

Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism

Na Liu^{1,2,4}, Yue Guo^{1,3,4}, Shangbo Ning^{1,2} & Mojie Duan¹  [✉]

Phosphorylation is one of the most common post-translational modifications. The phosphorylation of the kinase-inducible domain (KID), which is an intrinsically disordered protein (IDP), promotes the folding of KID and binding with the KID-interacting domain (KIX). However, the regulation mechanism of the phosphorylation on KID is still elusive. In this study, the structural ensembles and binding process of pKID and KIX are studied by all-atom enhanced sampling technologies. The results show that more hydrophobic interactions are formed in pKID, which promote the formation of the special hydrophobic residue cluster (HRC). The pre-formed HRC promotes binding to the correct sites of KIX and further lead the folding of pKID. Consequently, a flexible conformational selection model is proposed to describe the binding and folding process of intrinsically disordered proteins. The binding mechanism revealed in this work provides new insights into the dynamic interactions and phosphorylation regulation of proteins.

¹Key Laboratory of magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan 430071, People's Republic of China. ²School of biological and pharmaceutical engineering, Wuhan Polytechnic University, Wuhan 430023, People's Republic of China. ³University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China. ⁴These authors contributed equally: Na Liu, Yue Guo. ✉email: mjduan@wipm.ac.cn

As one of the most common post-translational modifications (PTMs), phosphorylation is important in regulating protein synthesis, cell cycle, growth, apoptosis, cell division, and signal transduction^{1–3}. Most of the phosphorylation sites are located in the intrinsically disordered proteins (IDPs) or intrinsically disordered regions^{4,5}. The disordered nature of these proteins or regions facilitate the access for post-translational modification. Meanwhile, the functions of IDPs are regulated by the phosphorylation^{6,7}. However, the detailed mechanism about the phosphorylation regulation on the structures and the interactions of IDPs remain elusive, which greatly impede the understanding of IDP function and the searching for “druggable” IDP targets^{5,8}.

In many cases, IDPs fold into the ordered structures upon binding to their function partner, which is termed as the “coupled folding and binding process”^{9–13}. The kinase-inducible domain (KID) is a typical example which carry out its biological function through coupled binding and folding mechanism. As a part of the cAMP-response element binding protein (CREB), KID function as an inducible transcriptional activator^{14,15}. Upon phosphorylation at Ser133, the phosphorylated KID (pKID) contacts with the KIX domain on the transcriptional coactivator CREB-binding protein (CBP) and modulates the target gene expression^{16,17}. Two transient helices are present on the free-state pKID, i.e., α_A (from residues 120 to 129) and α_B (from residues 132 to 144), however, the α_A and α_B regions would fold into stable helical structures in the KIX-bound state^{16,18}.

The phosphorylation on residue Ser133 would increase the binding affinity of KID and KIX by almost two orders of magnitude^{17,19–21}. Several models have been proposed to describe the phosphorylation modulation mechanism on the binding of pKID and KIX. One of them suggested that phosphorylation on Ser133 could shifts the conformation ensemble of pKID toward the configuration similar to the structure in pKID-KIX complex^{18,22}. On the other side, many researchers believed that the increased affinity of pKID is induced by the intermolecular interactions related to the di-anionic phosphate group^{17,23}. Another model proposed that the phosphorylation restrict the flexibility of the loop region on KID and therefore reduce the entropic cost for KIX binding²⁴. Although many progresses have been achieved, the detailed mechanism of phosphorylation regulation on the binding process of pKID and KIX remains obscure.

It is great challenge to determine the structures of the intermediate states and the binding process due to the short lifetime of the intermediates²⁵. One major debate about the pKID-KIX binding is the order of α_A and α_B folding in the intermediates and binding process. A recent kinetic experiment shows that the binding and folding of α_B region is prior to the binding and folding of α_A region²⁶. Most of the large Φ -value residues were located on the α_B region, indicate that the native contacts or binding of this region were completed in the transition state²⁶. Nevertheless, the NMR researches show that the chemical shift values of the residues in the C-terminus of α_A (residues 128–132) in the intermediates are close to the bound state, which means the C-terminus of α_A are almost folded in the intermediate. On the other side, the chemical shifts differences of the α_B region in the intermediates and bound state are larger than the α_A region, indicate the α_B region are less folding than α_A in the intermediates²⁷. To elucidate the above controversies requires the characterizing the binding intermediates and binding pathways.

In this work, the structure properties of free pKID or KID and the binding free energy surface of pKID/KID and KIX were characterized based on the enhanced sampling simulations. The results show that both free pKID and KID are mainly disordered with some transient helical structures on them. The secondary structure compositions of the pKID and KID are basically the

same to each other, however, more long-range residue–residue interactions were observed in the pKID. The contacts between the hydrophobic residues on pKID would form special hydrophobic residue cluster (HRC). Although both of the KID and pKID would form encounter complex with KIX, only the structures with HRC which might pre-formed in free pKID would binding to the correct binding sites on KIX and leading to the folding and binding of pKID to form final complex. We proposed that the binding mechanism of the intrinsically disordered pKID would follow a flexible conformational selection mechanism.

Results

Structure ensembles of free pKID and KID. PTMetaD-WTE method was employed to obtain the structure ensembles of free pKID and KID in solution. The quality of the simulated structure ensembles was evaluated by the prediction accuracy of secondary chemical shifts (δ_{cs}). As shown in Supplementary Fig. 1, the predicted chemical shifts are agree well with the NMR measurements¹⁸. The RMSE of $C\alpha$ δ_{cs} between simulated and experimental results is 0.44 ppm for pKID and 0.47 ppm for KID. The RMSE of $H\alpha$ chemical shift for pKID and KID are 0.08 ppm and 0.06 ppm, respectively. The RMSE values are close to the system errors of the chemical shift calculating tool²⁸. The results indicate that the structure ensembles obtained in our work provide a reasonable description of the structural properties of the IDPs. Based on the reliable structure ensemble, we found that both free pKID and KID are mainly disordered in solution, however, some transient helical structures were observed on the α_A (residues 120–129) and α_B (residues 134–144). The helical propensity of α_A and α_B were given in Fig. 1a. The N-terminal of KID and pKID have higher helical propensity than those on C-terminus. The helicity of α_A are about 50% in both pKID and KID. The average residue-helicity on α_B region of pKID (18.9%) is slightly higher than that on KID (14.6%), which is consistent with the experimental observations (Supplementary Table 1)¹⁸.

Although the secondary structure compositions in pKID and KID are similar to each other, some obvious differences were observed on their tertiary structures. Based on the residue-residue contact maps (Fig. 1b), more interactions between two terminals (marked by black circle in Fig. 1b) were observed in pKID. The larger amount of residue contact probability indicates more compact structures formed after the phosphorylation. The hydrophobic residues (Leu128, Tyr134, Leu137, Leu138) around the pSer133 in pKID are more likely to form hydrophobic interactions than in the KID. The spatial closed hydrophobic residues would form a HRC. The contact number between the side-chain heavy atoms on the residues Leu128, Tyr134, Leu137, and Leu138 were calculated to quantitatively define the formation of HRC. The conformations with contact number larger than 15 are defined to be HRC structures. As can be seen in Fig. 1c, the HRC structures present on more than 40% conformations of free pKID. On the other side, almost no HRC formed on free KID as the probability of conformations with contact number larger than 8 is close to zero.

The binding free energy landscape of pKID-KIX. To characterize the binding and folding mechanism of pKID and KIX, the free energy landscapes (FEL) along the reaction coordinates were constructed. Figure 2 gives the free energy landscape as a function of the center-of-mass (COM) distances and the number of native contacts between pKID and KIX. Sugase et al.²⁷ proposed that the folding and binding process of pKID-KIX can be described by four-site exchange model, i.e., the free state, the encounter complex, the folding intermediates, and the bound state. All the four stages were clearly present on our simulated

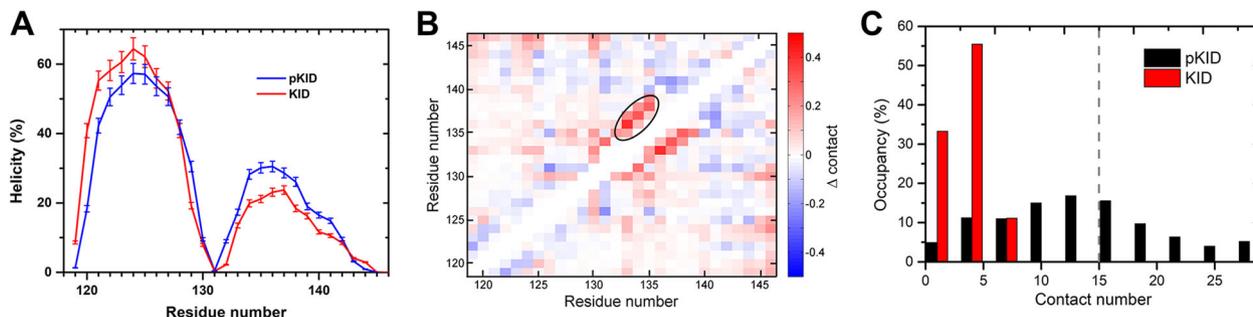


Fig. 1 Structure properties comparison of free KID and pKID. **a** The helicity of free pKID and KID. The errors are corresponding to the helicity fluctuation in the last 100 ns simulations, the helicity of residues were calculated from the initial time to different ending time, i.e., 220 ns, 240 ns, 260 ns, 280 ns and 300 ns. The standard deviations of the helicity are given as the error bars. **b** The intra-chain contact probability difference between the free pKID and KID. The contacts are calculated between the side-chain heavy atoms of the residues on the protein. Two residues are defined to be in contact if the distance between any two heavy atoms in different residues is <4.5 Å. Δ contact equals to the contact probability in pKID minus contact probability in KID. Red grid represents the residue-residue contact probability in pKID larger than that in KID, and blue grid represents the contact probability in KID larger than that in pKID. **c** Distributions of number of hydrophobic residue contact atoms in free pKID and KID. The hydrophobic residues include residues L128, Y134, L137, and L138.

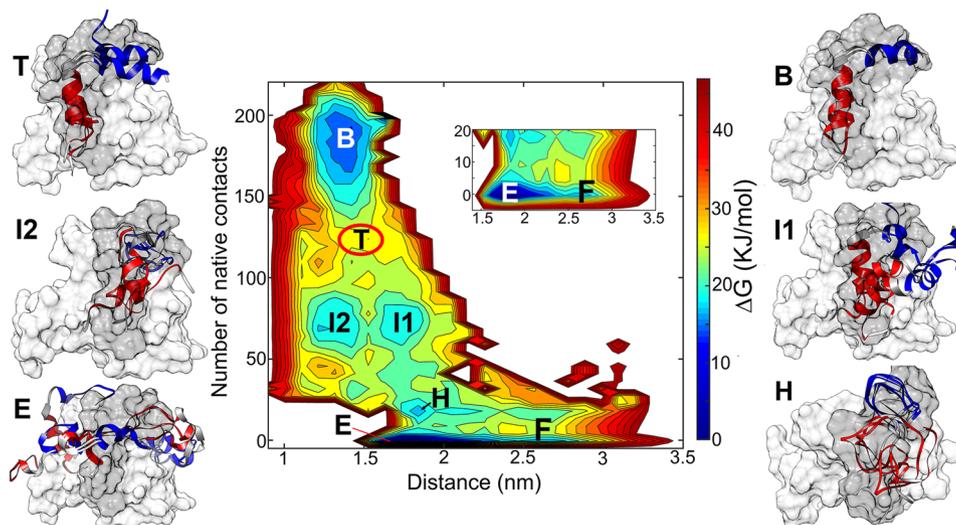


Fig. 2 Free energy landscape (FEL) of pKID and KIX binding process. The FEL is the function of the COM distance and native contacts number. Representative structures of the minima are given and overlapped with the experimental NMR structure. The NMR complex structure complex is shown in surface, and the pKID is colored in dark-gray and KIX is colored in gray. The simulated structures of pKID are represented in the cartoon style, and the α_A regions of simulated pKID are colored in blue, α_B regions are colored in red. The coordinates in PDB format of the representative structures is given in Supplementary Data 1.

FEL, i.e., the free state with the large distance and low contacts (the state marked by F), the encounter complex with distance close to 1.7 nm and the fraction of native contact Q close to 0 (marked by E), the intermediates with many native contacts formed (state I1 and I2) and the bound state with most of the native contact formed (state B). The structure properties of the important free energy minima are given in the following section.

The free state. In the free state, the pKID peptides are far away from KIX, the center of mass distances between pKID and KIX are larger than 25 Å. In the free-state conformations, neither native contacts nor non-native contacts are formed (Supplementary Fig. 2). The secondary structure composition of pKID in this state is similar to the apo-pKID, the helical contents on the α_A region is 52.0% and on α_B region is 10.2%.

The encounter complex state. Based on the ^1H - ^{15}N HSQC spectrum and ^{15}N R_2 dispersion experiments²⁷, Sugase et al.

found ensemble of weak complex which fast exchanging with the free pKID, and these states were defined to be encounter complex. Similar to the experimental observation, the interactions between pKID and KIX in the encounter complex state determined by our simulations are dynamic. Multiple hotspot binding sites present on the KIX and pKID (Fig. 3a), including the hydrophobic interactions between residue Leu128 on pKID with Leu664 on KIX, residue Leu141 on pKID with Val635, Met639 and Leu652 on KIX. In addition, the interactions between the charged residues also contribute to the formation of the encounter complex, such as Arg135 on pKID and Glu648 on KIX, Asp140 on pKID, and Lys659 on KIX. It should be noticed that most of the interactions in the encounter complex are non-native contacts, which indicate the non-native interactions are major force to drive the formation of pKID-KIX encounter complex. On the other side, the phosphorylated serine (pSer133) would not direct contribute to the formation of encounter complex since there is no interactions between the pSer133 and the residues on KIX.

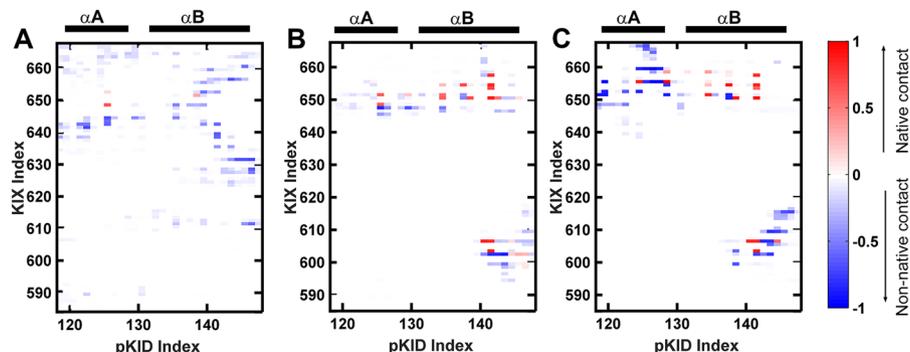


Fig. 3 Contact maps between residues on pKID and KIX. **a** Contact map between residues of pKID and KIX in the encounter complex (state E). **b** Contact map of intermediate 1 (state I1); **c** Contact map of intermediate 2 (state I2); In these contact map, red grid represents the native contacts are formed, blue grids are corresponding to the non-native contacts.

Intermediates and hidden intermediate. Two intermediates (I1 and I2) with similar native contact numbers but different COM distances were observed on the free energy landscape. Many residues of α_B regions are correctly anchoring with the native-binding sites on the KIX in both I1 and I2 intermediates, for example, the hydrophobic interactions between the residues Tyr134, Ile137, Leu138, Leu141 on pKID and the residues Tyr650, Ala654, Ile657, Tyr658 on KIX. Besides, the salt bridges interactions between residues Asp140 on pKID and Lys606 on KIX may also contribute to the binding. The residues on α_A region are more flexible and less contacts formed with KIX than residues in α_B region in the intermediates. The contacts between α_A region and KIX are different in I1 and I2, where Arg125 on pKID contact with Glu648 and His651 on KIX I1 (Fig. 3b), however, Arg124 mainly contact with Glu655 in intermediate I2 (Fig. 3c). The non-native contacts also contribute to the stabilization of the intermediates. However, more non-native interactions are formed in the intermediate I2, especially the residues on α_A region of I2 (Fig. 3c).

Although the native contact numbers are similar for intermediates I1 and I2, the secondary structure compositions are different in the two intermediates. In I1, the C-terminal of α_A are basically folded to helical structures (the helicity from 124 to 128 is close to 90%); the helicity of α_B in I1 is slightly lower than α_A (the helicity of these intermediates shown in Supplementary Fig. 3 and representative structures displayed in Fig. 2). This is consistent with the NMR characterization of the intermediate state²⁷ in which residues 124–128 in α_A nearly fully folded and the helicity of residues 133–138 and 141 in the α_B region is only about 70%. However, the pKID mainly adopt disordered structure in state I2, only several residues in α_B region have ~20% probability forming α -helix²⁷. By analyzing the FEL along α_A or α_B helicity and native contact number (Supplementary Fig. 4), we found that the structures of I2 are located on the off-pathway regions of the FEL.

In addition, a free energy minimum between the encounter complex and the intermediates was characterized. We denote the state as hidden intermediate (state H) since it has never been uncovered before. Compared with the encounter complex, some native contacts initially formed in state H. The contacts formed in this state might be corresponding to the primary driven force for the binding. The contact analysis showed that the intermolecular contacts between the aromatic residue Tyr658 on KIX and the hydrophobic residues Leu128 and Ile137 on pKID formed in state H, the contact probabilities are 80.3% and 62.1%, respectively. These residues are important to the pKID-KIX binding, which was proved in the single-residue mutagenesis experiment, the Y658A mutation would completely abrogate the complex

formation¹⁶, the L128A and I137A mutation would increase the dissociation constant (K_d) over tenfold and two orders of magnitude, respectively²⁶.

Fully bound state and free energy barrier. There is a high energy barrier (state T) between the intermediate state and the final-binding state. The pKID further folding to helical structures in this state compared with state I1 and I2, especially for the α_B region, which is almost fully folded in state T. The native contacts on the α_B regions are almost fully formed, however, the contacts in the α_A regions are partially formed with relatively low probability (Supplementary Fig. 2). The results mean that α_A haven't bound to the right position (binding sites) of KIX in the state T. It should be noticed that the native contacts between the phosphorylated serine on pKID and the residues Tyr658 and Lys662 on KIX are formed in this state, with the contact probability 66.8% and 64.2%, respectively, which might be the driving force for α_A region binding to the right position on KIX and the fully folding of α_A .

In the fully bound state, the residues on pKID bound to the native-binding sites of KIX. The α_B region and the C-terminal of α_A region (residue 128–132) completely folded, the helicity on these regions are close to 100%. The helicity of N-terminal of α_A increase to 40%. The incomplete folding of N-terminus of α_A was also demonstrated by Dahal et al.²¹, they found that the final bound complex of pKID-KIX is partially mobile, with α_A loosely bound to the KIX based on the kinetic experiments.

The free energy landscape for KID-KIX binding. Unlike the pKID, the unphosphorylated KID is hard to form stable complex with KIX. The binding affinity of unphosphorylated KID to KIX is about 100 times lower than the pKID²¹. To describe the binding behaviors of KID and KIX, the FEL as a function of KID-KIX COM distances and the fraction of native contacts (Q) was given in Fig. 4. The Q values is corresponding to the native contact atoms in the structure of pKID-KIX complex. Compared with pKID, KID is hard to form stable complex with KIX since the high free energy in the high Q region. Interestingly, KID and KIX also form encounter complex and some intermediate states with very low native contacts. For the structures in the encounter complex state, the binding sites are variable and transient, which is similar to the state E in pKID binding to KIX. The contact map shows the interactions between KID and KIX in the intermediates were mainly happened in α_B region (Supplementary Fig. 5), in which the hydrophobic interactions might dominant the interactions. In the binding process of pKID and KIX, two intermediates with native contacts number close to 80 were observed.

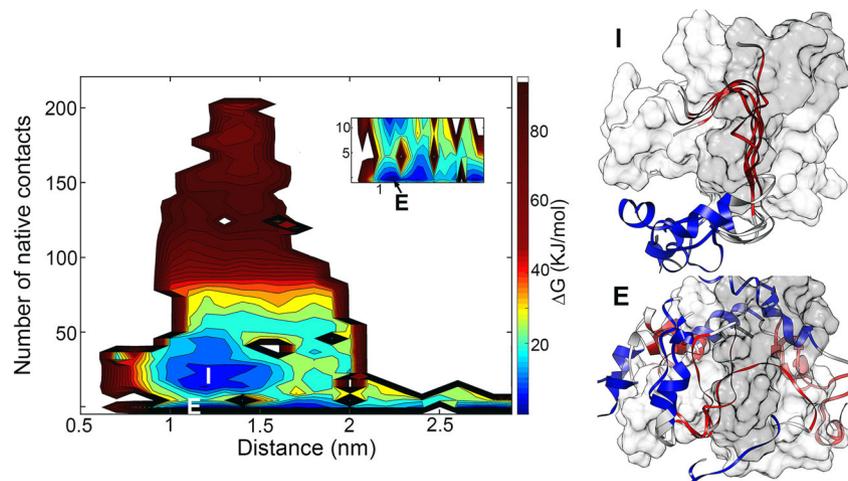


Fig. 4 Free energy landscape (FEL) of KID and KIX binding process. The FEL is the function of the KID-KIX distance and native contacts number, which formed in pKID-KIX. Representative structures of states E and I are shown. The structure of KID-KIX complex derived from pKID-KIX crystal structure and shown by surface in KID-KIX complex, KID color in dark-gray and KIX color in gray. The structures of KID in major states was shown in ribbon and α_A region color in blue, α_B region color in red. The coordinates in PDB format of the representative structures are given in Supplementary Data 2.

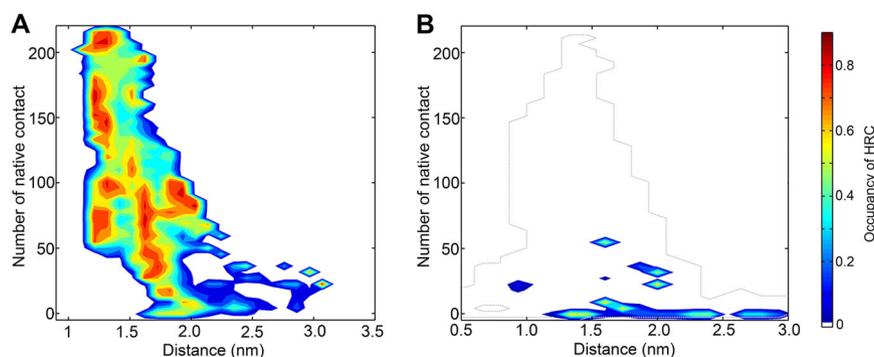


Fig. 5 The HRC formation propensity. **a** The formation probability of HRC as a function of the pKID-KIX distance and native contacts number which formed in pKID-KIX. **b** The formation probability of HRC as a function of the KID-KIX distance and native contacts number which formed in pKID-KIX. The formation of hydrophobic residue cluster (HRC) is defined by the contact atoms in residues L128, Y134, I137 and L138 larger than 15. Only the states with the free energy lower than 20 kJ/mol are given. The profile of free energy landscape of KID-KIX binding is given by dashed-line.

However, it is hard to form the similar intermediates in KID-KIX system. Besides, KID is basically incompletely folded and unstructured in all the states interact with KIX. Our results indicate that KID is transiently contacted with KIX and difficult to form stable and ordered complex.

The hydrophobic residue cluster formed in the binding process of pKID.

The mutagenesis and kinetic experiments demonstrated that the hydrophobic residues Leu128, Tyr134, Ile137, Leu138, and Leu141 are important in the pKID-KIX binding. Based on our simulations, we found the hydrophobic interactions related to these residues formed prior to the formation of pKID-KIX binding intermediates. On the other hand, the interactions between these residues are absent in the unphosphorylated KID and the binding and folding process of KID and KIX would not proceed after the encountering complex. The results demonstrate the interactions between the hydrophobic residues Leu128, Tyr134, Ile137, Leu138, and Leu141 play important roles in stabilizing and guiding pKID-KIX binding.

By analyzing the structure properties of conformations in the binding process, we found the special structure pattern, i.e., the HRC, also appeared in pKID. The HRC (the number of contact heavy atoms in the side-chain of Leu128, Tyr134, Ile137, and Leu138 are larger than 15) formation probability were projected

on the pKID-KIX and KID-KIX binding FEL (Fig. 5a and b), respectively. It can be seen that the average probability of HRC in the unphosphorylated KID is lower than 10%, indicate that the HRC structures are basically absent in the KID. On the other side, the conformations of pKID prefer to form the HRC, especially the conformations in the hidden state H. As the state H plays important role in the binding process of pKID and KIX, the formation of HRC and its anchoring to the binding site on KIX might provide the initial force of correct binding. We inferred that the HRC amplifies the hydrophobic interaction ability of pKID and facilitate the pKID to search for favorable binding sites on KIX and finally fold to the bound state.

Discussion

Phosphorylation promotes the formation of hydrophobic residue cluster. Phosphorylation on the Ser133 of KID dramatically increase the binding affinity of the peptide to KIX, however, the underlying mechanism is still unclear. The pSer133 is contact with the residues Lys662 and Tyr658 of KIX in the experimental complex structure, and the mutagenesis experiment shows that Y658A mutation on KIX completely abrogate the complex formation¹⁶. Therefore, the hydrogen bond interaction between the pSer133 and the Tyr658 might great contribute to the stabilization of the pKID-KIX complex. However, the kinetic

experiments show the binding Φ -values of residues around pSer133 are low or negative, indicates the phosphorylation has little effects in initiating the binding of pKID to KIX. Our simulation results are consistent with the experimental observations. Barely contacts between pSer133 and residues in KIX were formed at the early stage of the binding process, i.e., the free state, the encounter complex and the intermediates. The minimum distance between pSer133 and residues on KIX is larger than 10 Å in the encounter complex and intermediate states (Supplementary Fig. 6). The NMR experiment found a large chemical shift change on pSer133 caused by the encounter complex formation (the chemical shift change on backbone amide of pSer133 is 0.34 ppm). The chemical shifts of pSer133 also changed dramatically in our simulation, i.e., the chemical shift difference of backbone amide of pSer133 in the encounter complex and free state is about 0.2 ppm. The results indicate that even no obvious interactions formed between pSer133 and KIX, the local environment change when approaching KIX may induce the chemical shift perturbation on pSer133. However, it can't be excluded that the alternative pathways were observed in the NMR experiment which would induce larger chemical shift perturbation on pSer133.

On the other side, the phosphorylation on Ser133 promote the formation of HRC in pKID. Upon the phosphorylation, the phosphate group prefer to form hydrogen bonds with its nearby positively charged residues, such as Arg124, Arg125, Arg 130, Arg 131, and Lys136 in pKID. The phosphorylated group on pSer133 has high propensity to contact with the positively charged residues Arg131 (49.3%) and Lys136 (57.2%) in free pKID. In comparison, the contact probability of Ser133 with Arg131 and Lys136 in KID are 16.3% and 10.4%, respectively. (Supplementary Fig. 7). As the hydrophobic residues of the HRC are around pSer133 and the positively charged residues, the interactions would facilitate the approaching of these hydrophobic residues and the formation of HRC.

Hydrophobic residue cluster in the binding process. The HRC amplifies the hydrophobic interaction ability of KID and facilitates the searching of appropriate binding sites on KIX. In order to identify the interaction partner on KIX with the HRC, we calculated the interactions between the residues in HRC and residues on KIX in the hidden state. The results show that the HRC prefers to contact with the C-terminal helix (helix 3) of KIX; in particular, there are high interaction probabilities between residues Leu128 and Ile137 on pKID with Tyr658 on KIX (Fig. 6). The contact probability between L128 or I137 and Y658 are larger than 0.6 in the hidden state H, which is corresponding to the initial stage of the binding process. The results demonstrated

that the interactions between the HRC and residue Y658 play important role in the binding process. In fact, the important roles of Y658 in the binding process was observed by the mutagenesis experiments, Radhakrishnan et al. found that the single mutation Y658A on KIX would completely abolish the complex formation¹⁶.

Flexible conformation selection in the IDP binding. The HRC structures are observed in both of the free pKID and the pKID-KIX complex. More than 40% conformations with contact number larger than 15 in the free state pKID. The abundant amount of HRC structures of free pKID allow the direct binding of these conformations with KIX. In fact, the HRC structures of pKID in the hidden intermediate and free-state pKID are similar (Supplementary Fig. 8). As a result, we conclude that the binding and folding of pKID with KIX follows a flexible conformational selection mechanism, i.e., some pre-formed structure patterns on the free IDP is required for the effective binding (the HRC structure on pKID in the case of pKID-KIX binding). However, unlike the solid conformation selection model, which is common in the binding of enzyme and substrates, only some local structure features are required. There is no need to sample the pre-formed structure with the ordered final configuration. In the pKID-KIX binding process, the structures of regions except the HRC are distinct with each other (Supplementary Fig. 9). This binding model proposed here allows the binding of IDP with high efficiency.

The folding steps of α_A and α_B . One major debate about the pKID-KIX binding is the order of α_A and α_B folding, i.e., which part complete the folding firstly. The kinetic experiment shows that the binding and folding of α_B region prior to the binding and folding of α_A . Nevertheless, NMR results show that the chemical shifts differences of the α_B region in the intermediates and bound state are larger than the α_A region, indicate the α_B region are less folded than α_A in the intermediates.

The controversy can be explained and clarified by our simulations. By analyzing the residue helicity of intermediates (Supplementary Fig. 3), we found that the helix in the C-terminal region of α_A are basically folded in intermediate I1. The average helicity in this region (residue 124–128) is about 84%. The average helicity on the α_B region (residue 134–144) is near 40%, which is consistent with the NMR experiments. On the other side, both the α_A and α_B are basically finished the folding in the transition state. However, the α_A does not bind to the correct sites on KIX. The distributions of α_A - α_B angles in different states are given in Supplementary Fig. 9. The incorrect position of α_A might be the reason why the Φ -values of α_A region are much smaller

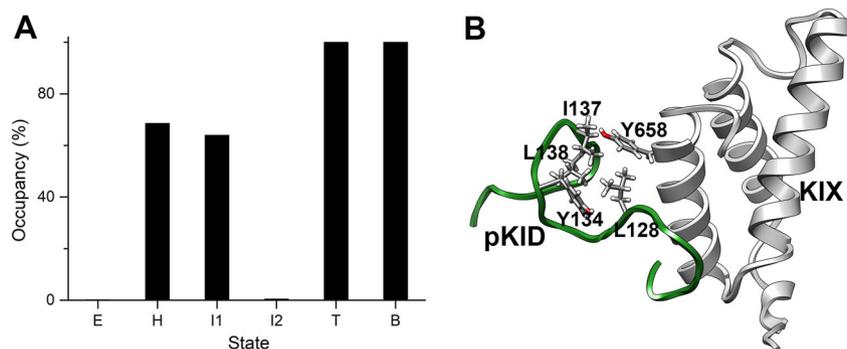


Fig. 6 The interactions between Y658 and HRC residues. **a** The contact probability between the hydrophobic residues in HRC with the Y658 (on KIX) in the encounter complex (state E), the hidden state (state H), intermediates (states I1 and I2), the transition state (state T) and the bound state (state B). **b** The representative structure of state H of pKID and KIX binding.

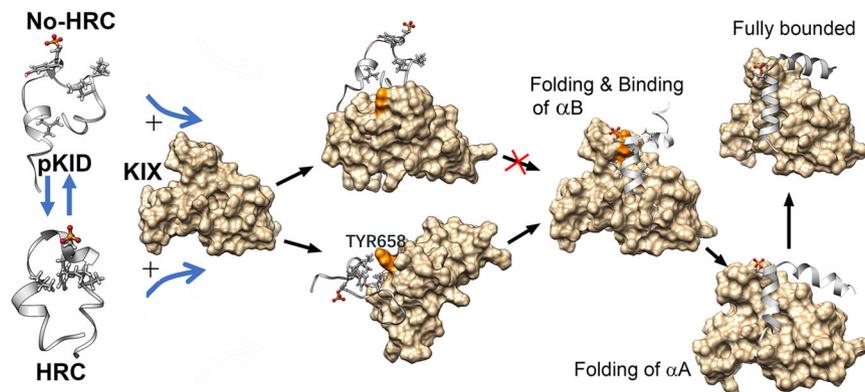


Fig. 7 The model of coupled folding and binding process of pKID and KIX. Before the encountering with KIX, the free pKID adopts flexible conformations and quick transiting between the conformations with HRC formed and the conformations without HRC. The pKID dynamically contact with KIX on multiple “hot-pot” binding sites to form the encounter complex, however, the pre-formed HRC structure would promote the correct binding of pKID to the Tyr658 on KIX. The accurate docking further promote the binding of α_B region. The binding with KIX also would induce the folding of C-terminal region of α_A and the α_B region of pKID. After that, the N-terminus of α_A would complete the folding. The folded α_A of pKID would rotate to the correct positions on KIX.

than the α_B region. Interestingly, the structures with incorrect α_A – α_B angles were also observed in a recent computational study, though they proposed these structures are in “misfolding” state²⁹.

The model of pKID-KIX binding process. A model is proposed to describe the binding and folding process of pKID-KIX based on the observations in this work (Fig. 7). In the free state, pKID adopts flexible conformations and quick transiting between these conformations. Due to the electrostatic interactions between pSer133 and positively charged residues on pKID, the hydrophobic residues around the positively charged residues incline to assemble together and form the HRC. Without the HRC structural pattern, pKID dynamic contact with KIX on multiple “hot-pot” binding sites to form the encounter complex, however, the pre-formed HRC structure would promote the pKID correct bind to the Tyr658 on KIX. The HRC is critical in the forming of intermediates. The folding and binding of pKID should obey the flexible conformational selection mechanism, which means the structure is high changeable except the HRC structure. The C-terminal region of α_A firstly complete the folding in the intermediates, meanwhile, the α_B region is more flexible and the N-terminal of α_A is unfolded. After the folding and binding of α_B region, the N-terminus of α_A complete folding. The folded α_A would rotate to the correct positions and finally bound with KIX by crossing the energy barrier.

Conclusions

The IDPs are usually binding with partners to carry out their biological functions. It is still ambiguous why IDP could adopt high specificity and efficient ability in the binding. In this study, the binding mechanism of pKID to KIX, which is regulated by phosphorylation on the serine in the central region of KID, was investigated by the computational simulations. The structure properties of free pKID and KID, as well as the binding process of pKID with KIX were characterized. The enhanced sampling results show that both free-state pKID and KID are disordered except with some transient helical structures on them. Similar to the experimental observation, no obvious differences were observed on the secondary structure composition of the pKID and KID. However, more hydrophobic interactions formed in the pKID, which promote the formation of the special HRC. Based on the binding free energy surface of pKID and KIX, the critical sites and important intermediates on the binding process were characterized. The HRC structures were also observed on the pKID in the present of KIX. Although both of KID and pKID would form

encounter complex with KIX, only the structures with HRC pre-formed in free pKID would bind to the correct binding sites on KIX and further fold to the final structure. The binding mechanism of the intrinsically disordered pKID follows a flexible conformational selection mechanism, i.e., the substrate protein specifically binding with the IDP with some special locally structures, meanwhile, most of the rest regions on the IDP are flexible and dynamic. The flexible conformational selection model proposed in this work give the explanation of the high specific and efficient of IDP binding. The binding mechanism proposed in this work provide new insights in the protein dynamic interactions and phosphorylation regulation.

Methods

System setup. The structures of free pKID and KID were built based on the experimental structure of pKID-KIX complex (PDB ID: 1KDX¹⁶). The phosphorylated pS133 was mutated back to the serine to build the structure of KID. The N- and C- terminus of pKID and KID were capped with acetyl (ACE) and amine (NH₂) groups, respectively. 4324 and 5546 water molecules were added to solvating pKID and KID, respectively. Sodium ions and chlorine ions were added to neutralize the systems and the final concentrations of the sodium chloride were set to be 100 mM. The amber99SB-ILDN force field³⁰ was employed for proteins, TIP3P model³¹ was used for water molecules. The force field parameters of phosphoserine were taken from Steinbrecher et al.’s work³². The unbiased molecular dynamics simulations were performed to equilibrate the conformations of free pKID and KID in aqueous solution. The systems were energy-minimized for 5000 steps using the steepest descent method. Then the NVT simulations with position constraints applied on protein heavy atoms and NPT simulation were performed to equilibrate the systems. All bonds length were constrained with LINCS algorithm³³. Isotropic scheme was utilized to couple the lateral and perpendicular pressures separately. The Particle-Mesh Ewald method was employed to calculate long-range electrostatics with a cutoff of 10 Å. The temperature was kept at 300 K with V-rescale method³⁴ and the pressure were coupled by Parrinello-Rahman barostat³⁵. 500 ns production runs were performed with 2-fs time step by GROMACS5.1.4^{36,37}.

The well-tempered metadynamics combined with parallel tempering. The parallel tempering simulations combined with well-tempered ensemble (PTMetaD-WTE) method^{38–40} was employed to obtain the structure ensemble of free pKID and KID. The initial structures of pKID and KID in the PTMetaD-WTE were obtained from the last snapshots of 500 ns unbiased molecular dynamics simulations. Twelve replicas were simulated spanning the temperature range of 288–508 K. The temperature distribution of these replicas was chosen according to the Ref. 40. The PTMetaD-WTE simulations were implemented in two-step scheme. First, the parallel tempered simulations in the well-tempered ensemble (PT-WTE) were employed to sample the conformations by using the potential energy as the collective variable. The bias factor γ was set to be 30. The height of initial bias energy is 1.0 kJ mol⁻¹ and the width is 300 kJ mol⁻¹. Exchange of configurations between adjacent replica was attempted every 150 fs. After the 30 ns simulations on each replica, the height of the bias energy decreased to the value close to zero and the exchange acceptance probability between adjacent replicas are about 0.3. The average potential energy in PT-WTE simulations remains close to the canonical

value with large fluctuations. In the second step, simulations in all replicas were performed with a static energy bias in the potential energy landscape which constructed in PT-WTE. The history-dependent energy bias is added on two collective variables to enhance the sampling of the structure on α_A and α_B regions of KID. The description and definition of the CVs are given in the supporting information. The PTMetaD-WTE simulation on each replica is 300 ns and the total simulation time is 3.6 μ s. All simulations were performed using GROMACS5.1.4 and PLUMED-2.1 plugin⁴¹. The average exchange acceptance ratio between the replicas is 0.35 for pKID and 0.39 for KID, respectively.

Bias-exchange metadynamics (BE-MetaD). BE-MetaD combined the ideas of replica exchange and metadynamics^{42,43}, the simulations are exchanged in different replicas, which could be present by different collective variables. In this way, a large number of different variables can be sampled simultaneously. In this work, BE-MetaD simulation performed in well-tempered ensemble³⁹. The initial structure was built based on the experimental complex structure (PDB ID: 1KDX). 10092 and 9746 water molecules were added to solvating pKID-KIX and KID-KIX complexes, respectively. Sodium and chloridion were added to neutralize the charged systems and the final concentrations of the sodium chloride were set to be 100 mM. Four biased replicas run along with four CVs for the BE-MetaD simulations, the descriptions and definitions of the CVs are given in supporting information. The exchanges between the replicas were attempted every 4 ps. 450 ns simulation was performed on each replica, and totally 1.8 μ s for each system. The convergence of these simulations is shown in Supplementary Fig. 10.

Analysis. Secondary structures were assigned using the STRIDE module in VMD^{44,45}. Chemical shifts were calculated by Shifftx2²⁸. Free energy landscapes as a function of CVs for pKID/KID-KIX system were calculated by weighted histogram analysis method by using the METAGUI program⁴⁶. The reweighted value of the observable property O were calculated based on the estimated free energies as following⁴³:

$$\langle O \rangle = \frac{\sum_{\alpha} O_{\alpha} \exp(-F_{\alpha}/k_B T)}{\sum_{\alpha} \exp(-F_{\alpha}/k_B T)}, \quad (1)$$

where F_{α} is the free energy of cluster α , k_B is the Boltzmann constant, T is the system temperature, the sums run over all the clusters of the free energy profile. O_{α} is the average value of O in cluster α .

Data availability

The data generated in this study are available from corresponding author upon request.

Received: 26 April 2020; Accepted: 11 August 2020;

Published online: 07 September 2020

References

- Xun, L. et al. Elucidating human phosphatase-substrate networks. *Sci. Signal.* **6**, rs10 (2013).
- Humphrey, S. J. et al. Protein phosphorylation: a major switch mechanism for metabolic regulation. *Trends Endocrinol. Metab.* **26**, 676–687 (2015).
- Sims, R. J. 3rd. & Reinberg, D. Is there a code embedded in proteins that is based on post-translational modifications? *Nat. Rev. Mol. Cell Biol.* **9**, 815–820 (2008).
- Iakoucheva, L. M. et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32**, 1037–1049 (2004).
- Bah, A. & Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *J. Biol. Chem.* **291**, 6696–6705 (2016).
- Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
- Theillet, F. X. et al. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem. Rev.* **114**, 6661–6714 (2014).
- Chen, J. Towards the physical basis of how intrinsic disorder mediates protein function. *Arch. Biochem. Biophys.* **524**, 123–131 (2012).
- Bah, A. et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106–109 (2015).
- Nishi, H. et al. Regulation of protein-protein binding by coupling between phosphorylation and intrinsic disorder: analysis of human protein complexes. *Mol. Biosyst.* **9**, 1620–1626 (2013).
- Shammas, S. L. et al. Insights into coupled folding and binding mechanisms from kinetic studies. *J. Biol. Chem.* **291**, 6689–6695 (2016).
- Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* **22**, 693–724 (2013).
- Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).
- Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2**, 599–609 (2001).
- Thakur, J. K. et al. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res.* **42**, 2112–2125 (2014).
- Radhakrishnan, I. et al. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752 (1997).
- Zor, T. et al. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277**, 42241–42248 (2002).
- Radhakrishnan, I. et al. Conformational preferences in the Ser(133)-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett.* **430**, 317–322 (1998).
- Matsuno, H. et al. Kinetic study of phosphorylation-dependent complex formation between the kinase-inducible domain (KID) of CREB and the KIX domain of CBP on a quartz crystal microbalance. *Chem.-A Eur. J.* **10**, 6172–6178 (2004).
- Shaywitz, A. J. et al. Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol. Cell Biol.* **20**, 9409–9422 (2000).
- Dahal, L. et al. Phosphorylation of the IDP KID modulates affinity for KIX by increasing the lifetime of the complex. *Biophys. J.* **113**, 2706–2712 (2017).
- Solt, I. et al. Phosphorylation-induced transient intrinsic structure in the kinase-inducible domain of CREB facilitates its recognition by the KIX domain of CBP. *Proteins-Struct. Funct. Bioinforma.* **64**, 749–757 (2006).
- Parker, D. et al. Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation. *Mol. Cell* **2**, 353–9 (1998).
- Ganguly, D. & Chen, J. H. Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding. *J. Am. Chem. Soc.* **131**, 5214–5223 (2009).
- Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**, 52–61 (2012).
- Dahal, L. et al. pKID binds to KIX via an unstructured transition state with nonnative interactions. *Biophys. J.* **113**, 2713–2722 (2017).
- Sugase, K. et al. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
- Han, B. et al. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
- Chong, S.-H. et al. Explicit characterization of the free energy landscape of pKID-KIX coupled folding and binding. *ACS Cent. Sci.* **5**, 1342–1351 (2019).
- Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins-Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
- Jorgensen, W. L. et al. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Steinbrecher, T. et al. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **8**, 4405–4412 (2012).
- Hess, B. et al. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
- Bussi, G. et al. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
- Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
- Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
- Páll, S. et al. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. in *Solving Software Challenges for Exascale*, Cham, 2015 (eds Markidis, S. & Laure, E.) 3–27 (Springer International Publishing, Cham, 2015).
- Deighan, M. et al. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theory Comput.* **8**, 2189–2192 (2012).
- Bonomi, M. & Parrinello, M. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **104**, 190601 (2010).
- Prakash, M. K. et al. Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations. *J. Chem. Theory Comput.* **7**, 2025–2027 (2011).
- Tribello, G. A. et al. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
- Piana, S. & Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **111**, 4553–4559 (2007).
- Marinelli, F. et al. A kinetic model of Trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **5**, e1000452 (2009).
- Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins-Struct. Funct. Genet.* **23**, 566–579 (1995).

45. Humphrey, W. et al. VMD: visual molecular dynamics. *J. Mol. Gr.* **14**, 33–38 (1996).
46. Giorgino, T. et al. METAGUI 3: a graphical user interface for choosing the collective variables in molecular dynamics simulations. *Computer Phys. Commun.* **217**, 204–209 (2017).

Acknowledgements

This work is supported by the National Natural Funding of Science (21773298).

Author contributions

D.M. conceived and supervised the project, L.N. performed the MD simulations and analyzed the data. G.Y. contributed to analysis. L.N., G.Y., N.S., and D.M. contributed to the scientific discussion. D.M., and L.N. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42004-020-00370-5>.

Correspondence and requests for materials should be addressed to M.D.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020