

Prior to analysis, the sample buffer was exchanged to 500 mM $CH_3CO_2NH_4$ pH 6.8 in 7 steps of concentration and dilution at 13,000 g and 4°C using centrifugal filter units with a 10 kDa molecular weight cutoff PES membrane (Sartorius). The protein solutions (estimated individual protein concentration = 5 µM) were loaded in gold-coated borosilicate glass needles and injected into an Orbitrap Exactive Plus EMR mass spectrometer (Thermo Fisher Scientific) using nano-electrospray ionization. Native mass spectra were acquired in positive ion mode with the following instrument settings: capillary voltage = 1.3–1.4 kV, source fragmentation voltage = 40 eV, ion injection time = 50 ms, higher-energy collision dissociation (HCD) energy = 10 eV, Noise threshold = 4.64, N_2 gas pressure in HCD cell = 3–

5×10^{-10} bar, mass resolution at 200 *m/z* = 17,500–35,000. Ion transfer settings were manually tuned to optimize transmission of the analyte ions and kept constant for all measurements to allow relative comparisons between samples. It was ascertained that the noise level remained below 1% relative signal intensity on a scan-to-scan basis. The mass spectrometer was externally calibrated using CsI clusters.

Native MS data analysis

To identify all species present in the mass spectra, they were converted in MassLynx-compatible RAW files using Databridge (Waters) and analyzed using MassLynx v4.1 (Waters). Species, which were consistently detected throughout the experiments, are monomeric Aur-A, monomeric Bora^{NT}, monomeric Plk1, the noncovalent Aur-A/Bora^{NT} 1:1 complex and the noncovalent Plk1/Bora^{NT} 1:1 complex. For the analysis of these species, the raw mass spectra were deconvoluted to zero charge state spectra using the ReSpect algorithm within Protein Deconvolution v3.0 (Thermo Fisher Scientific). The parameters used for deconvolution were as follows: mass tolerance = 0.05–0.12 Da, minimum peak significance = 1 standard deviation, noise rejection = 68–95%, number of iterations = 3, minimum adjacent charge states = 2, number of peak models = 1–2. All other parameters were either left at their default settings or manually optimized for the respective species. All deconvoluted spectra were manually compared to the raw mass spectra to ensure correct representation of all present charge states of each species.

In all cases, the molecular weights calculated with ReSpect were consistent (within 0.01%) with the theoretical molecular weights that were calculated based on the amino acid sequences of Aur-A, Bora^{NT} and Plk1 (see Figure S1), considering a 79.98 Da mass difference between the differentially phosphorylated isoforms of each species. The relative abundance of the thereby identified phospho-isoforms was derived from the ReSpect output and used to calculate the average phosphorylation state as described previously.[1] The relative protein complex abundances were respectively calculated based on the intensity ratio of Bora^{NT}-bound Aur-A phospho-isoforms to all Aur-A-containing phospho-isoforms ($\frac{\text{Aur-A}_{\text{bound}}}{\text{Aur-A}_{\text{bound}} + \text{Aur-A}_{\text{free}}}$) and the intensity ratio of Bora^{NT}-bound Plk1 phospho-isoforms to all Plk1-containing phospho-isoforms ($\frac{\text{Plk1}_{\text{bound}}}{\text{Plk1}_{\text{bound}} + \text{Plk1}_{\text{free}}}$).

Native IMS-MS

Samples were prepared and introduced into the mass spectrometer as described above for the native MS experiments. Analyses were performed on a Synapt G1 HDMS Q-TOF instrument (Waters) equipped with a traveling wave ion mobility cell. The instrument was operated with the settings listed below: source backing pressure = 5 mbar, capillary voltage = 1.3–1.4 kV, sampling cone voltage = 30 V, extraction cone voltage = 5 V, trap collision energy = 10 V, transfer collision energy = 20 V, trap gas

flow (N_2) = 8 ml/min, IMS gas flow (N_2) = 30 ml/min, IMS wave velocity = 150 m/s, IMS wave height linearly ramped from 4 V to 15 V over the full IMS cycle. Mass spectra were calibrated using CsI clusters and collisional cross section (CCS) calculation was facilitated by calibrating the drift time distributions as described previously,[2] comparing the drift times of denatured bovine ubiquitin, equine cytochrome C, and equine myoglobin to their reported CCS in N_2.[3] The data were processed using MassLynx v4.1 in combination with DriftScope v2.4 (Waters).

Top-down proteomics data acquisition

For top-down proteomics, Bora^{NT} was hyperphosphorylated with catalytic amounts of Aur-A and Plk1 (molar ratio 100:1:1) and subsequently denatured by buffer exchange into 0.1% (*v/v*) HCOOH during 6 centrifugations at 13,000 g using 5 kDa molecular weight cutoff centrifugal filter units with PES membranes (Sartorius). Denatured samples were diluted to 2 µM (Bora^{NT} concentration) in 40% (*v/v*) CH_3OH, 1% (*v/v*) HCOOH and directly infused into an Orbitrap Fusion instrument (Thermo Fisher Scientific) with a flow rate of 2 µL/min. Analyses were performed with 2 kV capillary voltage, 20 eV source fragmentation voltage, and an HCD cell pressure of 3 mTorr (intact protein mode). Precursor as well as fragment ion spectra were recorded in the Orbitrap mass analyzer with a mass resolution of 120,000 at 200 *m/z*. The 19+ charge states of 7–16-times phosphorylated Bora^{NT} were independently mass selected with the quadrupole mass analyzer at an isolation width of 2 *m/z* and fragmented using electron transfer dissociation (ETD) with the following settings: precursor automatic gain control (AGC) target value = 5×10^4, maximum precursor ion injection time = 200 ms, ETD reagent ion AGC target = 2×10^5, maximum ETD reagent ion injection time = 200 ms, ETD ion/ion reaction times = 4, 6 and 8 ms, no supplemental collisional activation.

Top-down proteomics data analysis

Fragment ion spectra were deconvoluted with Protein Deconvolution v3.0 (Thermo Fisher Scientific) applying the Xtract algorithm with the parameters listed below: S/N threshold = 1.1, fit factor = 40%, remainder threshold = 10%.

To map the phosphorylation sites, deconvoluted mass spectra were analyzed using an in-house developed program written in C#. In the first step, fragment ions of each acquisition were mapped on the Bora^{NT} amino acid sequence (Figure S1C), assigning them to the best matching theoretical fragment, considering b-, c-, and y-type fragment ions.[4] Subsequently, the masses of experimental and matched theoretical fragment ions were compared, for each spectrum individually calibrated around 0 ppm offset, and the resulting distribution of mass deviations was used to determine the allowed mass tolerance, which was found to be +/- 2 ppm. All fragment ions with mass deviations outside this window were assumed to be false positives and discarded.

Modified sites were then individually localized for each phospho-isoform by adding an increasing number of phosphorylations to the N- or C-terminus of the $Bora^{NT}$ sequence. Depending on the number of phosphorylations added, different sequence stretches are covered by fragment ions, indicating at which position a new phosphorylation occurs (Figure S8). Mapped phosphorylation sites were confirmed manually, assuming that a residue, which is unequivocally shown to be phosphorylated at a lower phosphorylation state, will be phosphorylated in all higher phospho-isoforms and, conversely, that a residue, which is unambiguously shown to be unmodified, cannot be modified in any lower phospho-isoform.

Chemical cross-linking and SDS-PAGE

Cross-linking reactions were conducted using bis(sulfosuccinimidyl)suberate (BS3, Thermo Fisher Scientific) in 20x molar excess (compared to the total protein concentration), incubated for 45 min at room temperature in 20 mM HEPES pH 7.5, 100 mM NaCl, and quenched by adding a surplus of Tris-HCl pH 7.5. Where indicated in the text, samples were subsequently dephosphorylated for 1 hr at 37°C with 1 U/µg calf intestinal alkaline phosphatase (New England BioLabs) in the presence of 10 mM $MgCl_2$. The reaction mixtures were analyzed by SDS-PAGE (Bio-Rad) and stained with GelCode Blue Stain Reagent (Thermo Fisher Scientific).

Bottom-up proteomics

Gel bands were excised and subjected to in-gel digestion according to the standard protocol.[5] Briefly, proteins were reduced with DTT, alkylated with iodoacetamide, digested with sequencing-grade modified porcine trypsin (Promega), and extracted from the gel with CH_3CN. The peptide mixtures were analyzed by nano-high performance liquid chromatography/tandem mass spectrometry (LC-MS/MS). LC-MS/MS was performed on an Orbitrap Elite instrument coupled to a Proxeon EASY-nLC 1000 (both Thermo Fisher Scientific) operating in reversed phase (columns packed ReproSil-Pur C18-AQ, 3 µm (Dr. Maisch GmbH)) at a flow rate of 300 nL/min. Peptides were loaded onto the analytical column using 100% solvent A (0.1% (*v/v*) HCOOH in H_2O) and eluted over 60 min by increasing the concentration of solvent B (0.1% (*v/v*) HCOOH in CH_3CN). Precursor as well as fragment ion spectra were recorded in the Orbitrap mass analyzer with a respective mass resolution of 60,000 or 15,000 at 200 *m/z*. Gas phase fragmentation was achieved by applying either electron-transfer/higher energy collision dissociation (EThcD) on the 10 most abundant precursor ions or sequential HCD and ETD on the 5 most abundant precursor ions, omitting singly charged precursor ions or, if cross-linked peptides were analyzed, singly and doubly charged precursor ions.

Peptide identification was performed using Proteome Discoverer v1.4 (Thermo Fisher Scientific). To identify potential contaminant proteins, a Mascot search was performed against the Swissprot database

(retrieved in August 2014), setting carbamidomethylation of Cys as fixed modification and oxidation of Met, attachment of DSS (equivalent to the BS3 cross-linker) to Lys (only for cross-linked samples) as well as phosphorylation of Ser, Thr, Tyr as dynamic modifications. Furthermore, the precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was set to 0.05 Da, trypsin was specified as protease and up to 2 missed cleavages were allowed. The identified peptide spectrum matches (PSMs) were verified using the Target Decoy PSM Validator, allowing a false-discovery rate (FDR) of 1%.

Relative quantitation of phosphorylation site occupancy

For phosphorylation site mapping, a smaller in-house database containing the Aur-A, Bora^{NT} and Plk1 sequences shown in Figure S1 was used for the Mascot search and phosphorylated residues were located by including the phosphoRS v3.1 node[6] in the Proteome Discoverer analysis workflow.

Relative quantitation of phosphorylation site occupancy was achieved by comparing the chromatographic peak areas of phosphorylated peptides and their unphosphorylated counterparts. Peak areas were determined with Xcalibur v2.2 (Thermo Fisher Scientific) or Progenesis QI for proteomics (Nonlinear Dynamics) and manually verified. The relative phosphorylation occupancy was then calculated as $\frac{[\text{peak areas of phosphorylated peptides}]}{[\text{peak areas of phosphorylated peptides + unphosphorylated counterparts}]} * 100.$

Identification of cross-linked peptides

Cross-linked peptides were identified with the in-house developed software XlinkX.[7] To this end, mass spectra were deconvoluted to charge state 1 using MS2 spectrum processor v0.9 within Proteome Discoverer v1.4 (Thermo Fisher Scientific). In XlinkX, a database of cross-linked peptides was generated based on the Aur-A, Bora^{NT} and Plk1 sequences (Figure S1) and the sequences of the *E.coli* proteins DnaK (UniProt: P0A6Y8) and GroEL (UniProt: P0A6F5), which were identified as minor contaminants in the initial Mascot database search. The XlinkX database was generated according to the following criteria: peptide length = 3–40 amino acids, allowed missed cleavages = 3, enzyme = trypsin, linked residue = Lys, fixed modification = Cys carbamidomethylation, variable modifications = ≤2 Met oxidations and ≤2 Ser, Thr, Tyr phosphorylations. Subsequently, the deconvoluted mass spectra were searched against the generated database in enumeration mode (for noncleavable cross-linkers), allowing a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 20 ppm. Cross-link validation was achieved through a target-decoy strategy as described previously,[7] searching the deconvoluted mass spectra against a database that contained 3 shuffled decoy sequences per target sequence. The cross-links were verified based on the total number of matched fragment ions (allowing 1% FDR) and the individual N scores of the linked peptides (allowing 5% FDR).[7,8] Mapping of the cross-links on structural models

was performed with Pymol v1.5 (Schrödinger, LLC) and schematic visualizations of cross-links were generated with xiNET.[9]

Extension (not in UniProt sequence) **Kinase Domain** Polo box domain

A Human Polo-like kinase 1 (Plk1), UniProt: P53350
expected molecular weight = 69713.48 Da

-12
RSMYPYDVPD YA

```
        10         20         30         40         50
MSAAVTAGKL ARAPADPGKA GVPGVAAPGA PAAAPPAKEI PEVLVDPRSR
        60         70         80         90        100
RRYVRGRFLG KGGFAKCFEI SDADTKEVFA GKIVPKSLLL KPHQREKMSM
       110        120        130        140        150
EISIHRSLAH QHVVGFHGFF EDNDFVFVVL ELCRRRSLLE LHKRRKALTE
       160        170        180        190        200
PEARYYLRQI VLGCQYLHRN RVIHRDLKLG NLFLNEDLEV KIGDFGLATK
       210        220        230        240        250
VEYDGERKKT LCGTPNYIAP EVLSKKGHSF EVDVWSIGCI MYTLLVGKPP
       260        270        280        290        300
FETSCLKETY LRIKKNEYSI PKHINPVAAS LIQKMLQTDP TARPTINELL
       310        320        330        340        350
NDEFFTSGYI PARLPITCLT IPPRFSIAPS SLDPSNRKPL TVLNKGLENP
       360        370        380        390        400
LPERPREKEE PVVRETGEVV DCHLSDMLQQ LHSVNASKPS ERGLVRQEEA
       410        420        430        440        450
EDPACIPIFW VSKWVDYSDK YGLGYQLCDN SVGVLFNDST RLILYNDGDS
       460        470        480        490        500
LQYIERDGTE SYLTVSSHPN SLMKKITLLK YFRNYMSEHL LKAGANITPR
       510        520        530        540        550
EGDELARLPY LRTWFRTRSA IILHLSNGSV QINFFQDHTK LILCPLMAAV
       560        570        580        590        600
TYIDEKRDFR TYRLSLLEEY GCCKELASRL RYARTMVDKL LSSRSASNRL

KAS
```

B Human Aurora kinase A (Aur-A), UniProt:O14965
expected molecular weight = 46066.68 Da

-1
RS

```
        10         20         30         40         50
MDRSKENCIS GPVKATAPVG GPKRVLVTQQ FPCQNPLPVN SGQAQRVLCP
        60         70         80         90        100
SNSSQRIPLQ AQKLVSSHKP VQNQKQKQLQ ATSVPHPVSR PLNNTQKSKQ
       110        120        130        140        150
PLPSAPENNP EEELASKQKN EESKKRQWAL EDFEIGRPLG KGKFGNVYLA
       160        170        180        190        200
REKQSKFILA LKVLFKAQLE KAGVEHQLRR EVEIQSHLRH PNILRLYGYF
       210        220        230        240        250
HDATRVYLIL EYAPLGTVYR ELQKLSKFDE QRTATYITEL ANALSYCHSK
       260        270        280        290        300
RVIHRDIKPE NLLLGSAGEL KIADFGWSVH APSSRRTTLC GTLDYLPPEM
       310        320        330        340        350
IEGRMHDEKV DLWSLGVLCY EFLVGKPPFE ANTYQETYKR ISRVEFTFPD
       360        370        380        390        400
FVTEGARDLI SRLLKHNPSQ RPMLREVLEH PWITANSSKP SNCQNKESAS

KQS
```

C Human Bora N-terminus (Bora^{NT}), UniProt:Q6PGQ7
expected molecular weight = 17470.62 Da

-4
GAASM

```
        10         20         30         40         50
MGDVKESKMQ ITPETPGRIP VLNPFESPSD YSNLHEQTLA SPSVFKSTKL
        60         70         80         90        100
PTPGKFRWSI DQLAVINPVE IDPEDIHRQA LYLSHSRIDK DVEDKRQKAI
       110        120        130        140        150
EEFFTKDVIV PSPWTDHEGK QLSQCHSSKC TNINSDSPVG KKLTIHSEKS

D
```

Figure S1. Amino acid sequences of the proteins used in this study. Residues with numbers below 1 (shown in italics) are part of the recombinantly produced protein constructs but do not belong to the corresponding UniProt sequence. Kinase domains are indicated by bold type and the Plk1 polo box domain is underlined.

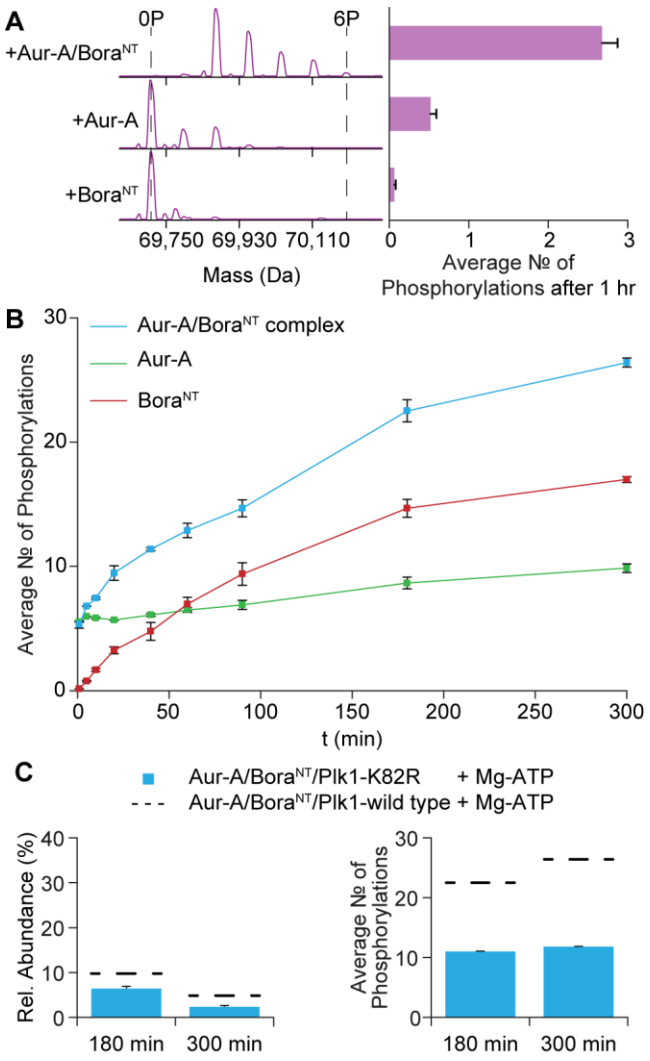

Figure S2. Native MS-based phosphorylation analysis. (A) Native MS of Plk1 after 1 hr incubation with stoichiometric amounts of Bora^{NT}, Aur-A or Aur-A/Bora^{NT} in presence of Mg-ATP. Mass spectra were deconvoluted from the *m/z* to the mass domain, yielding a zero charge state spectrum (left) and the relative phospho-isoform abundances were used to calculate the average phosphorylation state (right). Minimum and maximum number of Plk1 phosphorylations (P) are indicated in the spectra. (B) Phosphorylation progress curves of Aur-A, Bora^{NT} and the Aur-A/Bora^{NT} complex during the incubation of equimolar amounts of Aur-A, Bora^{NT} and Plk1 with Mg-ATP. (C) Relative abundance and average phosphorylation state of the Aur-A/Bora^{NT} complex after incubation with the kinase-inactive Plk1-K82R mutant and Mg-ATP. The phosphorylation and complex abundance levels that were obtained using Plk1-wild type (also shown in Figure 3) are displayed as black dashed lines. Error bars represent standard deviations from duplicates.

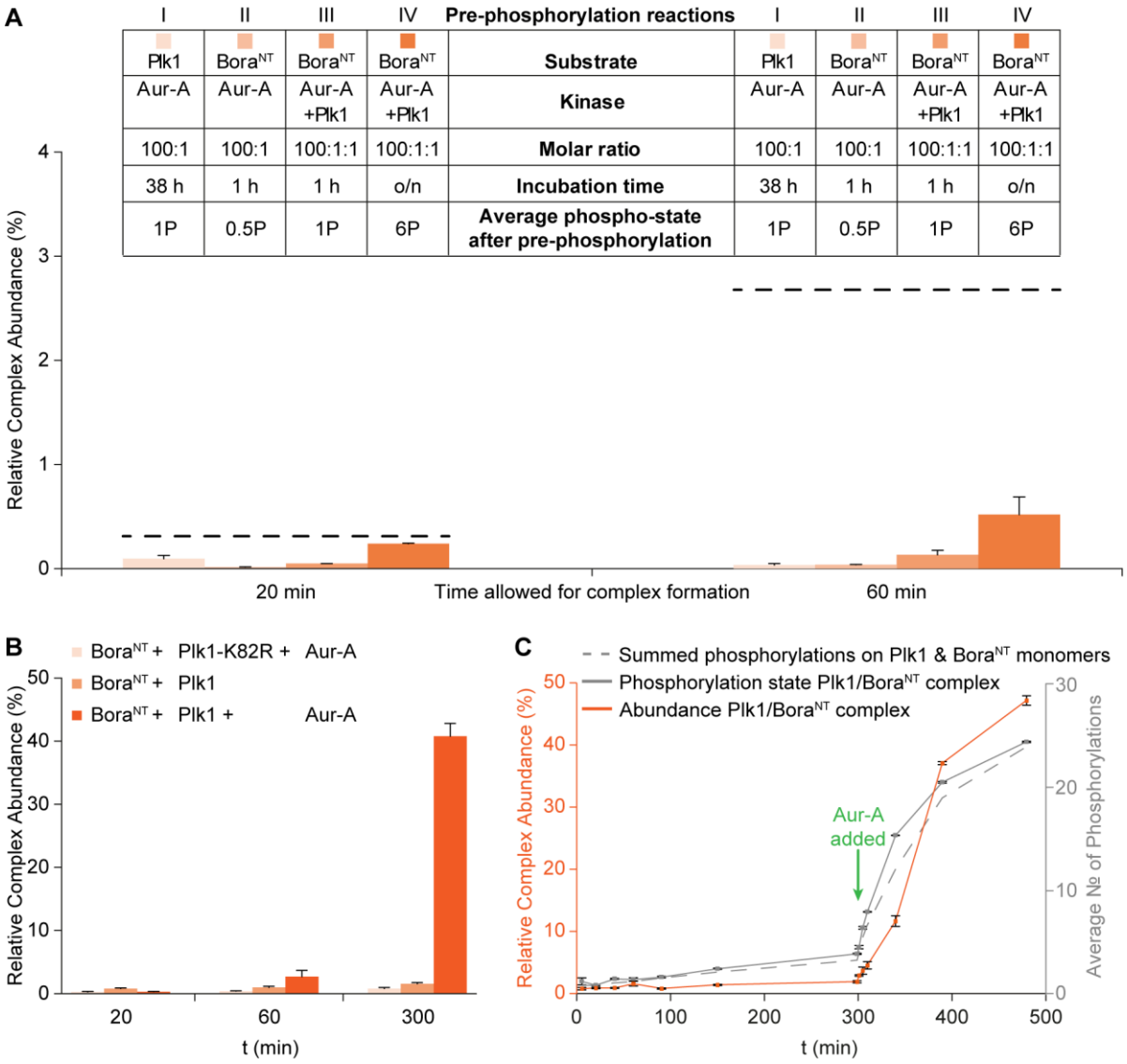

Figure S3. Determinants of Plk1/Bora^{NT} complex formation. (A) Complex formation after pprephosphorylation of Bora^{NT} or Plk1. The pprephosphorylation reactions (conditions indicated in the table, o/n = overnight) were followed by the addition of equimolar amounts of the unphosphorylated interaction partner and fresh Mg-ATP. Complex formation was monitored for up to 1 h. Results are compared to the relative Plk1/Bora^{NT} complex abundance observed in the initial experiment, where Aur-A, Bora^{NT} and Plk1 were incubated at a 1:1:1 molar ratio without pprephosphorylation (shown as black dashed line). (B) Complex formation in presence of different kinase combinations. Bora^{NT} was incubated with wild-type Aur-A and wild-type Plk1 together (see also Figure 4A), with wild-type Plk1 only, or with wild-type Aur-A in combination with the kinase-dead Plk1-K82R mutant. The extent of complex formation was monitored for up to 5 h. (C) Time-course of Plk1/Bora^{NT} complex abundance and phosphorylation state before and after addition of Aur-A. Substantial Plk1/Bora^{NT} complex formation was only detected after adding Aur-A to the reaction mix. Enhanced Plk1/Bora^{NT} complex formation was accompanied by phosphorylation levels that are clearly higher than the summed phosphorylation states

of the Plk1 and Bora^{NT} monomers (compare dashed vs. solid gray line). This is indicative for a preferred association of Plk1 with hyperphosphorylated Bora^{NT}. Error bars represent standard deviations from duplicates.

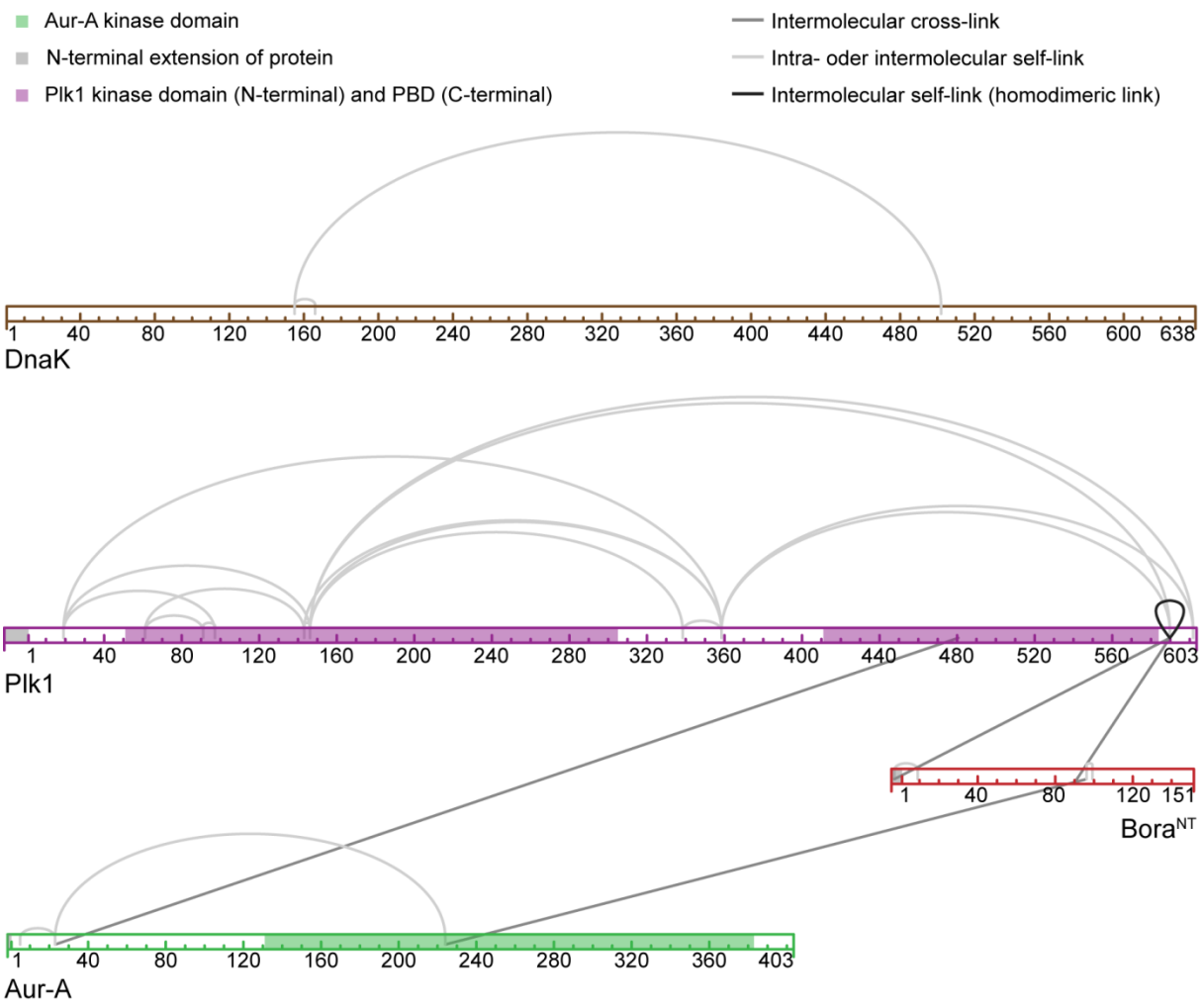

Figure S4. Cross-linking-MS analysis of the 120–160 kDa band identified by SDS-PAGE (Figure 5A). The corresponding gel band represents a mixture of different low abundant species. The identification of one intermolecular Plk1/Aur-A, Plk1/Bora^{NT} and Aur-A/Bora^{NT} cross-link may be explained by the (co-)existence of an Aur-A/Bora^{NT}/Plk1 1:1:1 complex (~134 kDa), a Plk1/Aur-A 1:1 complex (~116 kDa) and a Plk1/Bora^{NT} 2:1 complex (~157 kDa). The presence of Bora^{NT}-bound Plk1 dimers, and possibly also free Plk1 dimers (~139 kDa), is further supported by a homodimeric link connecting two Plk1 molecules via Lys589. Furthermore, the DnaK self-links point to a DnaK dimer of ~138 kDa.[10] All these possible noncovalent complexes showed very low signal intensities (<1%) in native MS and were only temporarily detected (data not shown). This, as well as the low number of identified intermolecular cross-links, indicates that these complexes originate from low-affinity or less stable interactions.

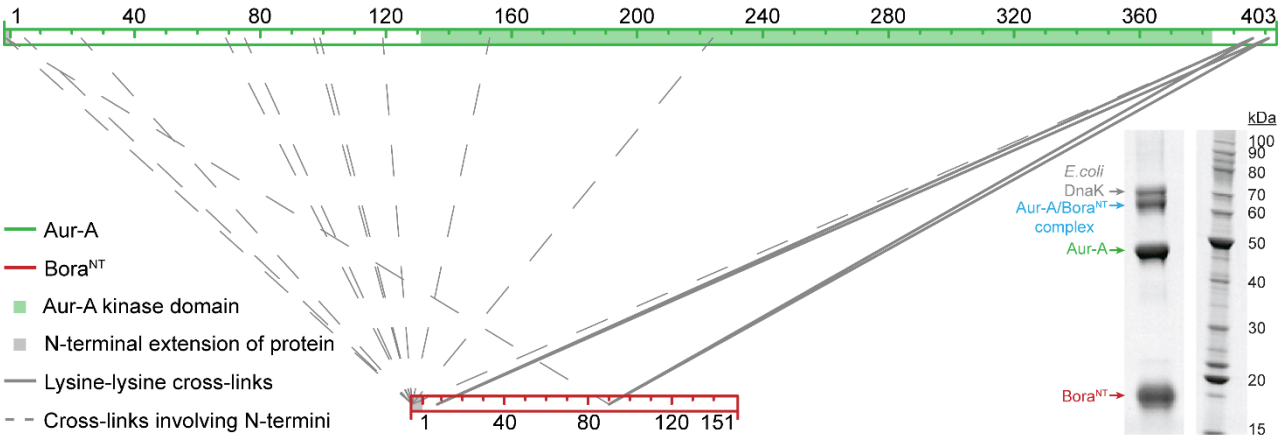

Figure S5. Cross-linking-MS analysis of the Aur-A/Bora^{NT} complex. Map of intermolecular Aur-A/Bora^{NT} cross-links identified from the gel band that represents the Aur-A/Bora^{NT} complex. The N-terminal extensions were introduced to facilitate the recombinant production of the proteins (Figure S1). The SDS-PAGE analysis of the corresponding Aur-A/Bora^{NT} reaction mix, which was cross-linked after 1 h incubation with Mg-ATP, is shown as an inset.

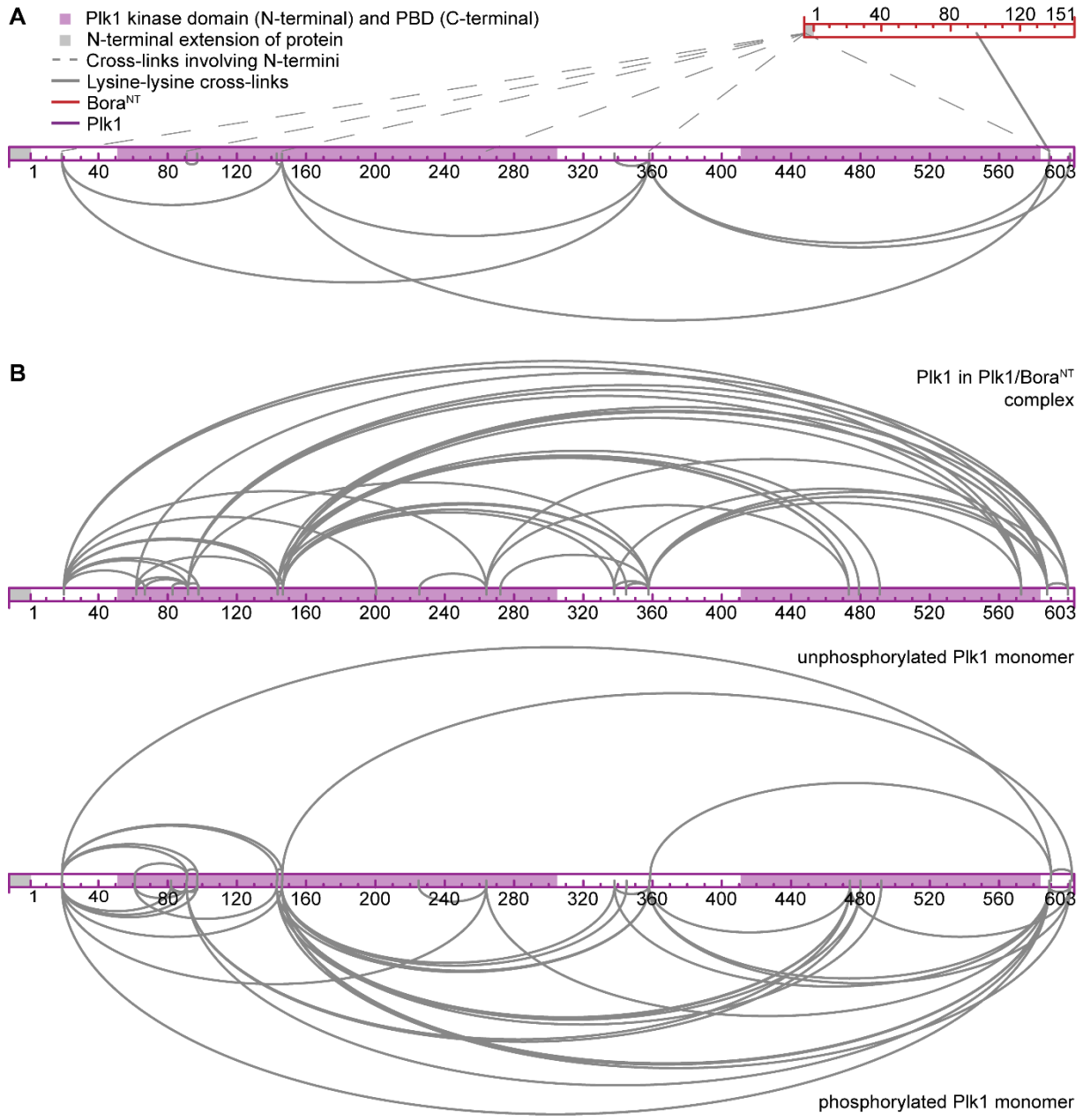

Figure S6. Structural analysis of Plk1 by cross-linking-MS. (A) Map of intramolecular Plk1 and intermolecular Plk1/Bora^{NT} cross-links identified from the corresponding gel band (indicated by an orange arrow in Figure 5A). The N-terminal extensions were introduced to facilitate the recombinant production of the proteins (Figure S1). (B) Intramolecular cross-links within Bora^{NT}-bound Plk1 after 5 h incubation with Aur-A and Mg-ATP (upper sequence bar) and within monomeric Plk1 after 5 h incubation with Aur-A, in the absence (cross-links above lower sequence bar) or presence (cross-links below lower sequence bar) of Mg-ATP.

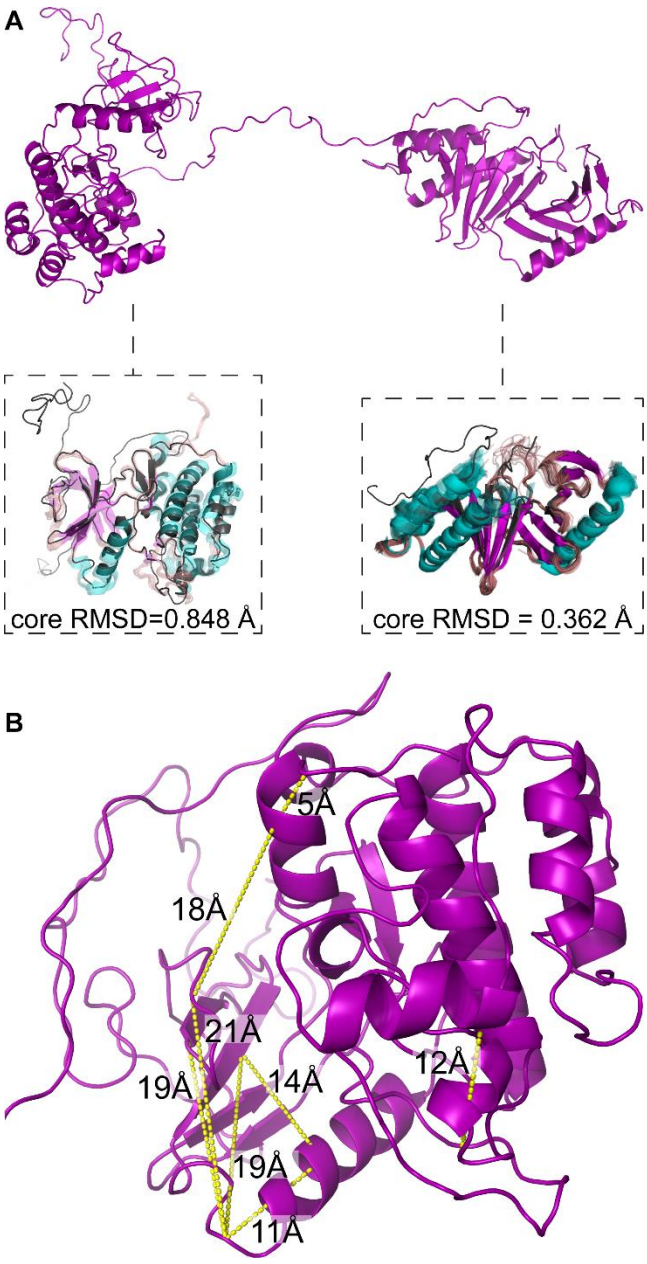

Figure S7. Structural mapping of intramolecular Plk1 cross-links. (A) Multi-template homology model of full-length Plk1 used for cross-link mapping in Figure 5B. The model was generated using the Phyre2 server.[11] The insets show structural alignments of the model (displayed as black cartoon representation) with published crystal structures of the Plk1 kinase domain (PDB: 2OU7, 2RKU, 2OWB, 2YAC, 4J52, 3FC2, 4J53, 4A41, 3KB7, 3THB, 4A4O, 2V5Q , left) and the Plk1 polo box domain (PDB: 4X9R, 4X9V, 3Q1I, 4O6W, 3P34, 3RQ7, 4DFW, 3FVH, 4RCP, 3P36, 4O56, 4LKL, 3HIH, 3P2W, 4WHK, 4E9C, 4O9W, 3HIK, 4X9W, 3P2Z, 2OGQ, 4WHH, 4H71, 4H5X, 4HY2, 3BZI, 4LKM, 1UMW.A, 3P35, 4E67, 1Q4O, 1Q4K, 3P37, 4WHL, 3C5L, 4HCO, 2OJX, 4HAB, 4E9D, right). Crystal structures that were used to build the homology model were excluded from the alignment. The crystal structures are depicted as half-transparent cartoons colored according to their secondary structure (α-helices = cyan, β-strands = purple, ω-loops = brown). The root mean square deviation (RMSD) of the

aligned regions was calculated using Pymol v1.5 (Schrödinger, LLC). (B) Mapping of cross-links identified within the Plk1 kinase domain.

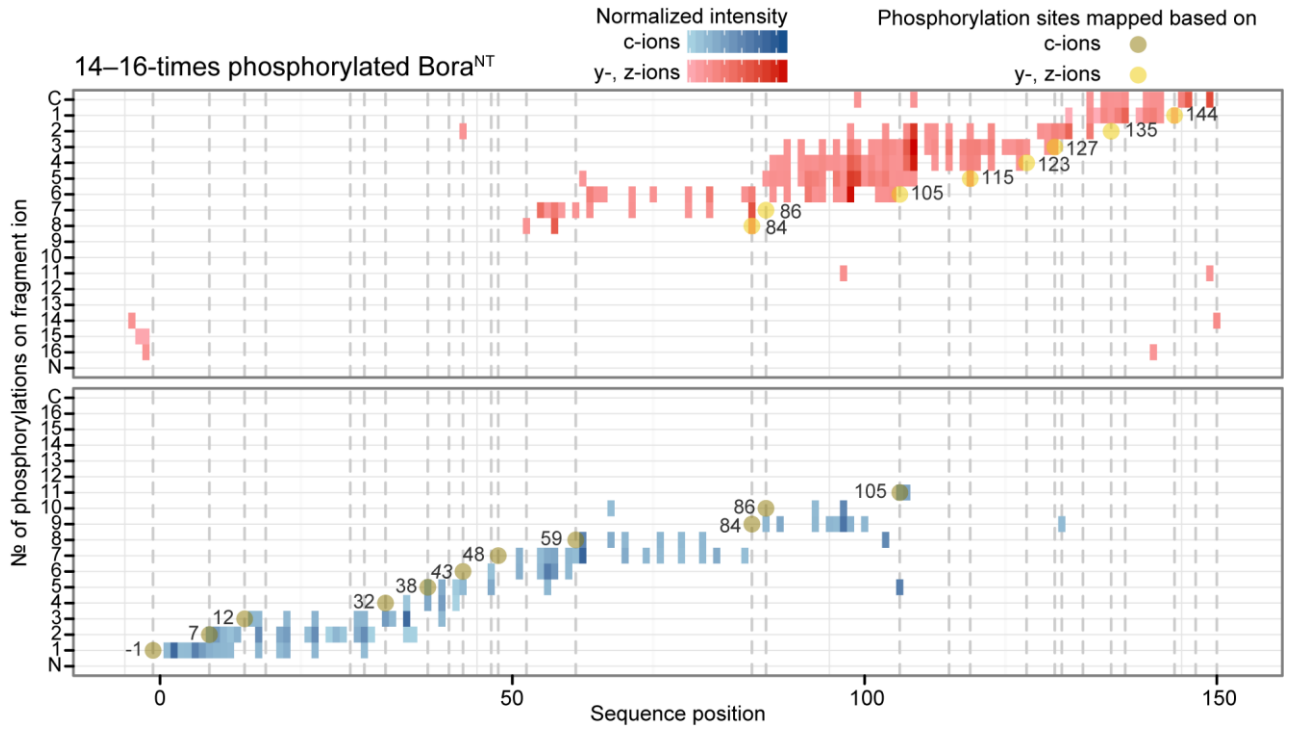

Figure S8. Top-down proteomics analysis of Bora^{NT} hyperphosphorylation. Exemplary heat map used for the localization of phosphorylation sites based on top-down proteomics. An increasing number of phosphorylations was added to the N- (lower plot) or C-terminus (upper plot) and the fragment ions harboring the respective amount of phosphorylations are plotted according to their sequence position (C-terminal y- and z-ions or N-terminal c-ions, respectively). This leads to a staircase-like arrangement that allows to identify the most likely phosphorylation site, i.e., the residue after which fragment ions with a higher phosphorylation state were detected. Ser43 is shown in italics, because this site was unambiguously localized based on lower phospho-isoforms, whereas the plot shown here would suggest phosphorylation of Ser47 instead, since no fragment ions between Ser43 and Ser47 were identified. However, the added information from all phospho-isoforms allows unequivocal mapping of all phosphorylation sites (Table S3).

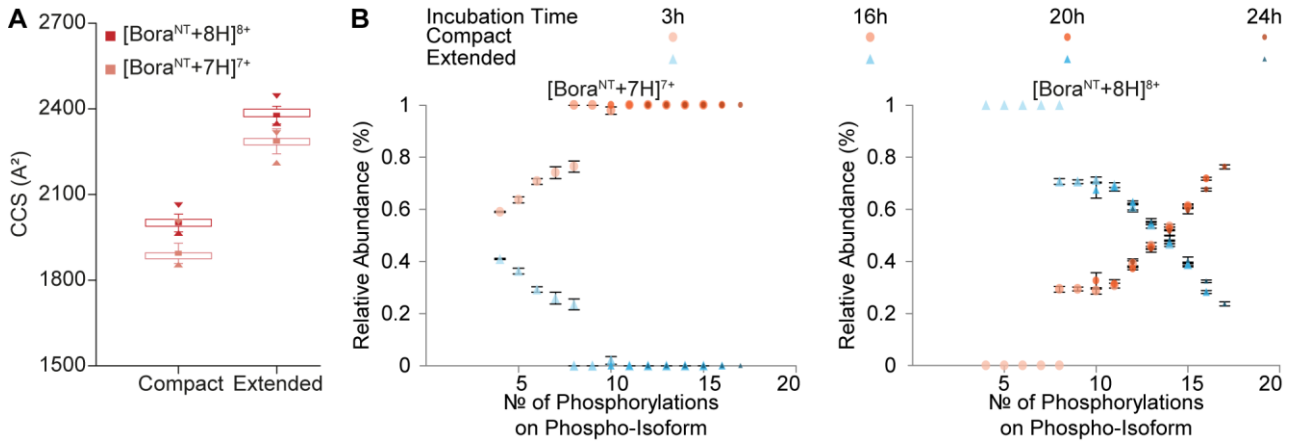

Figure S9. IMS-MS analysis of Bora^{NT} during hyperphosphorylation with catalytic amounts of Aur-A and Plk1. (A) Collisional cross sections (CCS) calculated for the compact and extended conformation of the 7+ and 8+ charge state of Bora^{NT}. The CCS were calculated for two independent reaction mixes at four time points. They are shown as box-and-whisker plots with the boxes harboring the 1st–3rd quartile and median (■), 1st percentile (▲), and 99th percentile (▼) being indicated. The error bars represent 1.5 standard deviations calculated over all detections of the respective Bora^{NT} conformer. (B) Relative abundance of compact and extended conformations in the different phospho-isoforms of 7- and 8-times protonated Bora^{NT}. Error bars represent standard deviations between two independent reaction mixes. No error bars are shown for a standard deviation of 0.

Table S1. Overview of all identified cross-linked peptides. The corresponding table is provided in a separate Excel document.

Table S2. Relative quantitation of the Bora^{NT} phosphorylation site occupancy by bottom-up proteomics. Samples were taken from the SDS-PAGE analysis of the Aur-A/Bora^{NT}/Plk1 reaction mix (Figure 5A). The columns correspond to the different incubation times, with asterisks denoting the slower migrating Bora^{NT} band at time points where intermediately and hyperphosphorylated Bora^{NT} were present in the same lane of the gel. Differential phosphorylation detected in these samples is indicated by bold type. Fields are colored with a gradient from light to dark red, according to the determined percentage of phosphorylation. Phosphorylation sites, for which no peptide was identified are indicated by gray shading and 'nf' (not found). If the phosphorylation occupancy could not be calculated, the respective site is listed as 'nd' (not determined).

Sequence motif	Residue Nº (UniProt: Q6PGQ7)	Phosphorylation site occupancy (%)							
		10 min	20 min	40 min	40 min*	60 min	60 min*	120 min	300 min
GAA **S** MMG	-1	10	33	76	98	91	100	99	100
VKE **S** KMQ	7	0	47	**50**	**97**	**34**	**100**	98	99
MQI **T** PET	12	0	0	0	2	0	2	13	57
TPE **T** PGR	15	0	0	0	0	0	0	0	0
PFE **S** PSD	27	0	0	0	nf	nf	nf	0	nf
ESP **S** DYS	29	0	0	0	nf	nf	nf	0	nf
SDY **S** NLH	32	0	78	50	nf	nf	nf	0	nf
HEQ **T** LAS	38	0	0	0	nf	nf	nf	0	nf
TLA **S** PSV	41	0	0	0	nf	nf	nf	0	nf
ASP **S** VFK	43	0	78	100	nf	nf	nf	0	nf
VFK **S** TKL	47	0	0	0	0	0	0	0	0
FKS **T** KLP	48	0	0	**0**	**55**	**1**	**83**	100	89
KLP **T** PGK	52	0	0	0	0	0	0	0	0
FRW **S** IDQ	59	nf	100	100	100	100	100	nf	100
QAL **Y** LSH	82	1	3	0	0	0	0	0	0
LYL **S** HSR	84	12	62	97	100	100	100	100	100
LSH **S** RID	86	4	60	**53**	**99**	**78**	**100**	99	96
EFF **T** KDV	105	0	0	0	5	0	2	24	5
IVP **S** PWT	112	0	0	0	0	0	nf	0	0
SPW **T** DHE	115	0	0	0	0	0	nf	0	10
KQL **S** QCH	123	52	89	100	nf	100	nf	100	nf
QCH **S** SKC	127	0	0	50	nf	0	nf	50	nf
CHS **S** KCT	128	0	0	50	nf	0	nf	50	nf
SKC **T** NIN	131	nd	0	1	0	10	0	89	nf
NIN **S** DSP	135	0	0	0	0	0	0	0	nf
NSD **S** PVG	137	0	0	0	0	0	0	0	nf
KKL **T** IHS	144	80	98	99	100	99	100	100	100
TIH **S** EKS	147	0	0	1	5	0	15	21	38
SEK **S** D	150	0	0	0	2	0	3	10	27

Table S3. Phosphorylation sites of 7- to 16-times phosphorylated Bora^{NT} determined by top-down proteomics. Most phospho-isoforms comprise several positional isomers with different combinations of phosphorylated residues. Therefore, phospho-isoforms often contain more (partially) phosphorylated residues than the number of phosphorylations on the intact protein would suggest. The initial phosphorylation sites of Aur-A and Plk1, which were determined by top-down proteomics of singly and doubly phosphorylated Bora^{NT},[12] are shown in bold type. The transition to hyperphosphorylated Bora takes place when the protein becomes 11–13× phosphorylated, as indicated by native MS (Figure 4B). The respective phospho-isoforms are surrounded by a bold black line.

		Nº of phosphorylations on Bora^{NT}									
Residue Nº (UniProt: Q6PGQ7)		7x	8x	9x	10x	11x	12x	13x	14x	15x	16x
Ser	-1	x	x	x	x	x	x	x	x	x	x
Ser	7		x	x	x	x	x	x	x	x	x
Thr	12				x	x	x	x	x	x	x
Thr	15										
Ser	27										
Ser	29									x	x
Ser	32										
Thr	**38**	x	x	x	x	x	x	x	x	x	x
Ser	41										
Ser	43	x	x	x	x	x	x	x	x	x	x
Ser	47										
Thr	48						x	x	x	x	x
Thr	52										x
Ser	**59**	x	x	x	x	x	x	x	x	x	x
Ser	**84**	x	x	x	x	x	x	x	x	x	x
Ser	86		x	x	x	x	x	x	x	x	x
Thr	105									x	x
Ser	112										
Thr	115	x	x	x	x	x	x	x	x	x	x
Ser	123	x	x	x	x	x	x	x	x	x	x
Ser	127							x	x	x	x
Ser	128										
Thr	131										
Ser	135	x	x	x	x	x	x	x	x	x	x
Ser	137										
Thr	**144**	x	x	x	x	x	x	x	x	x	x
Ser	147										
Ser	150										

SI References

(1) van de Waterbeemd, M. J.; Lössl, P.; Gautier, V.; Marino, F.; Yamashita, M.; Conti, E.; Scholten, A.; Heck, A. J. Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angew. Chem., Int. Ed.* **2014**, *53*, 9660–9664.

(2) Smith, D. P.; Knapman, T. W.; Campuzano, I.; Malham, R. W.; Berryman, J. T.; Radford, S. E.; Ashcroft, A. E. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom.* **2009**, *15*, 113–130.

(3) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **2010**, *82*, 9557–9565.

(4) Kruger, N. A.; Zubarev, R. A.; Horn, D. M.; McLafferty, F. W. Electron capture dissociation of multiply charged peptide cations. *Int. J. Mass Spectrom.* **1999**, *185-187*, 787–793.

(5) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1*, 2856–2860.

(6) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **2011**, *10*, 5354–5362.

(7) Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **2015**, *12*, 1179–1184.

(8) Liu, F.; van Breukelen, B.; Heck, A. J. Facilitating protein disulfide mapping by a combination of pepsin digestion, electron transfer higher energy dissociation (EThcD), and a dedicated search algorithm SlinkS. *Mol. Cell. Proteomics* **2014**, *13*, 2776–2786.

(9) Combe, C. W.; Fischer, L.; Rappsilber, J. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **2015**, *14*, 1137–1147.

(10) Sarbeng, E. B.; Liu, Q.; Tian, X.; Yang, J.; Li, H.; Wong, J. L.; Zhou, L.; Liu, Q. A functional DnaK dimer is essential for the efficient interaction with Hsp40 heat shock protein. *J. Biol. Chem.* **2015**, *290*, 8849–8862.

(11) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858.

(12) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteoform characterization. *Anal. Chem.* **2015**, *87*, 4152–4158.