OXFORD

Briefings in Bioinformatics, 2024, **25(2)**, 1–9
https://doi.org/10.1093/bib/bbae026
Problem Solving Protocol

Enhancing protein dynamics analysis with hydrophilic polyethylene glycol cross-linkers

Min Sun †, Jing Chen†, Chang Zhao, Lihua Zhang, Maili Liu, Yukui Zhang, Qun Zhao and Zhou Gong

Corresponding authors. Qun Zhao, CAS Key Laboratory of Separation Science for Analytical Chemistry, National Chromatographic R. & A. Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, No. 457, Zhongshan Road, Dalian, Liaoning 116023, China. Tel.: +86-411-84379720; Fax: +86-411+84379720; E-mail: zhaoqun@dicp.ac.cn; Zhou Gong, State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, Innovation Academy for Precision Measurement Science and Technology Chinese Academy of Sciences, Division of NMR Spectroscopy in Biological Systems, West No. 30 Xiao Hong Shan, Wuhan, Hubei 430071, China. Tel.: +86-27-87197358; Fax: +86-27-87199101; E-mail: gongzhou@wipm.ac.cn

†Min Sun and Jing Chen contributed equally to this work.

Abstract

Cross-linkers play a critical role in capturing protein dynamics in chemical cross-linking mass spectrometry techniques. Various types of cross-linkers with different backbone features are widely used in the study of proteins. However, it is still not clear how the cross-linkers' backbone affect their own structure and their interactions with proteins. In this study, we systematically characterized and compared methylene backbone and polyethylene glycol (PEG) backbone cross-linkers in terms of capturing protein structure and dynamics. The results indicate the cross-linker with PEG backbone have a better ability to capture the inter-domain dynamics of calmodulin, adenylate kinase, maltodextrin binding protein and dual-specificity protein phosphatase. We further conducted quantum chemical calculations and all-atom molecular dynamics simulations to analyze thermodynamic and kinetic properties of PEG backbone and methylene backbone cross-linkers. Solution nuclear magnetic resonance was employed to validate the interaction interface between proteins and cross-linkers. Our findings suggest that the polarity distribution of PEG backbone enhances the accessibility of the cross-linker to the protein surface, facilitating the capture of sites located in dynamic regions. By comprehensively benchmarking with disuccinimidyl suberate (DSS)/bis-sulfosuccinimidyl-suberate(BS3), bis-succinimidyl-$(PEG)_2$ revealed superior advantages in protein dynamic conformation analysis *in vitro* and *in vivo*, enabling the capture of a greater number of cross-linking sites and better modeling of protein dynamics. Furthermore, our study provides valuable guidance for the development and application of PEG backbone cross-linkers.

Keywords: poly ethylene glycol (PEG); cross-linking; NMR; MD simulations; protein dynamics

INTRODUCTION

Protein structure and dynamics are fundamental in unraveling the functions and mechanisms of biological macromolecules. Several structural biology techniques, including X-ray crystallography [1], nuclear magnetic resonance (NMR) spectroscopy [2–4], cryo-electron microscopy (cryo-EM) [5–8] and small-angle X-ray scattering (SAXS) [9, 10], have been widely employed to study protein structure and dynamics [11, 12].

In recent years, chemical cross-linking mass spectrometry (XL-MS) has emerged as a potent tool for investigating the protein structure and interactions, particularly the protein structure in the cellular microenvironment [13–17]. XL-MS utilizes cross-linkers that react and link specific residues within a defined range, allowing for the identification of cross-linked residues through high-resolution mass spectrometry. This, in turn, furnishes valuable distance constraints within the three-dimensional structure, thereby enhancing our comprehension

Min Sun is a doctoral candidate at the Innovation Academy for Precision Measurement Science and Technology, CAS. The current research direction involves the dynamic structure of biomolecules by integrating NMR technology with MD simulations.
Jing Chen is a doctoral candidate jointly trained by the University of Science and Technology of China and the Dalian Institute of Chemical Physics, CAS. The main research direction is new methods for *in situ* analysis of protein dynamics and interactions based on chemical cross-linking mass spectrometry.
Chang Zhao completed her master's degree at the Innovation Academy for Precision Measurement Science and Technology, CAS. She is currently a doctoral candidate at the Hong Kong Polytechnic University. The main research direction is to evaluate protein dynamics and protein–protein interactions with HDX and MD simulations.
Lihua Zhang is the group leader of High Separation and Characterization of Biomolecules at the Dalian Institute of Chemical Physics, CAS. Her research focuses on developing new technologies for proteome analysis, providing technical support for research in life sciences, precision medicine, synthetic biology and other fields.
Maili Liu is a member of Chinese Academy of Sciences and a Professor at the Innovation Academy for Precision Measurement Science and Technology, CAS. His research focuses on the development and application of novel methods in biological magnetic resonance and analytical chemistry.
Yukui Zhang is a member of Chinese Academy of Sciences and a professor of Dalian Institute of Chemical Physics, Chinese Academy of Sciences. Prof. Zhang's research is focused on the fundamental, new techniques and methods for separation science. He developed systematic thermodynamic and kinetic research methods for chromatography and made great contribution to the establishment of expert systems for liquid chromatography.
Qun Zhao is a professor at the Dalian Institute of Chemical Physics, CAS. She is interested in developing new technologies for qualitative and quantitative analysis of proteomes and protein–protein interactions.
Zhou Gong is an associate professor at the Innovation Academy for Precision Measurement Science and Technology, CAS. His research area primarily involves the combination of various experimental and computational methods to study the structure, dynamics and interactions of biomacromolecules.
Received: October 30, 2023. **Revised:** December 30, 2023. **Accepted:** January 11, 2024

of protein structure [18, 19]. Moreover, to enhance the capacity of cross-linkers for capturing dynamic protein conformations, another approach involves improving their reactivity and biocompatibility within physiological cellular environments. This can be accomplished by developing cross-linkers with hydrophilic groups to enhance their solubility under biologically orthogonal conditions [13]. Traditional cross-linkers like DSS possess a hydrophobic methylene backbone, which limits the solubility with the increased chain length. To address this limitation, alternative cross-linkers have been developed, such as bis-sulfosuccinimidyl-suberate (BS^3), which is an ionizable succinimidyl ester sulfonic acid–based compound that improves water solubility (Figure S1 available online at http://bib.oxfordjournals.org/) [20–22]. However, it's worth noting that BS^3 might have restricted membrane permeability. The other strategy is to modify the structural composition of the cross-linker arms. Polyethylene glycol (PEG), a versatile polymer renowned for its exceptional water solubility, has wide application in the delivery of therapeutic biomolecules due to its hydrophilic and non-toxic nature [23]. Additionally, by utilizing the excellent water solubility of PEG, longer PEG chains can be designed to enhance the capture of sites with longer spatial distances.

Therefore, considering the remarkable versatility of PEG backbone cross-linkers in biological systems, an intriguing question arises regarding the capture capability of methylene backbone cross-linkers compared to PEG backbone cross-linkers for protein structure and dynamics. However, to date, there hasn't been a systematic comparative study between hydrophobic and hydrophilic backbone cross-linkers. Hence, in this study, we conducted a comparison study of the performance between BS^3/DSS and bis-succinimidyl-$(PEG)_2$ [referred to as $BS(PEG)_2$, with two PEG units] cross-linkers in terms of protein structure and dynamic characterization (Figure S1 available online at http://bib.oxfordjournals.org/). We selected four multi-domain proteins: calmodulin (CaM) [24], adenylate kinase (AdK) [25], maltodextrin binding protein (MBP) [26] and dual-specificity protein phosphatase (PTEN) [27] as the subjects. Inter-domain dynamics of these proteins were investigated using these two different cross-linkers. Furthermore, the thermodynamic and kinetic properties of $BS(PEG)_2$ and BS^3 were explored through quantum chemical calculations and all-atom molecular dynamics (MD) simulations. Additionally, solution NMR was employed to evaluate the interactions between protein and the cross-linkers. We comprehensively revealed the advantages of $BS(PEG)_2$ in protein dynamic analysis by benchmarking its physical and chemical characteristics, *in vitro* and *in vivo* cross-linking capabilities and the mechanism underlying its stronger capture ability for dynamic regions against BS^3/DSS.

RESULTS

PEG cross-linker improves the capture of protein dynamics *in vitro*

To evaluate the protein dynamic characterization of methylene backbone and PEG backbone cross-linkers, we focused on comparing the differences between BS^3/DSS and $BS(PEG)_2$, taking into account their similar arm lengths but distinct backbone structures (Figure S1 available online at http://bib.oxfordjournals.org/). Firstly, we investigated the cross-linking performance of these cross-linkers on three proteins: CaM, AdK and MBP, which exhibit typical inter-domain dynamics. These proteins were cross-linked in their *apo* form *in vitro*, without binding to any substrate ligands or target proteins. To better compare the cross-linking capabilities of different backbones, we chose the more soluble BS^3 over DSS for the *in vitro* cross-linking experiments.

For the identified number of cross-linked sites, whether the analysis is conducted using only target proteins or by incorporating a yeast database as a trap library for cross-linking data retrieval, $BS(PEG)_2$ exhibited capturing more cross-linked pairs compared to BS^3 for all three proteins with high spectra quality (Figure 1, Table S1, Figure S2, Supplementary File—Instances of MS Spectra available online at http://bib.oxfordjournals.org/). Specifically, in the case of CaM, it adopts a dumbbell-like extended conformation in the absence of target protein binding (*apo* state). The residues between N and C domain (residues 77–81) exhibit flexibility, allowing CaM to adopt a more closed conformation in *holo* state upon binding to interacted proteins (Figure S3A available online at http://bib.oxfordjournals.org/). $BS(PEG)_2$ captured 36 cross-linked pairs, 4 more than BS^3. Notably, three of the four additional cross-linked pairs were between the N and C domains, suggesting PEG captured more cross-linking information representing the inter-domain dynamics of CaM (Figure 1A).

For AdK, the crystal structure of the *apo* form revealed an open conformation, with the ATP and AMP domains separated (Figure S3B available online at http://bib.oxfordjournals.org/), while the ATP and AMP domains will move toward the central core domain upon binding to the substrate AP_5A (Figure S3B available online at http://bib.oxfordjournals.org/). Here, $BS(PEG)_2$ captured 129 cross-linking pairs compared to 122 using BS^3. Notably, $BS(PEG)_2$ captured an additional 15 unique inter-domain cross-linking pairs involving the ATP-core, AMP-core and ATP-AMP residues, indicating its ability to investigate the different dynamic states of the three domains in *apo* AdK (Figure 1B). Importantly, these findings are consistent with previous findings, which suggest that the three domains in AdK undergo collective motions even in the absence of ligand binding [25]. In the case of MBP, the protein also consists of two domains (N domain and C domain) connected by three flexible linker residues. Crystal structures suggest that MBP undergo ~35° domain bending upon sugar binding (Figure S3C available online at http://bib.oxfordjournals.org/). Our study identified a total of 176 residue pairs cross-linked by $BS(PEG)_2$. Among these, 72 inter-domain cross-linking pairs (including 27 unique pairs) provided evidence of the bending motion between the two domains in *apo* MBP (Figure 1C), which is consistent with the previous NMR data [26].

Taken together, the above results demonstrate that XL-MS can capture inter-domain dynamics in solution, even in the absence of ligand or target protein binding. These dynamic characteristics are consistent with experimental data in solution. Compared to BS^3, $BS(PEG)_2$ was able to capture more cross-linking sites, particularly inter-domain sites, highlighting its suitability for investigating protein dynamics. These additional cross-linking sites provide valuable information for further modeling the dynamic structure of the protein.

Intrinsic features of PEG cross-linker make it easier to approach protein surface

To further explore the intrinsically properties of different cross-linkers with different backbone, we performed all-atom MD simulations for $BS(PEG)_2$ as well as DSS/BS^3. In the following computational and simulation studies, we selected DSS as comparative agent with the same NHS ester group as $BS(PEG)_2$. By calculating the distances between the oxygen atoms (O_α–$O_{\alpha'}$) at the ends of the cross-linker from the simulation trajectories, we can statistically obtain the conformation distribution of the two cross-linkers. The results demonstrate that both DSS and $BS(PEG)_2$

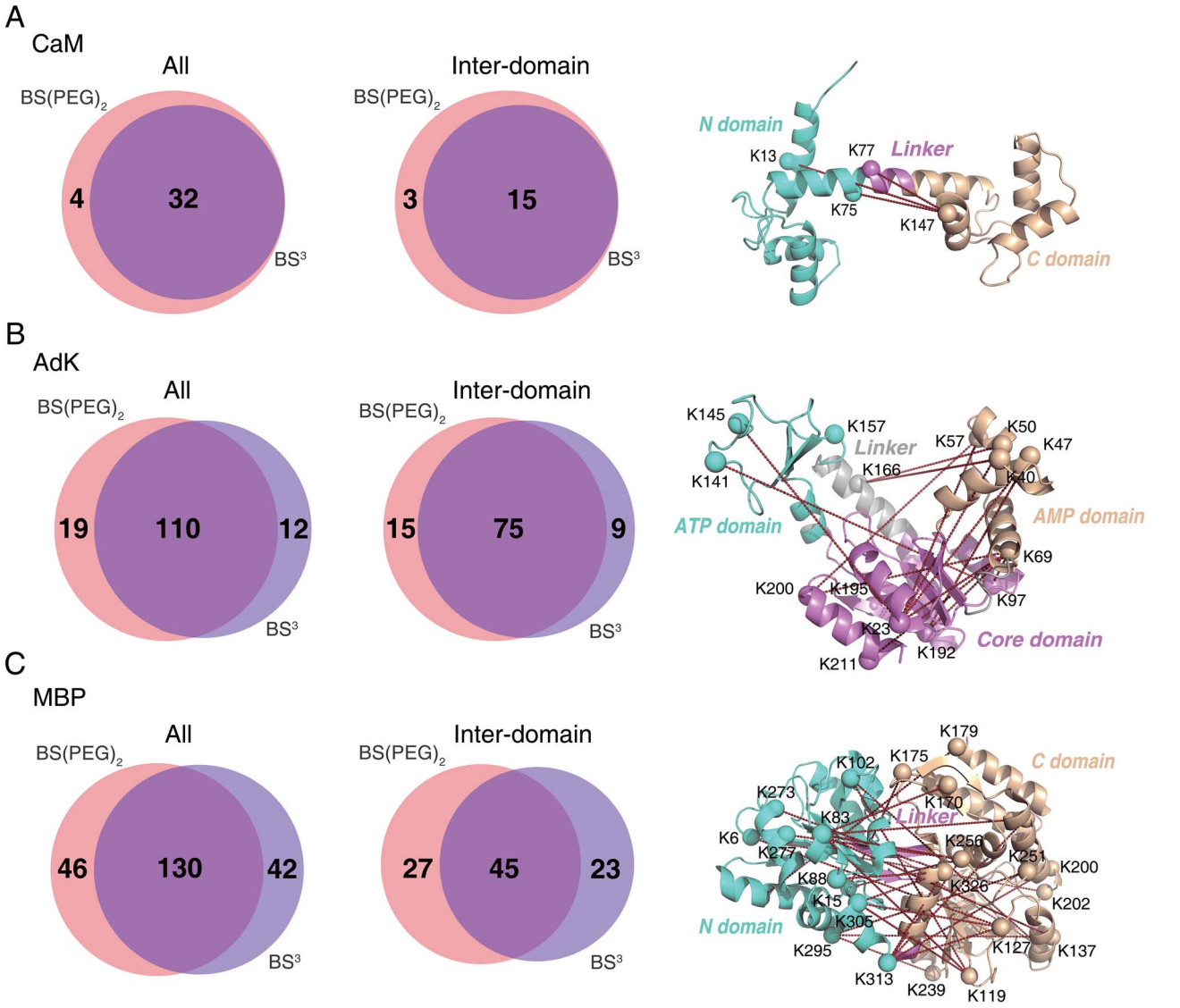

Figure 1. XL-MS analysis was conducted on three proteins *in vitro* respectively (**A**: CaM, **B**: AdK, **C**: MBP). The total number and the inter-domain cross-linked pairs were indicated with the Venn diagrams. The cross-linked pairs captured by $BS(PEG)_2$ and BS^3 were indicated, respectively. The structures of these proteins in the *apo* state are shown in the right panel, with different colors indicating the different structural domains and the linker between the structural domains. The inter-domain cross-linked sites are indicated by spheres, and the dotted lines between the spheres indicate the cross-links formed between them.

exhibit very low probability of being in a fully extended state, indicating that the fully extended state of the cross-linker is an energetically unstable state (Figure 2A and B). Both cross-linkers exhibit two different conformation states, with the junction of the two states around 7–8 Å. Therefore, we treat distances less than 7 Å as the closed state and distances greater than 8 Å as open states. The DSS spends 23.9% of the time in the open states (Figure 2A). In comparison, $BS(PEG)_2$ exhibited a higher proportion of closed states (33.9%) and relatively lower proportion of open states (53.2%) (Figure 2B) than DSS. The represent structures from MD trajectory indicate the closed states correspond to a folded-back compact conformation, while the open states correspond to a more extended conformation (Figure 2C).

Moreover, the kinetic parameters delineated the distinct free energy landscapes of the two cross-linkers (Figure 2D). The results showed that the dwell time of the open state was similar for both DSS and $BS(PEG)_2$, while the closed state of $BS(PEG)_2$ had a significantly longer dwell time (Figure S4 available online at http://bib.oxfordjournals.org/). This is consistent with the previously mentioned the open state of DSS is similar to that of $BS(PEG)_2$, while the $BS(PEG)_2$ has a greater proportion of closed states, indicating a lower energy of the closed states. Additionally, the exchange rate between the open and closed state was faster for $BS(PEG)_2$ ($1635 \pm 26/\mu s$) compared to DSS ($1253 \pm 67/\mu s$). Therefore, the energy barrier between the open and closed states is lower for $BS(PEG)_2$ than for DSS, facilitating easier transition between the two states and resulting in a faster exchange rate (Figure 2D). On the other hand, we utilized quantum chemical calculation to calculate the partial charges of the cross-linker backbone atoms. Statistical analysis revealed significant differences in partial charges of different atoms, despite the overall electrically neutral state of the cross-linker. The DSS backbone tended to exhibit a more neutral distribution (Figure 2E), whereas $BS(PEG)_2$ displayed greater polarity, with most atoms carrying higher charges (Figure 2F).

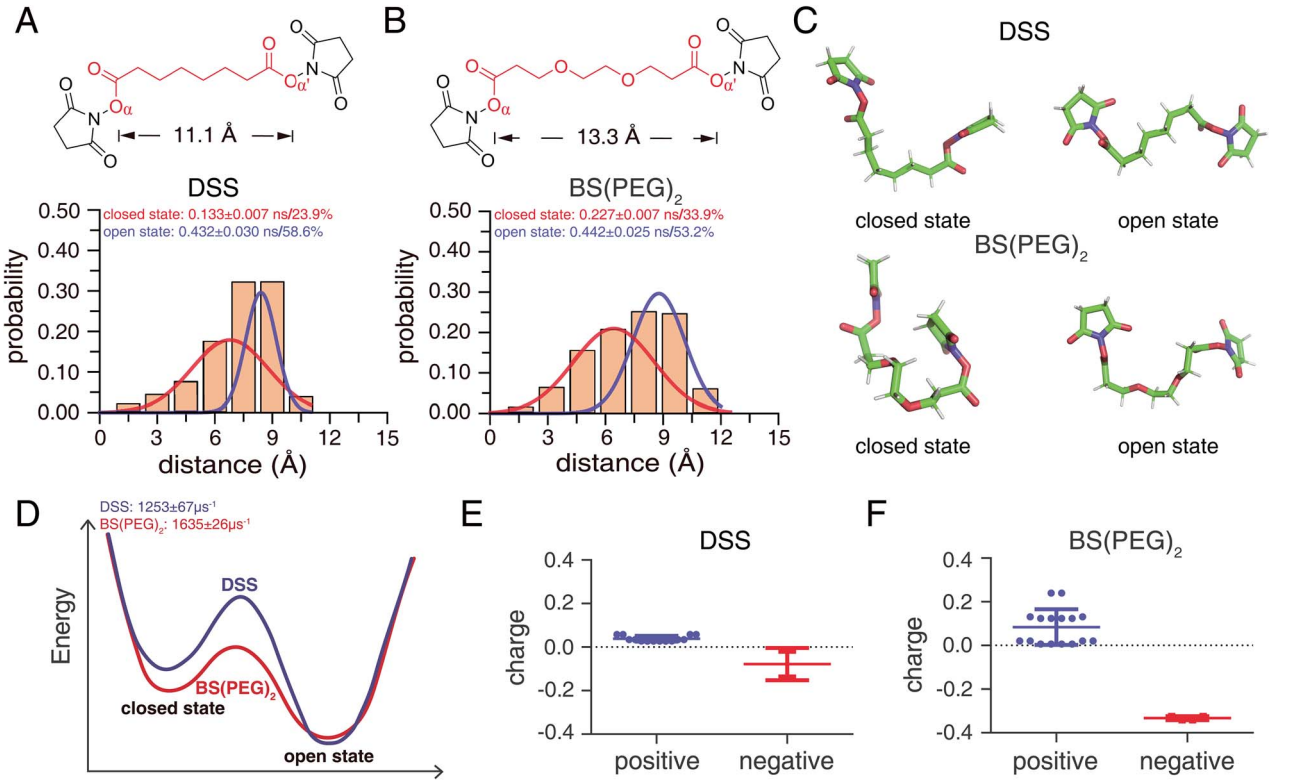

Figure 2. Dynamic and kinetic analysis of DSS and $BS(PEG)_2$. (**A**, **B**) The distance between O_α and $O_{\alpha'}$ was used to evaluate the conformation distributions of the DSS and $BS(PEG)_2$, two-dimensional structures of the two cross-linkers with the theoretical maximum distance between these two atoms (when the cross-linker is fully extended) were indicated. The statistics of the distance distribution between two atoms can be fitted with two Gaussian functions, collectively representing two states. The proportion and dwell time of closed and open states were also indicated. (**C**) Representative conformers from MD trajectories of cross-linkers in closed and open states. (**D**) Schematic free energy landscape for two cross-linkers. Exchange rates for the two cross-linkers were indicated. (**E**, **F**) Charge distribution of the cross-linkers backbone computed with Gaussian.

As a result, the more polar $BS(PEG)_2$ had a propensity for adopting the closed state, and the enhanced intramolecular electrostatic interactions facilitated conformational dynamics, reducing the energy barrier and resulting in a faster exchange rate between different conformational states. The lower energy barriers indicate that the PEG backbone exchanges between different conformational states more frequently, thereby increasing the likelihood of capturing sites at different distances. Meanwhile, more polar $BS(PEG)_2$ is more likely to approach the polar protein surface.

The cross-linking reaction typically involves one end of the cross-linker initially forming a covalent connection with the protein, and a greater number of closed states suggests that the $BS(PEG)_2$, after forming a covalent connection at one end, is more likely to approach the protein surface with the other end. To explore this, we patched one end of DSS and $BS(PEG)_2$ to the lysine side chain (K41 and K137) of the MBP protein, while allowing the remaining parts to move freely (Figure S5 available online at http://bib.oxfordjournals.org/). All-atom MD simulations were performed with the MBP protein connected to the cross-linker to examine the conformational changes of the cross-linker upon binding to the protein. To visually depict the movement of the cross-linker, the terminal carbon atoms of the cross-linkers (represented as a pink sphere in Figure S5 available online at http://bib.oxfordjournals.org/) in all snapshots from the simulation trajectories were rendered as spheres. The results revealed that regardless of the attachment sites on the protein, the structural distributions of $BS(PEG)_2$ were much closer to the protein surface and more concentrated (Figure 3A and D).

Meanwhile, we also calculated the distance distribution between the ends of different cross-linkers during the simulation process. Similar to the conformational distribution in its free form described earlier, when one end of $BS(PEG)_2$ is linked to the protein, the conformation still exhibits a bimodal distribution, appearing a larger proportion of the closed state (Figure 3B and E). In comparison, the structure of the DSS predominantly exhibits extended conformational states (Figure 3C and F). Additionally, we calculated the distances between the terminal carbon atom (Cz) of the cross-linker and the terminal Nz atom of lysine side chains. The results indicated that when one end was attached to the protein, the other terminal of $BS(PEG)_2$ connected to the lysine side chain was significantly closer to the protein surface (Figure 3G–L). The results indicate that when one end of the $BS(PEG)_2$ is connected to the protein, the other end has a higher probability of approaching the side chains of other lysine residues, thereby forming a cross-link. This finding further explains why $BS(PEG)_2$ can capture more cross-linking information.

Furthermore, we conducted NMR experiments to confirm the enhanced propensity of the $BS(PEG)_2$ backbone to approach the protein surface. We employed two compounds with backbones similar to BS^3/DSS and $BS(PEG)_2$, namely, *N*-[ε-maleimidocaproyloxy] succinimide ester (EMCS) and *N*-hydroxylsuccinimide (NHS)-polyethylene glycol-maleimide (NHS-PEG-MAL) (Figure 4A). These cross-linker analogs featured a maleimide group at one end and an NHS ester at the other end. Initially, we reacted an excess of Tris with the small molecules to convert the NHS ester into an inactive amine. After that, these small molecules were attached to the D41C site of the MBP

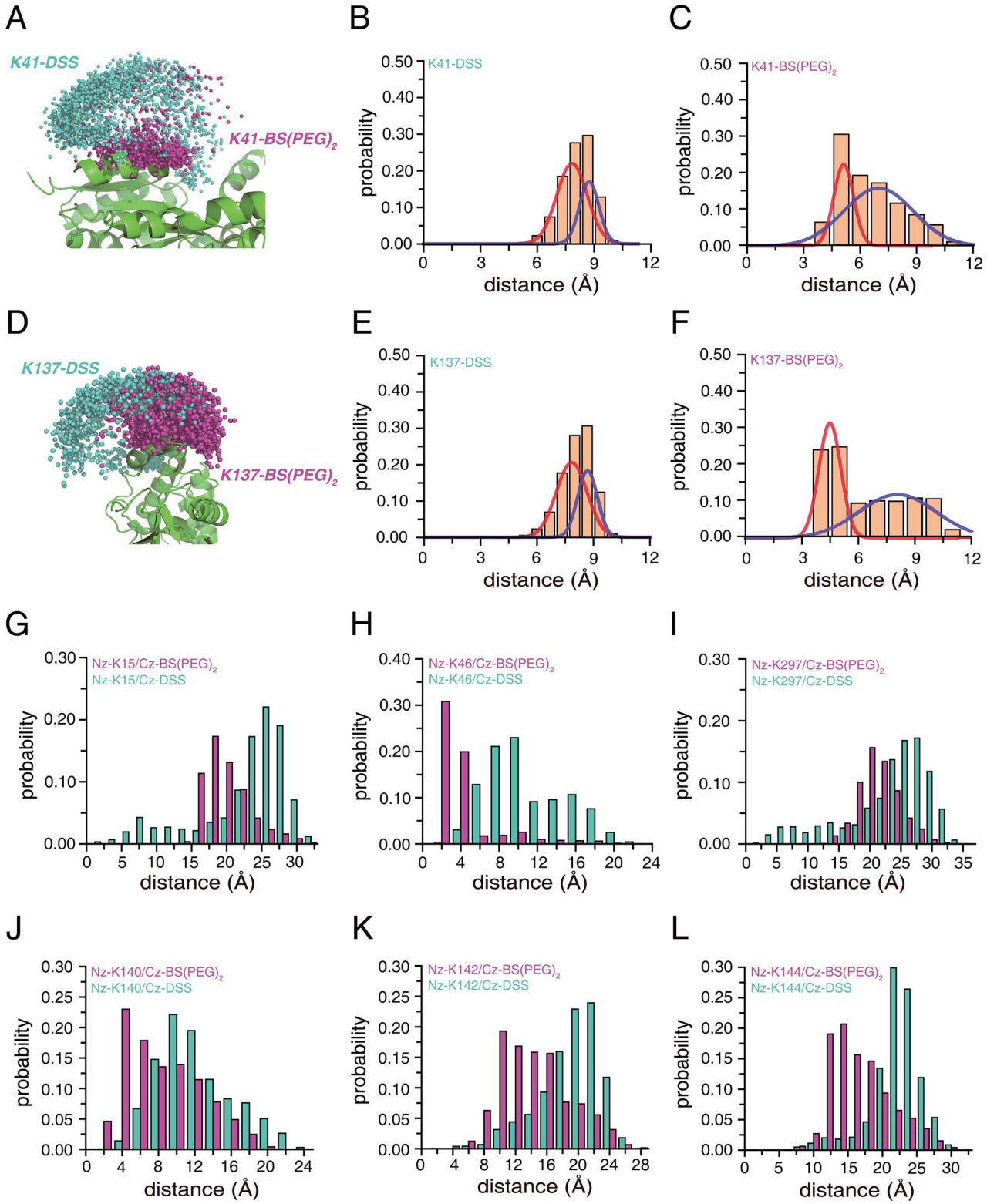

Figure 3. MD simulation results for MBP patched with two cross-linkers. (**A**, **D**) Depict the distribution of terminal atom of DSS and $BS(PEG)_2$ attached to K41 and K137 of MBP, represented by spheres for each cross-linker. (**B**, **C**, **E**, **F**) The distance distribution of patched DSS and $BS(PEG)_2$, which calculated with the Nz atom of the patched sites (K41 and K137) to the terminal carbon atom of the cross-linkers. (**G**–**L**) The distance distribution between the neighboring Nz atom of lysine side chain and the terminal carbon atom of cross-linkers.

protein (Figure S6 available online at http://bib.oxfordjournals.org/). The previous NMR experiments had already confirmed the D41C mutation did not affect the structure of the MBP protein [26]. The ^{1}H-^{15}N HSQC spectra of MBP protein with or without the attached cross-linker analogs were recorded. The chemical shift perturbation (CSP) was employed to evaluate the interactions between the cross-linker analogs and the amino acids on the protein.

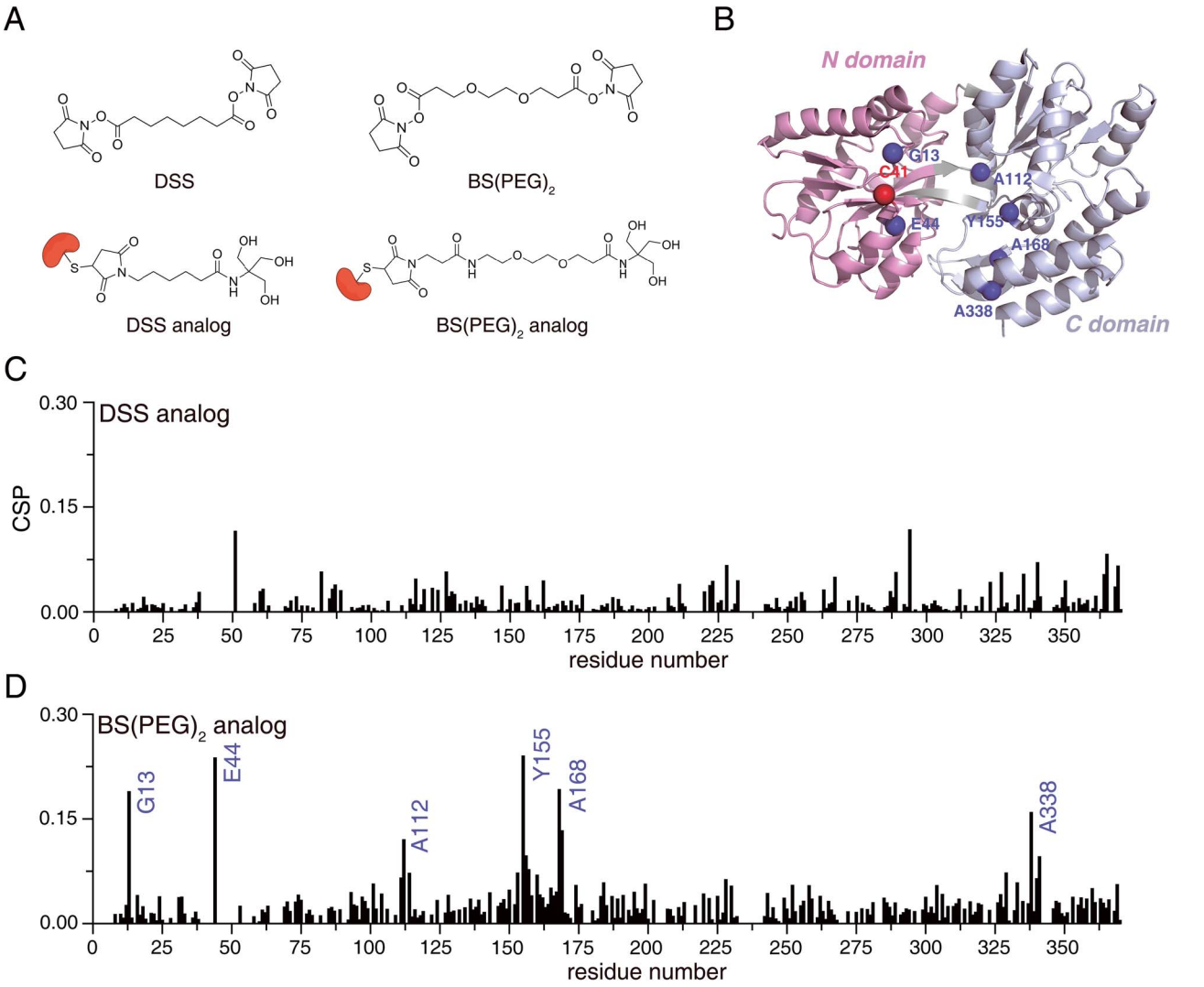

Figure 4. NMR experiments to investigate the interaction of cross-linker analogs with MBP. (**A**) The chemical structures of DSS and $BS(PEG)_2$ with their analogs used in NMR experiments, respectively. (**B**) Structure of MBP protein with different domains represented by distinct colors. Residue (D41C) used for attaching cross-linker analogs are indicated. Residues that exhibited significant chemical shift perturbations (CSPs) after binding with $BS(PEG)_2$ analogs are shown as spheres. (**C**, **D**) The CSPs of different cross-linker analogs. CSPs were calculated by subtracting the chemical shift of the MBP without analog from the chemical shift of the MBP with cross-linker analog connected.

The results indicate that there was almost no CSP observed before and after connecting the MBP with the DSS analog, with most of CSP values below 0.05 ppm (Figure 4C and Figure S7A available online at http://bib.oxfordjournals.org/). In contrast, when connecting the $BS(PEG)_2$ analog to MBP, more significant CSPs were observed, with specific amino acids showing chemical shifts exceeding 0.15 ppm (Figure 4D and Figure S7B available online at http://bib.oxfordjournals.org/). These notable CSPs were not limited to the vicinity of the connection site, such as G13 and E44, but were also observed in amino acids located at a greater distance, especially within the C domain of the protein, including A112, Y155, A168 and A338 (Figure 4B). These perturbations in the amino acid chemical shifts indicate that $BS(PEG)_2$ analog approached these amino acids. These results further support the notion that compared to BS^3/DSS, $BS(PEG)_2$ has a greater tendency to approach the protein surface, facilitating the capture of dynamic changes in protein conformation.

Taken together, the lower energy barriers enable $BS(PEG)_2$ to switch more easily between different conformational states and to form relatively closed state, both in its free form and when one end is connected to the protein. Additionally, the stronger polarity facilitates $BS(PEG)_2$ to move closer to the protein surface. These factors collectively enable $BS(PEG)_2$ to effectively capture a greater amount of cross-linking information, particularly within the structural domains responsive to protein dynamics.

Advantages of PEG cross-linker in studying protein dynamics *in vivo*

In addition to *in vitro* and *in silico* experiments, we conducted *in vivo* XL-MS analysis to compare the effectiveness of different backbone cross-linkers (Figure 5A). Due to the inability of BS^3 to penetrate the cell membrane, we utilized DSS and $BS(PEG)_2$, both of which are membrane-permeable, for *in vivo* XL-MS experiments. These two cross-linkers were individually employed to examine the structure and dynamics of PTEN within cells. PTEN comprises two domains: the phosphatase domain and the C2 domain. The crystal structure of PTEN shows that these two domains are in close proximity to each other, forming an interaction interface.

Consistent with the *in vitro* experimental results, $BS(PEG)_2$ demonstrated a higher efficiency in capturing cross-linked sites

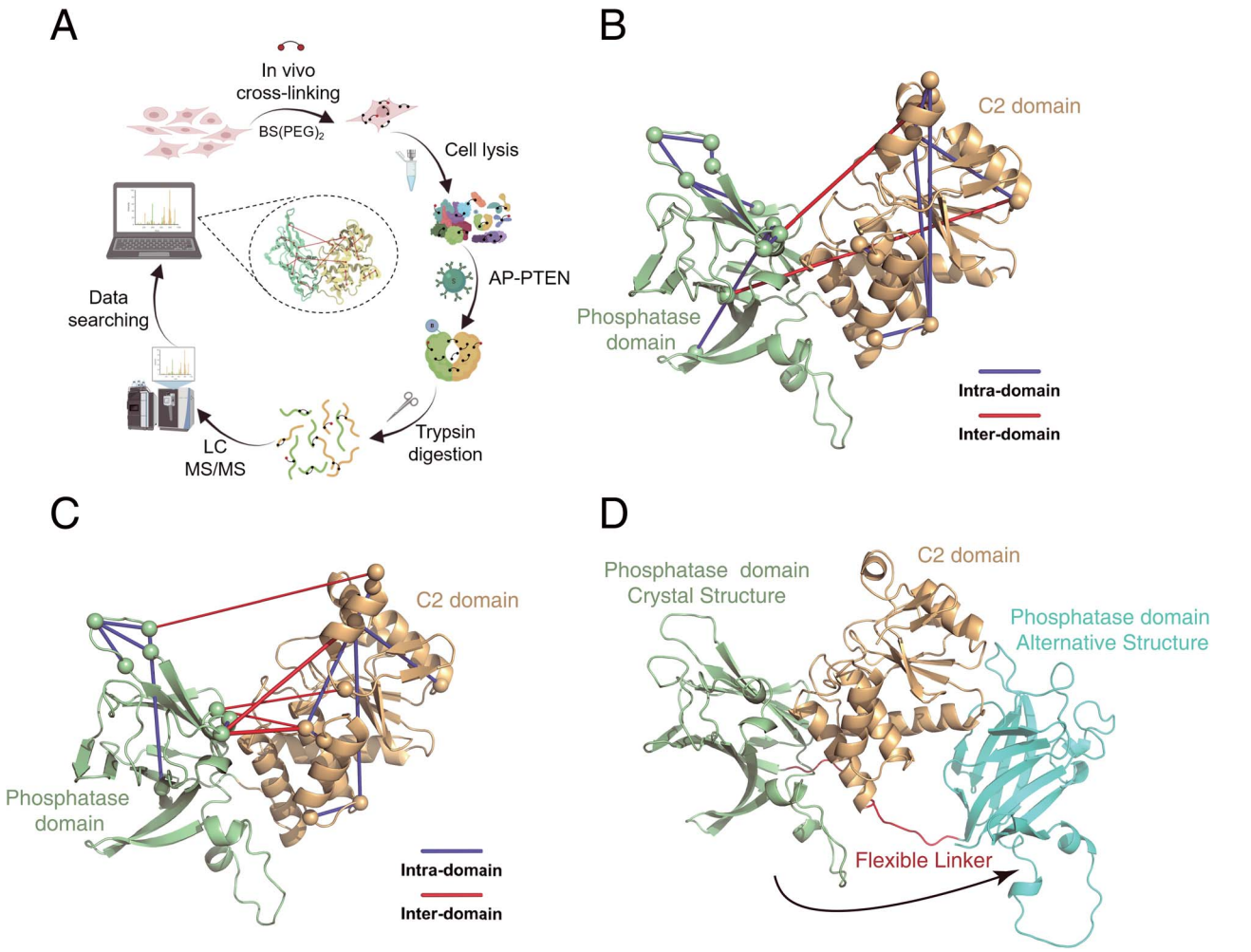

Figure 5. Structure and dynamics of PTEN protein in cells analyzed by two cross-linkers *in vivo*. (**A**) Schematic diagram of the flow of *in vivo* XL-MS experiments. (**B**, **C**) Crystal structure of PTEN with different domains depicted in distinct colors. Intra- and inter-domain cross-link pairs captured by DSS and $BS(PEG)_2$ were highlighted using straight lines. (**D**) Alternative structure of PTEN refined with *in vivo* cross-linking information. The C2 domain (blue) was fixed during structure refinement, the phosphatase domain can move as a rigid body and the flexible linker (red) between two domains has complete freedom. The positions of the phosphatase domain in crystal structure and alternative structure calculated with XL-MS data were indicated, respectively.

compared to DSS, regardless of whether only target proteins were used or a human database was incorporated for cross-link retrieval with high spectra quality (Table S2, Supplementary File—Instances of MS Spectra available online at http://bib.oxfordjournals.org/). Specifically, in the case of PTEN-targeted database searching, 16 cross-linked pairs were identified using DSS, which included 2 inter-domain cross-links. On the other hand, $BS(PEG)_2$ identified 17 cross-linked pairs, with 4 inter-domain cross-links (Figure 5B and C). To gain a more precise understanding of various conformational states of PTEN *in vivo*, we employed a higher concentration of 2 mM for $BS(PEG)_2$. It is important to note that the maximum concentration of DSS is typically limited to 0.25 mM due to its hydrophobicity. In contrast, $BS(PEG)_2$ allows for significantly higher experimental concentrations, leading to a greater coverage of cross-links. At this increased concentration, a total of 35 cross-linked pairs were identified, including 8 pairs of inter-domain cross-linked sites. Upon comparing the crystal structures of PTEN, we discovered that there are four pairs of inter-domain cross-linked sites where the distances in the crystal structure exceed the maximum arm length of the cross-linker. This observation suggests that PTEN may adopt additional alternative conformations in the cellular environment.

The two domains of PTEN are connected by a flexible linker region consisting of nine amino acids (Figure 5D). This means that the relative positions of the two domains can undergo dynamic changes. Using the cross-linking information as distance restraints, we performed structural refinement to obtain the conformational states of PTEN *in vivo*. During this refinement process, the C2 domain and phosphatase domain were treated as rigid bodies, respectively. The results revealed a structural ensemble consisting of two conformations that satisfy all the cross-linking restraints. One conformation is similar to the crystal structure, while the other one showed the phosphatase domain moving to the opposite side of the C2 domain (Figure 5D). Given that the C2 domain is responsible for binding to other proteins or lipids, once anchored in place, the dynamic nature of the phosphatase domain allows it to exert its dephosphorylation function in different directions. Hence, the multiple conformations of the PTEN protein confer it with versatile and flexible functionalities.

Overall, due to its ability to achieve higher cross-linking concentrations, $BS(PEG)_2$ can effectively capture useful information that reflects conformational dynamics of proteins within cells. Utilizing inter-domain cross-linking data, we characterized additional conformations of the PTEN protein *in vivo*, distinct from

the crystal structure observed outside the cellular environment. The dynamic structural analysis of intracellular proteins through *in vivo* XL-MS can also contribute to further research on protein functionality.

DISCUSSION

In this study, we systematically compared BS^3/DSS and $BS(PEG)_2$ cross-linkers to assess their effects on protein structure and dynamics. BS^3 and DSS share similar backbones and reactivity, but the sulfonic acid group of BS^3 improves solubility, enabling higher reaction concentrations. For *in vitro* experiments, BS^3 was used for its solubility, ensuring a fair comparison with PEG-based cross-linkers. For MD simulations and charge calculations, DSS was selected for consistent reactive groups. We discovered that the $BS(PEG)_2$ cross-linker, with its C-O backbone, exhibited stronger charge polarity compared to the alkyl backbone of BS^3/DSS. This difference allowed $BS(PEG)_2$ to adopt various conformations when interacting with protein surfaces containing charged residues, resulting in a greater capacity to capture cross-linked sites, especially inter-domain cross-linking information reflecting protein dynamics. Moreover, $BS(PEG)_2$ exhibited superior amphipathic properties, enabling it to penetrate cell membranes at higher concentrations and capture a more comprehensive set of protein structural information within cellular environments. It should be noted that the temperature is also a key factor regulating the protein dynamic as well as the backbone flexibility of the cross-linker. Moreover, due to the varying system scale, the temperature fluctuation may affect the dynamic properties of proteins and cross-linkers differently. To objectively compare the cross-linking data of different cross-linkers, all cross-linking experiments and various MD simulations were conducted at room temperature (25°C) to minimize the impact of temperature on the results.

In addition, cross-linkers with a PEG backbone offer scalability, meaning they can be lengthened while maintaining solubility. This raises the question of whether longer cross-linkers capture more useful cross-linking information. To investigate this, we employed longer cross-linkers with more PEG units, namely, $BS(PEG)_5$ and $BS(PEG)_9$ (Figure S8A available online at http://bib.oxfordjournals.org/). The MD simulations showed that even though the theoretical maximum arm length of $BS(PEG)_5$ is nearly twice as long as that of $BS(PEG)_2$, the actual simulations only stretched it up to 20.7 Å (distance between O_α and $O_{\alpha'}$), while $BS(PEG)_2$ reached 12.7 Å. Similarly, the maximum distance during MD simulation for $BS(PEG)_9$ was only 27.9 Å, much smaller than the theoretical maximum arm length of 38.4 Å (Figure S8B available online at http://bib.oxfordjournals.org/). XL-MS experiments using these longer cross-linkers indicated that the number of cross-linking sites captured by $BS(PEG)_5$ and $BS(PEG)_9$ are similar to that of $BS(PEG)_2$. For the MBP protein, the number of cross-linking sites captured by $BS(PEG)_9$ is even slightly less than that of $BS(PEG)_5$ (Table S3 available online at http://bib.oxfordjournals.org/). This suggests that the actual length distribution of the cross-linker does not increase proportionally with the lengthening of the PEG backbone unit.

Furthermore, it is important to note that a longer cross-linker is not always advantageous when studying protein dynamics. On one hand, longer cross-linkers imply longer distance restraints in structural calculations, which in turn reduce the overall precision of the structure. On the other hand, the so-called 'over-length cross-links', where the distance between cross-linked sites in the known structure exceeds the maximum arm length of the cross-linker, often indicate the presence of multiple conformations during protein dynamics. Such over-length cross-linked pairs hint at more compact conformational states. If the cross-linker arm lengths are excessively long, they may not provide similar structural information. Therefore, selecting an appropriate cross-linker length based on the protein size is crucial for accurately assessing dynamic alterations.

In conclusion, our comprehensive approach, integrating *in silico*, *in vitro* and *in vivo* XL-MS experiments, conclusively shows that $BS(PEG)_2$, a PEG-based cross-linker, outperforms BS^3/DSS in studying protein structure and dynamics. PEG-based cross-linkers offer superior capture capabilities, enhanced solubility and adjustable linker lengths tailored to research needs. Our findings offer valuable insights for future cross-linker design and optimization. First, for more comprehensive cross-linking, a cross-linker's backbone should primarily consist of polar groups like PEG and may benefit from incorporating mildly charged entities such as hydroxyl or thioether ester groups. Second, customizing the cross-linker's backbone length to match the size of the system and the dynamic range of motion is crucial for accurately capturing protein dynamics.

Key Points

- A comprehensive comparative analysis of two cross-linkers with different backbone properties, the hydrophilic $BS(PEG)_2$ and hydrophobic DSS/BS^3, carried out through a combination of *in vitro*, *in silico* and *in vivo* experiments.
- Clear evidence from both *in vitro* and *in vivo* cross-linking experiments strongly advocates for the superior capability of $BS(PEG)_2$ in capturing the intricate dynamics of multi-domain proteins.
- Profound insights derived from all-atom molecular dynamic simulations and quantum chemical calculation offer a deeper understanding of $BS(PEG)_2$, highlighting its intrinsic characteristics, notably heightened hydrophilicity and increased polarity, facilitating its close proximity to the protein surface.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/bib.

FUNDING

The work has been supported by the National Natural Science Foundation (31971155, 21991081, 22322411 and 22074139) and the Youth Innovation Promotion Association of the Chinese Academy of Sciences under grant No. 2020329 and No. 2020184.

DATA AVAILABILITY

All data generated or analyzed in this study are included in this article.

REFERENCES

1. Jackson RW, Smathers CM, Robart AR. General strategies for RNA X-ray crystallography. *Molecules* 2023;**28**(5):2111.

2. Gronenborn AM, Polenova T. Introduction: biomolecular NMR spectroscopy. *Chem Rev* 2022;**122**:9265–6.
3. Luchinat E, Cremonini M, Banci L. Radio signals from live cells: the coming of age of in-cell solution NMR. *Chem Rev* 2022;**122**: 9267–306.
4. Vedel IM, Papagiannoula A, Naudi-Fabra S, *et al.* Nuclear magnetic resonance/single molecule fluorescence combinations to study dynamic protein systems. *Curr Opin Struct Biol* 2023;**82**:102659.
5. Chari A, Stark H. Prospects and limitations of high-resolution single-particle Cryo-electron microscopy. *Annu Rev Biophys* 2023;**52**:391–411.
6. Dandey VP, Budell WC, Wei H, *et al.* Time-resolved cryo-EM using Spotiton. *Nat Methods* 2020;**17**:897–900.
7. Kikkawa M, Yanagisawa H. Identifying proteins in the cell by tagging techniques for cryo-electron microscopy. *Microscopy* 2022;**71**:i60–5.
8. Fan X, Wang J, Zhang X, *et al.* Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 angstrom resolution. *Nat Commun* 2019;**10**:2386.
9. Byer AS, Pei XK, Patterson MG, Ando N. Small-angle X-ray scattering studies of enzymes. *Curr Opin Chem Biol* 2023;**72**:102232.
10. Tants JN, Schlundt A. Advances, applications, and perspectives in small-angle X-ray scattering of RNA. *Chembiochem* 2023;**24**(17):e202300110.
11. A celebration of structural biology. *Nat Methods* 2021;**18**:427–7.
12. Cerofolini L, Fragai M, Ravera E, *et al.* Integrative approaches in structural biology: a more complete picture from the combination of individual techniques. *Biomolecules* 2019;**9**(8):370.
13. Chen J, Zhao Q, Gao H, *et al.* A Glycosidic-bond-based mass-spectrometry-cleavable cross-linker enables *in vivo* crosslinking for protein complex analysis. *Angew Chem Int Ed Engl* 2023;**135**:e202212860.
14. Gao H, Zhao LL, Zhong BW, *et al.* In-depth *in vivo* crosslinking in minutes by a compact, membrane-permeable, and Alkynyl-Enrichable Crosslinker. *Anal Chem* 2022;**94**:7551–8.
15. Kaake RM, Wang XR, Burke A, *et al.* A new *in vivo* crosslinking mass spectrometry platform to define protein-protein interactions in living cells. *Mol Cell Proteomics* 2014;**13**: 3533–43.
16. Steigenberger B, Pieters RJ, Heck AJR, Scheltema RA. PhoX: an IMAC-Enrichable cross-linking reagent. *ACS Cent Sci* 2019;**5**: 1514–22.
17. Wang JH, Tang YL, Gong Z, *et al.* Characterization of protein unfolding by fast cross-linking mass spectrometry using di-ortho-phthalaldehyde cross-linkers. Nature. *Communications* 2022;**13**(1):1468.
18. Britt HM, Cragnolini T, Thalassinos K. Integration of mass spectrometry data for structural biology. *Chem Rev* 2022;**122**: 7952–86.
19. O'Reilly FJ, Rappsilber J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* 2018;**25**:1000–8.
20. Lv L, Chen PH, Cao LZ, *et al.* Discovery of a molecular glue promoting CDK12-DDB1 interaction to trigger cyclin K degradation. *Elife* 2020;**9**:e59994.
21. Rappsilber J, Siniossoglou S, Hurt EC, Mann M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* 2000;**72**: 267–75.
22. Sun Q, Zhu X, Qu J, *et al.* Molecular architecture of the 90S small subunit pre-ribosome. *Elife* 2017;**6**:e22086.
23. Ibrahim M, Ramadan E, Elsadek NE, *et al.* Polyethylene glycol (PEG): the nature, immunogenicity, and role in the hypersensitivity of PEGylated products. *J Control Release* 2022;**351**:215–30.
24. Ding YH, Gong Z, Dong X, *et al.* Modeling protein excited-state structures from "over-length" chemical cross-links. *J Biol Chem* 2017;**292**:1187–96.
25. Gong Z, Gu XH, Guo DC, *et al.* Protein structural ensembles visualized by solvent paramagnetic relaxation enhancement. *Angew Chem Int Ed* 2017;**56**:1002–6.
26. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in *apo* maltose-binding protein observed by paramagnetic NMR. *Nature* 2007;**449**:1078–82.
27. Smith SL, Pitt AR, Spickett CM. Approaches to investigating the protein Interactome of PTEN. *J Proteome Res* 2021;**20**:60–77.