JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

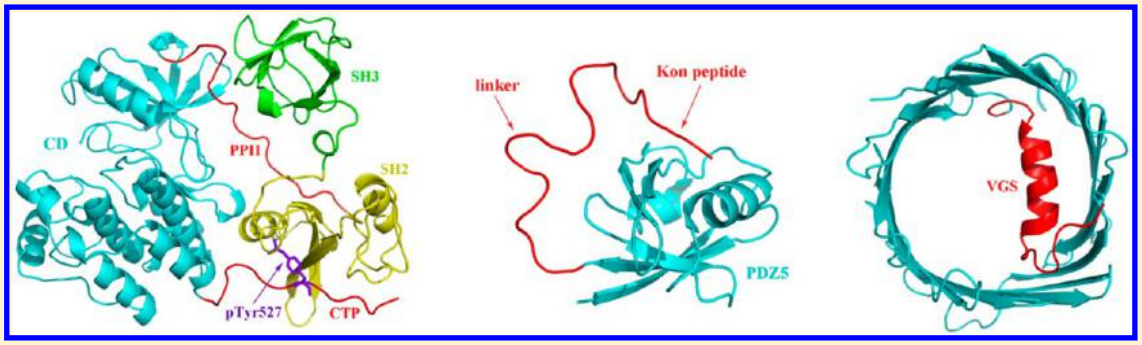
pubs.acs.org/jcim

Self-Binding Peptides: Folding or Binding?

Chao Yang,[†] Shilei Zhang,[†] Ping He,[‡] Congcong Wang,[†] Jian Huang,[†] and Peng Zhou*[,†]

[†]Center of Bioinformatics (COBI), Key Laboratory for Neuroinformation of the Ministry of Education, Center for Information in BioMedicine, School of Life Science and Technology, University of Electronic Science and Technology of China (UESTC), Chengdu 610054, China

[‡]Department of Cardiothoracic Surgery, Southwest Hospital, Third Military Medical University, Chongqing 400038, China



ABSTRACT: Self-binding peptides (SBPs) represent those short peptide segments within monomeric proteins to fulfill their biological functions by dynamically binding to/unbinding from their target domains in the same monomers. They are frequently found in disordered or unstructured regions of proteins that are now known to be not simply helices, loops, or linkers and adopt a well-defined structure upon binding. In this study, we systematically explored the structural basis, energy landscape, and thermodynamic properties of SBP-mediated biological mechanisms by dissecting dynamic interactions of SBPs with their cognate targets. In order to identify whether the formation of the SBP−target bound state is naturally a folding or a binding, we carried out atomistic molecular dynamics (MD) simulations to investigate both the native bound state of eight representative SBPs in complex with their cognate domains and the same bound state but where the SBPs were split artificially from the domains in primary sequence. Results showed that the splitting did not influence essentially the interaction behavior of SBPs with their target domains, suggesting that the SBP−domain interaction is almost a binding phenomenon, although the polypeptide linker between the SBP and the domain seems to facilitate the interaction. In addition, the sequence pattern of SBPs is very similar to those of short linear motifs (SLiM) found in many protein-binding peptides where few key residues contribute predominantly to protein−peptide recognition. All of the above comes together to suggest that the SBP is a novel biomolecular phenomenon spanning between folding and binding.

1. INTRODUCTION

Protein−protein interactions (PPIs) orchestrate various cellular processes predominantly in signaling and regulatory networks. It has recently become clear that many PPIs, particularly those that are transient, low-affinity, or related to post-translational modification events are mediated by the binding of a short linear peptide stretch in one protein to a globular domain in another.[1,2] Russell and co-workers estimated that up to 40% of all PPIs in the cell are mediated by short peptide fragments located at the interaction sites.[3,4] Due to their significance and estimated abundance, many researchers have attempted to design synthetic peptide drugs that have great potential for therapeutic intervention when these processes are deregulated. Peptide drugs can also disrupt protein−protein interfaces that cannot be inhibited by the available small molecules or even act as allosteric modulators.[5] Although therapeutic peptides have spurred increasing interest in the drug development community owing to their strong specificity, high biocompatibility and low toxicity profile,[6,7] a considerable number of protein−peptide interactions lack defined structures, especially for many peptide-mediated interactions in isolated monomer, which deserves us to do further exploration.

Traditionally, peptide-mediated interactions involve the binding of a peptide ligand to a separated globular protein receptor. Such interactions are very prevalent in cellular networks, orchestrating many key biological processes such as signal transduction,[8] protein transport,[9] and immune response.[10] Apart from above separated interaction between protein and peptide, peptide can also mediate the biological function of its parent monomeric protein by adopting a so-called "self-binding" approach, that is, a peptide segment, on the one hand, specifically recognizes and interacts with its cognate target to establish a dynamic balance between the bound complex and unbound state; on the other hand, the segment is integrated into the target in primary sequence via a

Received: August 29, 2014
Published: February 2, 2015

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

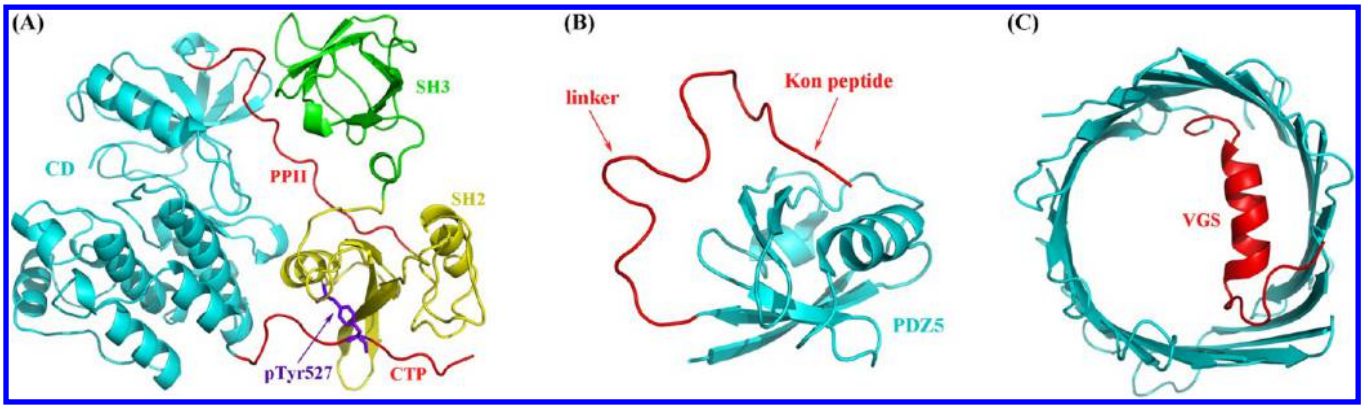


Figure 1. Some examples of SBPs. (A) *c*-Src kinase. This is a monomer that consists of an SH2 domain (colored in yellow), an SH3 domain (in green), and a CD catalytic domain (in cyan) as well as two SBPs, namely CTP and PPII (in red) that locate at C-terminus of the protein and between CD and SH2, respectively. The phosphorylation of Tyr527 residue induces CTP binding to SH2 domain, which causes an allosteric effect that promotes PPII−SH3 interaction and, in such a way, the CD domain is locked in an inactive state (PDB: 1FMK). (B) Artificial protein fused by a PDZ domain (in cyan) and a SBP (in red) composed of a Kon peptide and a linker. The Kon peptide prevents native peptide ligands from the peptide-binding pocket of PDZ when it is bound (PDB: 2LA8). (C) The voltage-dependent anion channel containing a SBP VGS (in red) at its N-terminus that controls the channel switching between open and closed states (PDB: 2JK4).

Table 1. Eight Representative SBPs in Seven Monomeric Proteins

PDB	core binding sequence (CBS)	target domain	monomeric protein
2LA8	^{100}RRNQYWV^{106}	PDZ5 domain	INAD PDZ5-Kon peptide fusion protein[15]
1FMK	^{521}TSTEPQ**pY**QPGENL^{533}[a]	SH2 domain	*c*-Src kinase[37]
1FMK	^{246}PTSKPQTQGLAKDA^{259}	SH3 domain	*c*-Src kinase[37]
3EFZ	^{259}CSGLLT**pS**AFF^{268}[a]	peptide-binding pocket	14−3−3 protein Cp14b[49]
2LBD	^{407}PMPPLIREMLENP^{419}	helix 12 binding pocket	human retinoic acid receptor hRARγ[27]
2JK4	^{1}MRGSAVPPTYADLGKSARDVFTKG^{24}	barrel wall	voltage-dependent anion channel[16]
1Z98	^{186}KRSARDSHVP^{195}	N-terminus	spinach aquaporin SoPIP21[38]
1RHZ	^{56}FEFWQTITAS^{65}	pore ring	SecYEβ protein-conducting channel[46]

[a] **pY**, phosphorylated tyrosine; **pS**, phosphorylated serine.

flexible polypeptide linker. Here, we call such biomolecular phenomenon *self-binding peptide (SBP)*; a typical SBP should satisfy following three features: (i) SBP is connected to its target in primary sequence and binds independently to the target in advanced structure, (ii) the SBP−target recognition is highly specific so that SBP can regulate specific biological process, and (iii) SBP can bind to/unbind from its target under a controllable mechanism such as pH value, phosphorylation, or ligand regulation.

As shown in Figure 1A, human nonreceptor tyrosine kinase *c*-Src is a monomeric protein consisting of a catalytic domain CD and two regulatory domains SH2 and SH3 as well as two SBPs, namely CTP and PPII that locate at C-terminus of the protein and between the CD and SH2, respectively. The phosphorylation of Tyr527 residue induces CTP binding to SH2 domain, which causes an allosteric effect that promotes the reversible interaction of PPII with SH3 domain and, in such a way, the CD domain is ultimately locked in an inactive conformation; the autoinhibitory state of *c*-Src can be released by Tyr527 dephosphorylation.[11−13] In addition to acting as molecular switcher to regulate enzyme's activity, SBPs exhibit a broad spectrum of biological functions such as improving protein thermostability,[14] acting as autoinhibitory peptide to prevent native ligands from occupying protein active pocket[15] (Figure 1B), and blocking ion channel's pore[16] (Figure 1C).

If an SBP can bind to its cognate target in a manner similar to that of conventional protein−peptide interaction but it is always connected to the target in primary sequence, the following question would arise: what is the driving mechanism underlying such biomolecular phenomenon, folding or binding? It is well recognized that folding and binding are two distinct strategies adopted by biomolecules to fulfill their biological functions. Proteins can work normally only if they are folded into proper structures and, then, bind specifically to their cognate partners. Otherwise, like some disordered proteins, they take a similar mechanism so-called coupled folding and binding to implement functions.[17,18] However, all of currently known binding mechanisms to which folding is coupled or not, including protein−peptide interaction, have two or more participants that are separated in sequence. In this regard SBP is an exception, where the self-regulated mechanism can be classified as either folding outcomes or binding effects. Of course, folding and binding may also be coupled in the SBP-mediated mechanism. Yet, no further study shows the structural characterization of the interaction of such a peptide with the target domain in the same monomers.

Elucidating the dynamics property and energy landscape of SBP−target interaction is a fundamentally important step to understand the molecular mechanism and biological implication underlying SBP. In the current work, we attempted to answer the questions like "is SBP a folding or a binding phenomenon?", "does the polypeptide linker contribute to the SBP−target interaction", and "what are the determinants to make a peptide SBP?". Here, a total of eight representative SBPs from seven monomeric proteins, including four soluble and three transmembrane proteins, were investigated in detail

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

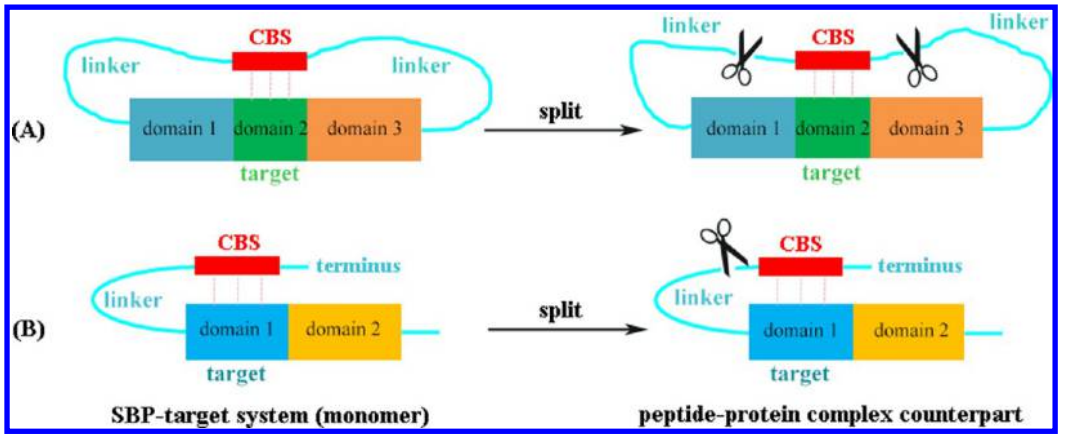


Figure 2. Schematic representation of splitting CBS from the rest of the monomer to model the peptide−protein complex counterpart of the corresponding monomeric SBP−target system. The SBP is made up of CBS and polypeptide linker(s) flanking the CBS, which can be located at either the middle (A) or terminus (B) of the monomer.

with respect to their structural, energetic, and dynamic aspects by using atomistic molecular dynamics (MD) simulations and post binding free energy analysis, particularly focusing on the comprehensive comparison of these SBPs with their counterparts where the polypeptide linkers between the SBPs and targets were split artificially. We also gave a systematic exploration on the molecular mechanism and biological significance of SBPs functioning in their parent proteins. This study would help to establish SBP as a new and special biomolecular phenomenon spanning between folding and binding.

2. MATERIALS AND METHODS

2.1. SBP Collection. The high-resolution crystal structures of seven monomeric proteins were retrieved from the PDB database,[19] from which eight typical SBPs were extracted according to the three basic features defined above (Table 1). The SBPs can be located at either the middle or terminus of their parent monomers. We manually examined these crystal structures and found that all the eight SBPs are in bound state, that is, the SBPs are complexed with their target domains in the crystallized structures. Therefore, the core binding sequence (CBS) and polypeptide linker of SBPs can be readily differentiated; the CBS is the peptide fragment that directly contacts its bound domain, which provides primary binding free energy for the CBS−domain system, while the polypeptide linker is the loop region between CBS and domain. Here, the contacts were defined as atomic van der Waals interactions at the CBS−domain interface, including wide contact, close contact, small overlap, and bad overlap,[20] which were identified using in-house 2D-Gralab program.[21]

There were two kinds of objects concerned in this study: the monomeric SBP−target system and its peptide−protein complex counterpart. The counterpart was modeled from the corresponding SBP−target system by manually breaking the polypeptide linker between the CBS and its target domain. In the procedure the CBS and its few flanking residues were split artificially from rest of the monomer (Figure 2).

2.2. Atomistic Molecular Dynamics (MD) Simulation. All MD simulations were carried out with Amber13 package[22] using AMBER FF99SB force field.[23] AmberTools were used to create topology and coordinate files for the structures of SBP−target systems and peptide−protein complexes. The particle mesh Ewald[24] (PME) and SHAKE algorithm[25] were utilized to treat long-range electrostatic interactions and to constrain all hydrogen atoms, respectively. Each system was equilibrated for 40 ns with time step of 2.0 fs, during which atomic coordinates saved every 10 ps for post analysis.

Soluble Protein System. Four proteins (PDB: 2LA8, 1FMK, 2EFZ, and 2LBD) investigated in this study are soluble globulin; they were solvated in a rectangular water box full of TIP3P water molecules[26] so that the boundary of the box is at least 12 Å away from any solute atom. The ligand binding domain of human retinoic acid receptor hRARγ (PDB: 2LBD) was cocrystallized with a small-molecule ligand, the all-trans retinoic acid, which has been found to be responsible for receptor's stability.[27] Thus, the ligand was kept during the MD simulations and parametrized with general amber force field (GAFF).[28] Counterions of Na^+ or Cl^- were placed based on the Columbic potential to keep the whole system neutral. The cutoff distance was set to 10 Å for short-range Coulomb and van der Waals interactions. Each system was first relaxed by 1000 steps of structure minimization, and then equilibrated at 300 K for 100 ps under a constant volume ensemble (NVT) with all heavy atoms restrained by harmonic model, followed by a 200 ps simulation under constant pressure of 1 atm without any constraint. The Langevin thermostat was applied using a collision frequency time of 1.0 ps^{-1} to maintain the production simulations at a constant temperature of 300 K, and the pressure was maintained by coupling to the reference pressure of 1 bar; relaxation time was considered for the simulations in bulk water.

Transmembrane Protein System. For the three investigated transmembrane proteins (PDB: 2JK4, 1Z98, and 1RHZ), each was embedded in a POPC membrane bilayer with preset orientation obtained from the OPM database,[29] by which the spatial arrangements of membrane proteins with respect to hydrocarbon core of the lipid bilayer were calculated empirically. Here, the lipids were assigned with the LIPID11 force field,[30] which is a modular force field specific for dynamics simulation of phospholipids and was designed to be compatible with the pairwise additive Amber force field. To avoid atomic collisions at the protein−lipid interface, lipid molecules that overlap on the embedded protein and channel pore were removed manually. During the simulation procedure, the cell was a rectangular periodic box hydrated with TIP3P water[26] and neutralized by counterions of Na^+ or Cl^-. The minimum distance between the protein and cell boundaries on each

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

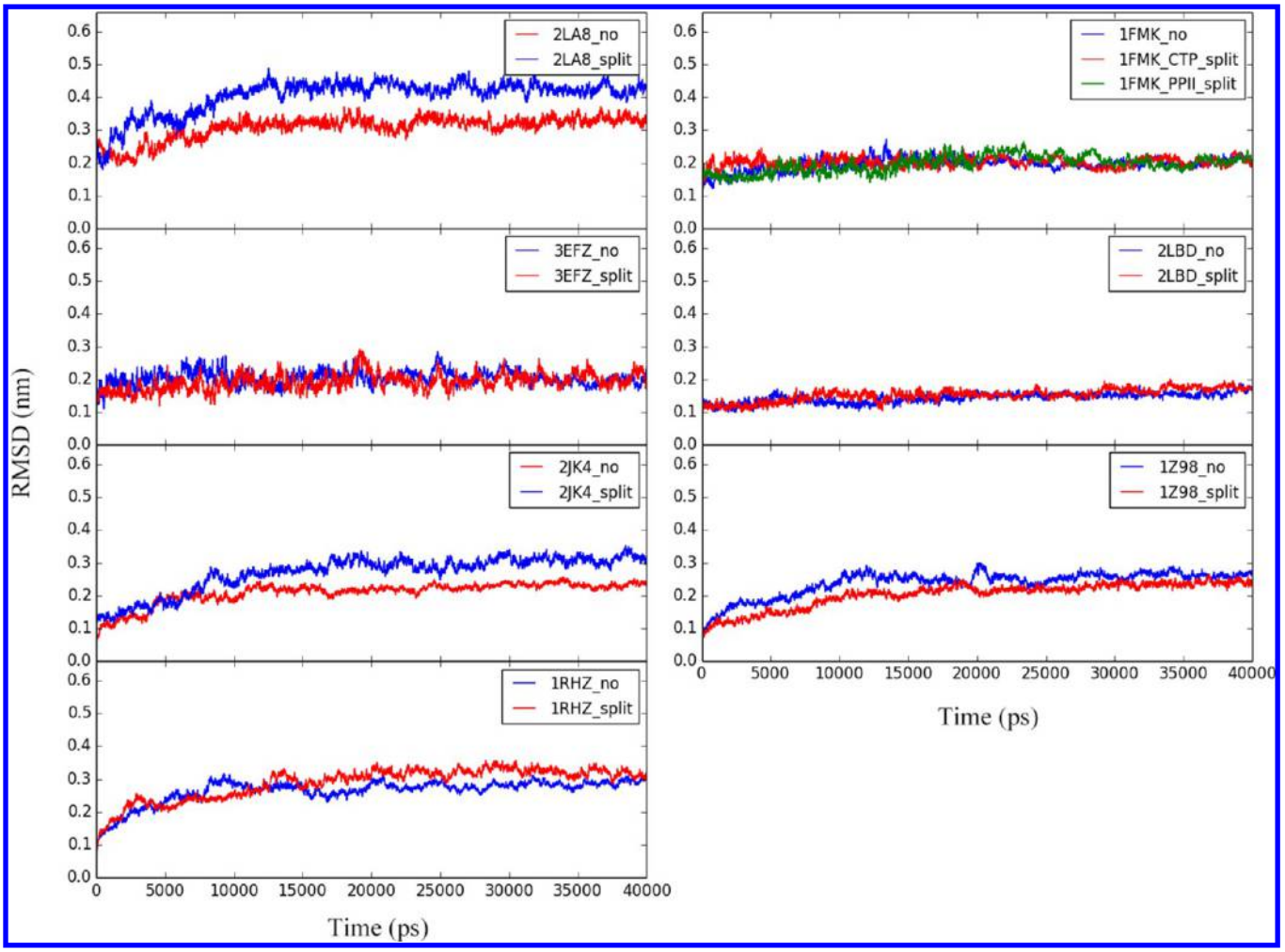


Figure 3. RMSD fluctuation profiles of seven monomeric proteins (PDB codes with tag "suffix_no") as well as eight peptide−protein complex counterparts (with "suffix_split") over 40 ns MD simulations: INAD PDZ5-Kon peptide fusion protein (PDB: 2LA8), *c*-Src kinase (PDB: 1FMK), 14−3−3 protein Cp14b (PDB: 3EFZ), human retinoic acid receptor hRARγ (PDB: 2LBD), voltage-dependent anion channel (PDB: 2JK4), spinach aquaporin SoPIP21 (PDB: 1Z98), and SecYEβ protein-conducting channel (PDB: 1RHZ).

direction was set to at least 20 Å so that the protein did not directly interact with its own image given in cutoff 10 Å.

To ensure appropriate disorder of lipid tails, each system was minimized with 2000 steps of steepest descent and then equilibrated at 300 K for 1 ns under an NVT ensemble, with all atoms (except those in lipid tails) being harmonically restrained by a force constant of 10 kcal·mol^{-1}·Å^{-2}. Further, above procedure was repeated for two times; first round, only protein atoms were restrained by harmonic model and, in second round, no restraint was imposed. Subsequently, a 40 ns production simulation with time step of 2 fs was performed to equilibrate the whole system at a constant temperature of 300 K and a constant pressure of 1 atm; the temperature and pressure were maintained using Berendsen barostat and Langevin thermostat methods, respectively. During the simulations SHAKE[25] was applied to constrain all bonds involving hydrogen atoms, PME[24] was used to calculate electrostatic potentials at a cutoff of 10 Å, and system coordinates were saved every 10 ps.

2.3. Binding Free Energy Analysis. After MD simulations post molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) analysis[31] implemented with the *mmpbsa* program[22] was carried out to calculate total binding free energy (ΔG_{total}) for the artificial peptide−protein complex counterparts, which can be computed from the absolute free energies of complex ($G_{complex}$), protein ($G_{protein}$), and peptide ($G_{peptide}$) in solution according to the formula:

$$\Delta G_{total} = G_{complex} - (G_{protein} + G_{peptide}) = \Delta E_{mm} + \Delta G_{dslv} - T \times \Delta S \quad (1)$$

where ΔE_{mm} is the molecular mechanics interaction energy between the peptide and protein, ΔG_{dslv} is the desolvation free energy upon peptide−protein binding, and ΔS is the conformational entropy change. The ΔE_{mm} consists of three components, i.e. the internal energy term (ΔE_{int}), the electrostatic energy term (ΔE_{elc}) and the van der Waals energy term (ΔE_{vdw}). The ΔG_{dslv} is a cocontribution from polar and nonpolar effects (ΔG_{sol_pol} and ΔG_{sol_npol}) due to draining out of water molecules from the complex interface upon peptide−protein binding; ΔG_{sol_pol} and ΔG_{sol_npol} were described using Poisson−Boltzmann (PB) approach and surface area (SA) model, respectively. The entropic penalty (ΔS) was estimated with normal-mode analysis (NMA)[32] using the *nmode* program.[22] For transmembrane protein systems, lipid molecules were removed from the trajectory in order to avoid the

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

influence of membrane environment on binding free energy calculations.

The binding free energy of SBP to its target in a monomeric protein can also be calculated using eq 1, but here the complex is the whole monomer, the peptide is the CBS and its flanking residues, and the protein is rest of the monomer. The snapshots extracted from the last 10 ns MD trajectories were used to perform binding free energy analysis for the monomeric SBP−target systems and their peptide−protein complex counterparts.

3. RESULTS AND DISCUSSION

3.1. Overview of the Simulation Procedure. We performed 40 ns atomistic MD simulations for each of the 7 monomeric proteins (corresponding to 8 SBP−target systems) and 8 peptide−protein complex counterparts. During the simulation process several important physics quantities including temperature, volume, total energy and mass density were monitored and found to be relatively stable over the simulation period. Here, the root-mean-square deviation (RMSD) fluctuations of backbone heavy atoms as a function of time for different systems are shown in Figure 3. As can be seen, all systems have achieved to equilibration state in the later stage of simulation procedure. It is suggested from the RMSD profiles that after simulations all the investigated systems retained in ordered structure similar to their initial state; modeled structures generally had a slight increase in RMSD as compared to corresponding original structures:

- The average RMSDs of modeled (split) 2LA8 protein and unmodeled (unsplit) 2LA8 protein are 4.17 ± 0.31 and 3.24 ± 0.28 Å, respectively; the two systems were convergent after ∼14 ns simulations.
- The average RMSDs of modeled 1FMK_CTP protein and 1FMK_PPII protein as well as unmodeled 1FMK protein are 2.06 ± 0.26, 2.13 ± 0.31 and 1.98 ± 0.17 Å, respectively; the three systems were convergent after ∼16 ns simulations.
- The average RMSDs of modeled 3EFZ protein and unmodeled 3EFZ protein are 2.04 ± 0.30 and 1.98 ± 0.33 Å, respectively; the two systems were convergent after ∼22 ns simulations.
- The average RMSDs of modeled 2LBD protein and unmodeled 2LBD protein are 1.72 ± 0.14 and 1.67 ± 0.11 Å, respectively; the two systems were convergent after ∼9 ns simulations.
- The average RMSDs of modeled 2JK4 protein and unmodeled 2JK4 protein are 3.04 ± 0.26 and 2.29 ± 0.20 Å, respectively; the two systems were convergent after ∼27 ns simulations.
- The average RMSDs of modeled 1Z98 protein and unmodeled 1Z98 protein are 2.34 ± 0.18 and 2.61 ± 0.23 Å, respectively; the two systems were convergent after ∼25 ns simulations.
- The average RMSDs of modeled 1RHZ protein and unmodeled 1RHZ protein are 3.08 ± 0.20 and 2.87 ± 0.16 Å, respectively; the two systems were convergent after ∼22 ns simulations.

Due to translation and rotation in the simulations, it is considered reasonably that there are no essential difference between the split and unsplit systems. The reason that the split systems generally showed a slightly larger RMSD profile than that of unsplit systems can be explained by structure disorder contributed from breaking polypeptide linkers. Thus, it is readily concluded that no significant conformational rearrangement can be found during MD simulations, suggesting that binding, not folding, dominates the structural stability of SBP−target interactions.

For transmembrane proteins the electron density profile provides a time-averaged measurement of the density of electrons through lipid bilayer of the proteins (Figure 4).

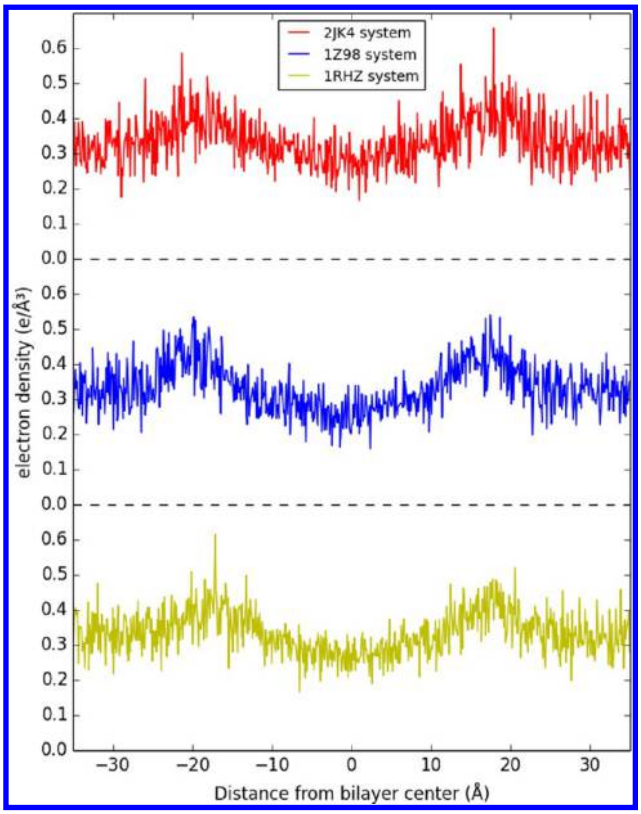


Figure 4. Electron density profiles of the three transmembrane protein systems over 0−20 ns MD simulations.

Here, the electron density averaging over 0−20 ns trajectory for each system was calculated using the Density Profile Tool of VMD.[33] The maximal extension in the electron density profile was simply used as bilayer thickness, which often represents the distance between the mean positions of phosphorus atoms in each leaflet.[34,35] The average thickness of POPC bilayers in three membrane systems was estimated to be ∼38 Å, which is very close to the experiment value (39.1 Å at 30 °C) reported previously,[35] suggesting an appropriate feature of membrane setup in our MD simulations.

3.2. Dynamic Analysis of SBPs in Split Peptide−Protein Complex Systems. Considering that the dynamic feature of SBP (the self-binding peptide fragments) may not be clear with respect to the RMSD profile of whole protein−peptide complex system, two-dimensional RMSD heatmap focused on the SBP portions of different modeled (split) systems are shown in Figure 5, which reveals a stable dynmaic behavior for these SBPs in split structures, with only slight variations through the course of simulations. Typically, there is a small variation in the split 2LA8 system at approximately 5 ns, which, however, does not change significantly the structural conformation of SBP; this variation can be attributed to the adaptation of modeled structure due to spliting the SBP from its binding partner. After the adaptation is completed, a stable

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

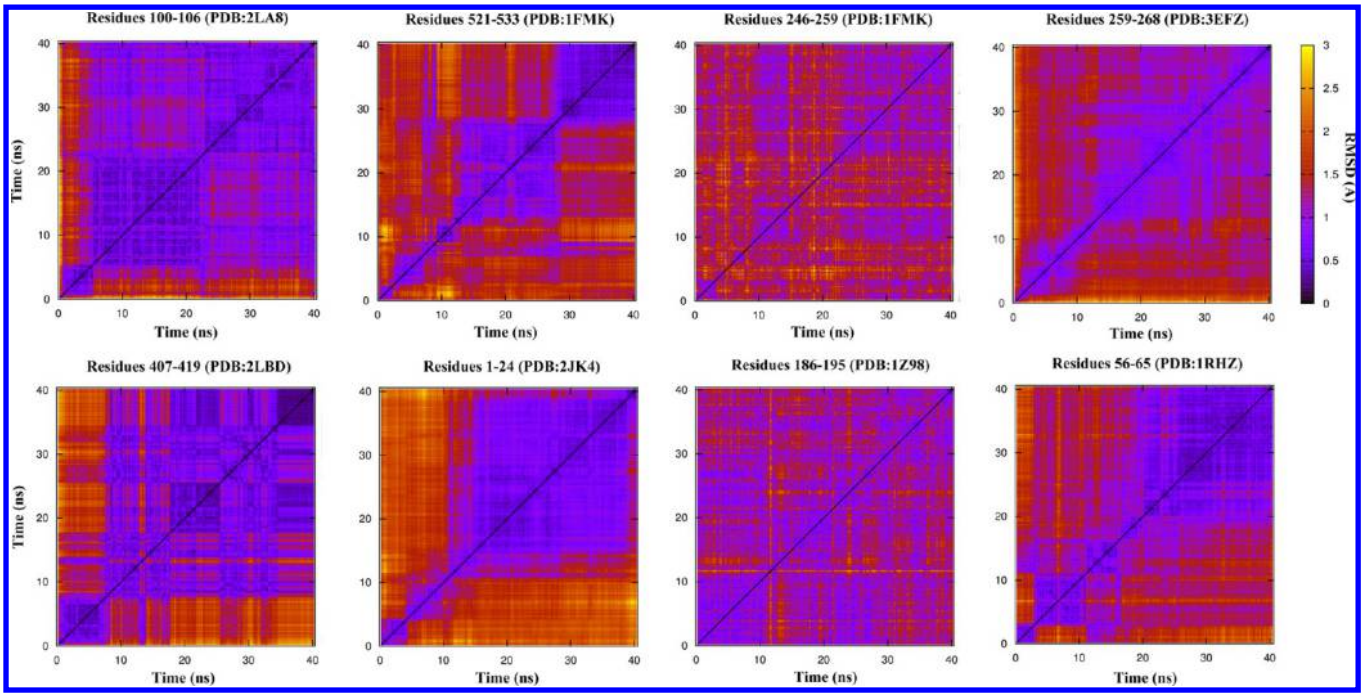


Figure 5. RMSD versus time correlation plots derived from 40 ns MD simulations of SBPs in different split protein−peptide systems.

conformational state was reached until the end of simulations. Similar RMSD plots of 1FMK (CTP), 3EFZ, 2LBD, and 2JK4 showed small changes occurring no more than 30 ns simulations.

3.3. Structural Analysis of SBP-Linkers of Intact SBP−Target Systems vs Split Peptide−Protein Complex Counterparts. Secondary structure evolution of SBP with polypeptide linker in all the eight monomeric SBP−target systems during 40 ns MD simulations was tracked to explore whether the presence of polypeptide linkers can influence SBP−target interactions in monomeric proteins. For comparison, the eight corresponding peptide−protein complex counterparts that the SBPs and their targets were split artificially in sequence were also investigated. The averaged secondary structure propensities of SBP-linker residues in monomeric proteins were analyzed using the DSSP method of Kabsch and Sander[36] implemented with the *cpptraj* program,[22] which assigned secondary structure classes for each residue based on backbone atom position fluctuation in the MD simulations.

It is evident from Figure 6 that the secondary structure propensities of SBP in all the seven monomeric proteins can be well-maintained during the whole MD simulation procedure, implying a high stability of the SBP−target bound state. For the SBP−target systems 2LA8, 3EFZ, 1FMK (CTP), 2LBD, and 2JK4, where the SBPs locate at protein terminus, the secondary structure classes of polypeptide linkers are as diverse as 3_{10}-helix, β-sheet, *t*-turn, and random coil, in which the random coil seems to be dominant due to its large flexibility and spatial extension. However, for the systems 1Z98, 1RHZ, and 1FMK (PPII) with SBPs at the middle of monomeric proteins, the linkers tend to fold into *t*-turn and/or antiparallel β-sheet structures. Such feature makes polypeptide linkers highly flexible so that the SBPs can convert swiftly between bound and unbound states. The residue Trp260 locates within a β-sheet at one side of the SBP peptide PPII in *c*-Src kinase (PDB: 1FMK), which anchors C-terminal portion of the PPII by packing tightly against a neighboring "C" helix.[13,37] The PPII touches on the recognition surface of *c*-Src SH3 domain to form weak nonspecific contacts between them, and thus the location of PPII is crucial for its intramolecular interactions with the SH3 domain (Figure 1A).

Another example is the loop D in spinach plasma membrane aquaporin SoPIP21 (PDB: 1Z98) (Figure 7). The water flux channel of plant aquaporin is in closed conformation under drought stress or flooding conditions. The SBP-regulatory mechanism is that the loop D interacts with the N-terminus of aquaporin, inducing Leu197 residue to cap on the channel from cytoplasm environment.[38,39] The polypeptide linkers at both sides of the loop D are antiparallel β-sheet structure, which anchors the loop D into an appropriate position for its functional activity. In addition, the linkers of 1RHZ system connect transmembrane helices TM2a (SBP segment) with TM1 and TM2b, which can also be recruited to form a new helical plug in the TM2a deletion mutant structure.[40] This process would disrupt β-hairpin loop and induce the formed plug rotating away from pore ring. Thus, although a similar helix structure of TM2a is shaped, the plug has lost interactions with the pore ring and can no longer stabilize lateral gate.

The SBP peptides in 8 peptide−protein complex counterparts showed a high consistency of bound conformation with corresponding SBP−target systems during the 40 ns MD simulations (Figure 6), except few linker regions that were broken in the former. For example, the linker Arg97-Asn102 of INAD PDZ5-Kon peptide fusion protein (PDB: 2LA8) loses its 3_{10}-helix structure in artificial peptide−protein complex; the antiparallel β-sheet in the linker region of spinach aquaporin SoPIP21 (PDB: 1Z98) become random coil in split counterpart. As the CBS is flanked by few linker residues in the counterpart, while the secondary structure of linker region is disrupted, the SBP maintains its original conformation by interacting with target. In this respect, it is suggested that the structural profile has no essential difference between the intact SBP−target systems and corresponding split peptide−protein

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

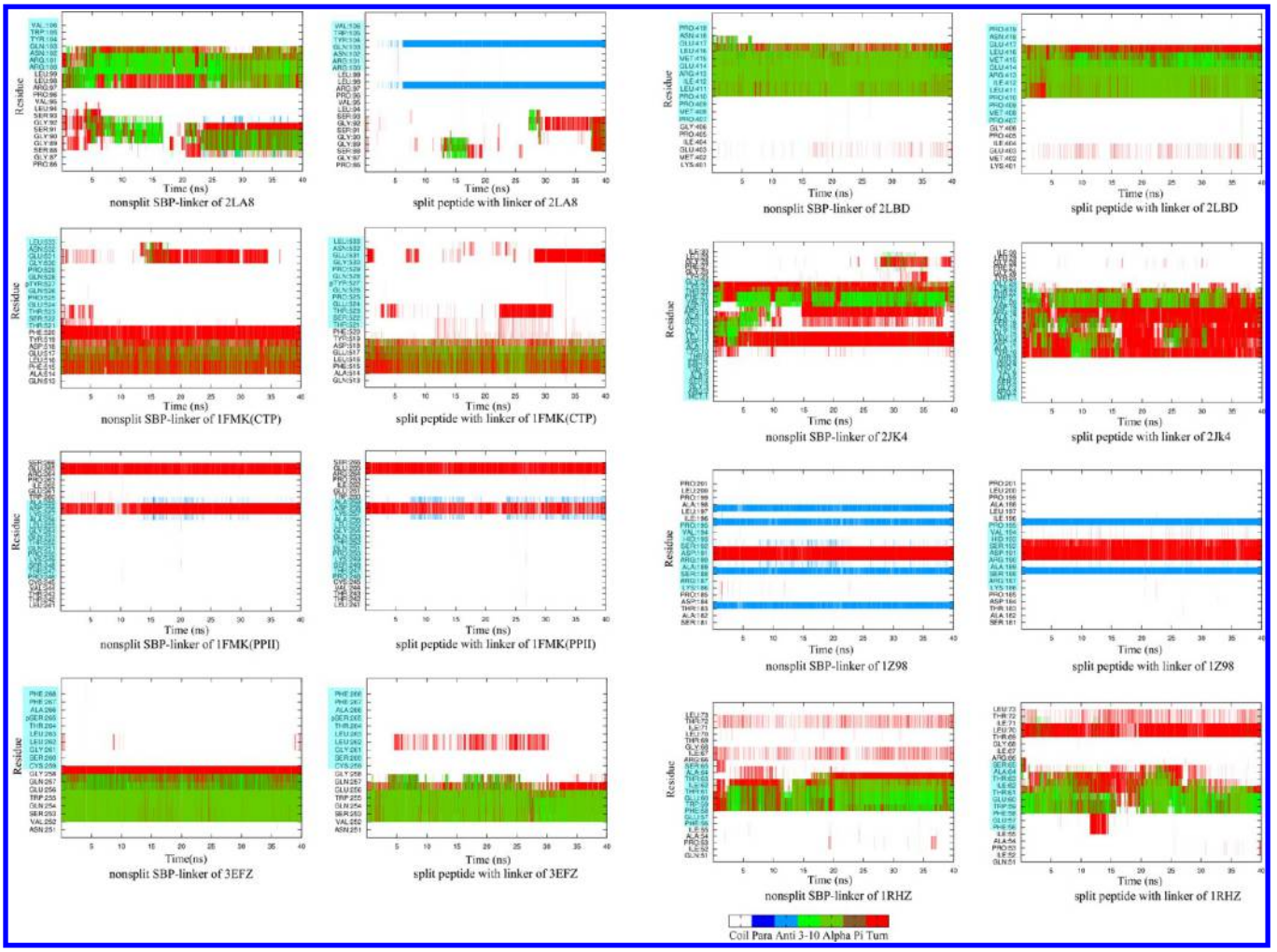


Figure 6. Secondary structure evolution of SBP-linker in 8 SBP−target systems as well as their peptide−protein complex counterparts during 40 ns MD simulations. In each plot the location of SBP segment or corresponding peptide is highlighted (cyan): para, parallel β-sheet; anti, antiparallel β-sheet; 3−10, 3−10 helix; alpha, α-helix; pi, π-helix.

complex counterparts, supporting that breaking polypeptide linkers does not influence essentially the binding behavior of SBPs to their target domains.

3.4. Binding Features of Intact SBP−Target Systems vs Split Peptide−Protein Complex Counterparts. Dynamic analysis of the residue contacts between SBPs and domains were carried out to explore if the binding mode of split peptide−protein complexes is in line with corresponding intact SBP−target systems. Here, the minimum distance between the heavy atoms of pairwise residues was considered as residue contact distance, and the contact map of each system was derived from the last 10 ns MD trajectories. As shown in Figure 8, very similar contact profiles of most SBP−target systems with their peptide−protein counterparts are clearly emerged, except the voltage-dependent anion channel (PDB: 2JK4). The channel has a helical voltage-dependent anion gating switch (VGS) that locates horizontally within the channel pore[16] (Figure 1C); the broken linker seems to have a considerable influence on the contact interactions between VGS's edge region (residues 18−24) and the four N-terminal β-strands (residues 29−78) of barrel wall. However, it is worth noting that these regions only constitute weak and nonspecific random contacts that do not influence substantially the VGS-barrel wall binding behavior.

The most frequent conformations in the last 10 ns snapshots were retreived and shown in Figure 9. As can be seen, both the modeled and unmodeled structures retained in stable conformation that SBPs interact with native domain area in crystal structure. The modeled complexes exhibit slight structural motion due to the conformational flexiblilty of split polypeptide linkers, which agrees well with the dynamic analysis of SBPs in Figure 5. Thus, it can be readily concluded that the SBP segments bind specifically to their target domains with high specificity to establish functional interactions and adopt a well-defined structure upon the binding, even they are totally split from the targets in primary sequence. As might be expected, the core binding sequence of SBP contributes largely to the basic functional characteristic in SBP−target interaction. The above dynamic results of the split peptide−protein complexes show a unified structure−function relationship of SBPs without folding effect. Similar to those conventional peptide-mediated interactions that a short linear motif of about 10 residues from one protein is recognized by a globular domain in another with specific interaction mode,[41,42] the SBP segments also bind specifically to their target domains with high specificity to establish functional interactions and adopt a well-defined structure upon the binding, even they are totally split from the targets in sequence. Thus, the folding effect in the

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

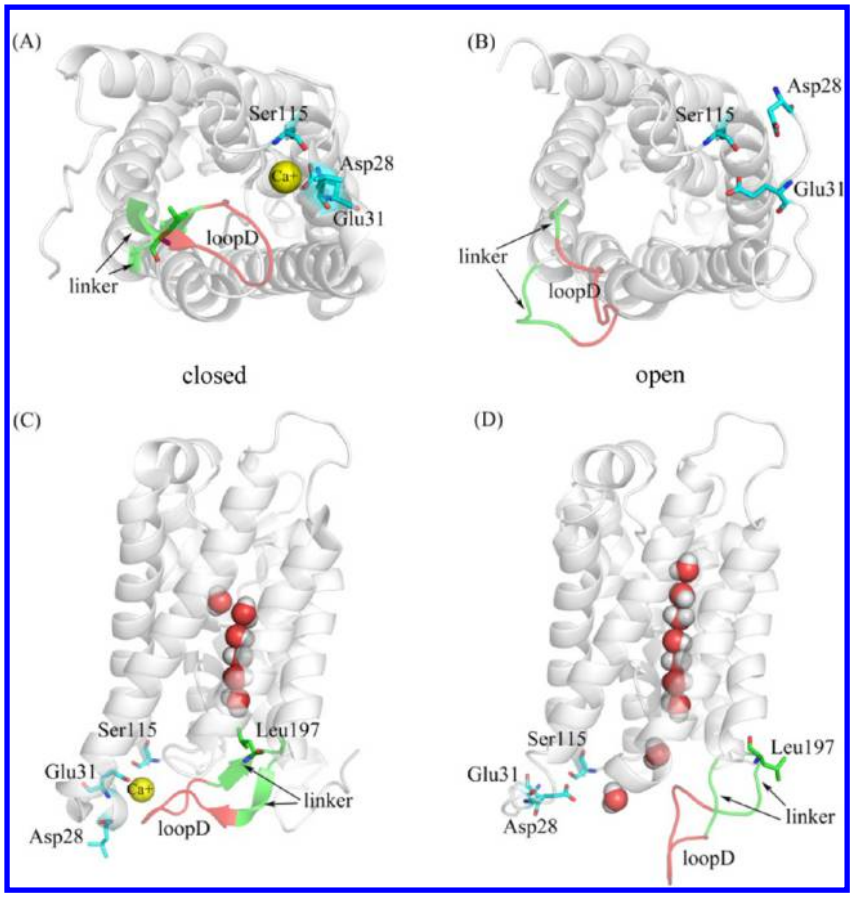


Figure 7. Cartoon representation of the closed and open conformations of spinach plasma membrane aquaporin SoPIP21 (PDB: 1Z98). The protein transforms from open (B) to closed (A) states under drought stress or flooding conditions. (C and D) Lateral views of the closed and open state, respectively. The loop D interacts with key residues Asp28 and Glu31 at the N-terminus of SoPIP21 and then caps on water flux channel, while the residue Ser115 acts as the regulatory button of SBP−target binding/unbinding by its phosphorylation/dephosphorylation.

SBP−target recognition can be largely ignored and only binding interaction is sufficient to explain the regulatory mechanism of SBP.

3.5. Energetic Analysis of Intact SBP−Target Systems vs Split Peptide−Protein Complex Counterparts. The total binding free energies (ΔG_{total}) of 8 SBP−target systems and corresponding peptide−protein complex counterparts were calculated empirically using the sophisticated MM/PBSA approach. The ΔG_{total} can be decomposed into different energy components and listed in Table 2. Given the fact that SBP converts from a flexible peptide to a rigid, well-defined structure upon its binding to target, the configuration entropy penalty (ΔS)[43,44] was also considered here and calculated via normal-mode analysis approach[32] based on MD trajectory. It is seen from Table 2 that there is no significant difference between the total binding free energies of SBP−target systems and their counterparts ($|\Delta G^{split}_{total} - \Delta G^{no}_{total}| < 5$ kcal·mol^{-1}), which could be considered within the reasonable error range of calculations due to random noise and structural disorder. In addition, although electrostatic nonbonded interaction seems to have a quite favorable contribution ($\Delta E_{elc} < 0$) to SBP−target binding in vacuum owing to lack of solvent screening, which would be largely offset by the unfavorable polar solvent effect ($\Delta G_{sol\_pol} > 0$). Consequently, the nonpolar aspect ($\Delta E_{vdw} + \Delta G_{sol\_npol}$) seems to dominate the binding in solvent as compared to polar facet ($\Delta E_{elc} + \Delta G_{sol\_pol}$). Overall, the eight SBP−target systems and their counterparts exhibit a similar profile in both enthalpy contribution (ΔH) and entropy lost ($-T\Delta S$), suggesting that breaking polypeptide linker does not have a substantial influence on SBP−target interaction.

As can be seen from Table 2, the internal energy term (ΔE_{int}) of SBP−target systems is primarily dominated by torsion angle energy (ΔE_{tors}), which showed to be increasing during the whole MD simulations. Different from split peptide−protein complexes, the polypeptide linkers in intact SBP−target systems may be distorted due to the binding of SBPs to their target domains; this is particularly notable if the linkers are short and/or structured. The distortion is apparently unfavorable to the binding, but this unfavorable effect is not significant if considering that most polypeptide linkers are in loop regions that are highly flexible and deformable. According to our calculations the internal energy changes upon the breaking of different polypeptide linkers range between 5 and 20 kcal·mol^{-1}, suggesting that the linkers would indeed influence the interaction behavior of SBPs with targets, albeit the influence appears to be quite modest for most systems as reflected in total binding free energies. This finding supports that, in addition to binding, a marginal folding effect is also involved in the SBP−target association process owing to the existence of polypeptide linkers.

Next, the total binding free energy of SBP−target interaction was assigned into each residue pair at the interaction interface and, in this way, hot-spot residues that contribute predominantly to SBP−target binding can be readily identified. In the

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

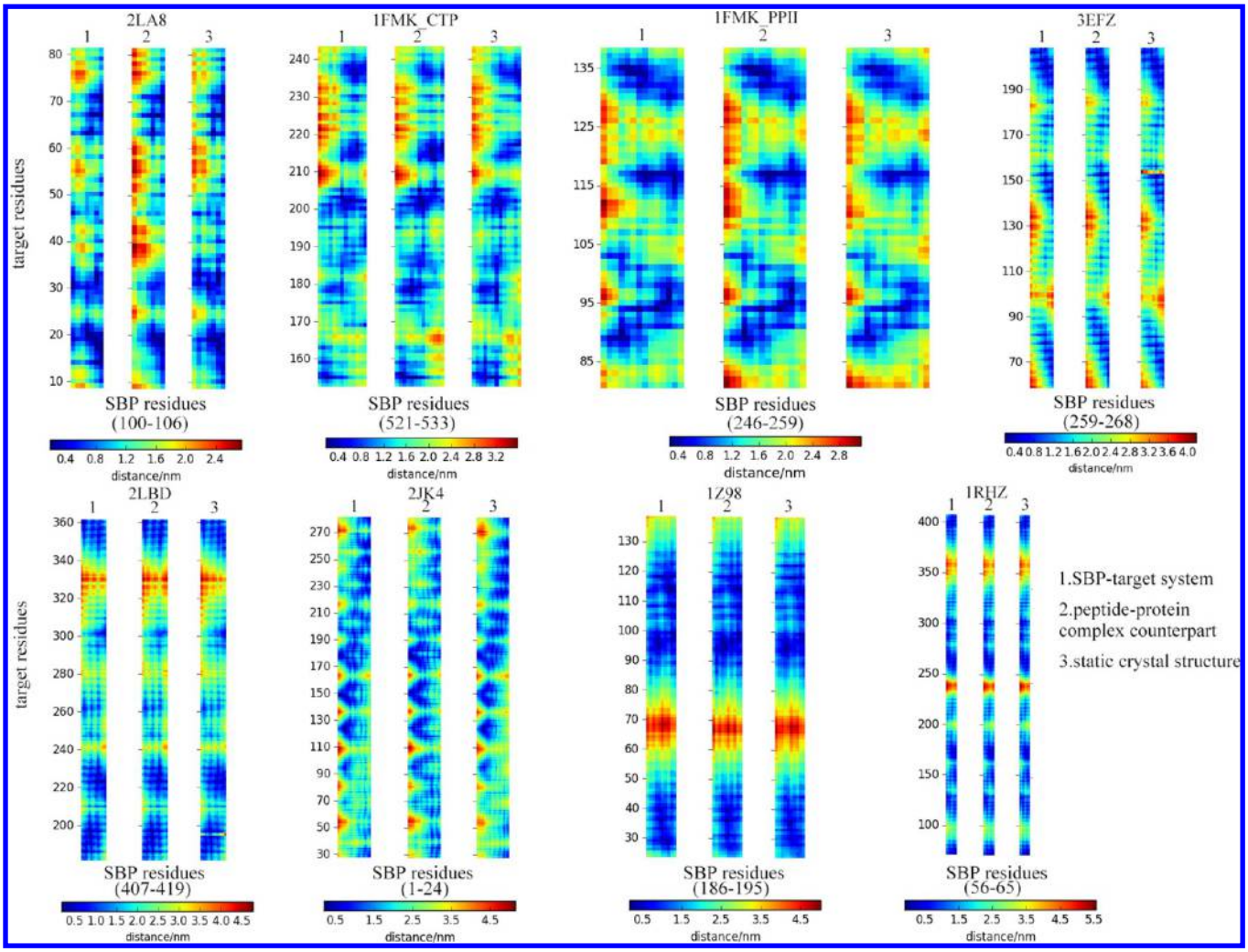


Figure 8. Contact maps between the SBPs and their target domains for the eight SBP−target systems and their peptide−protein complex counterparts as well as those in corresponding static crystal structures. The maps were derived from the last 1 ns MD trajectories.

procedure, only those two residues separately from SBP and target and with apparent contributions were selected as residue pairs. The results of binding energy assignment are shown in Figure 10, from which it is evident that most hot-spot residues are shared by both the intact SBP−target systems and their split counterparts, except few in INAD PDZ5-Kon peptide fusion protein (PDB: 2LA8) and voltage-dependent anion channel (PDB: 2JK4) that are nearby the broken sites of polypeptide linkers. From the residue pair interaction energy maps (Figure 10) we can straightforwardly identify a number of shared hot spots and exemplified some as follows: (i) phosophorylation of Tyr527 and Ser265 are the key events in the SBP−target binding of, respectively, *c*-Src kinase (PDB: 1FMK) and 14−3−3 protein Cp14b (PDB: 3EFZ), as these two modified residues confer a large stability for the bound SBP−target architectures; (ii) the sequence region 249−253, in particular residues Lys249 and Pro250, contributes significantly to the binding affinity of SBP peptide PPII to SH3 domain in *c*-Src kinase (PDB: 1FMK), which is in line with the defined feature of class II SH3-binding peptide motif;[37] in addition, the C-terminal residues of the PPII (256−259) form a β-turn (shown in Figure 6), which also contribute a lot to PPII-SH3 interactions; (iii) the residues Glu414 and Glu417 in human retinoic acid receptor hRARγ should play a critical role in the self-binding process of SBP peptide helix12 (Figure 9E); this finding can be supported by a previous experimental evidence that point mutation of these two residues would considerably impair the structural stability and biological functionality of hRARγ,[27] and (iv) the loop D residues 190−193 of spinach aquaporin SoPIP21 (PDB: 1Z98) also exhibit favorable interaction with the N-terminal region of the protein.

3.6. Two Examples of Marginal Folding Effect Involved in SBP−Target Systems. It is known that the four C-terminal residues ^{103}QYWV^{106}-COOH of Kon peptide (Figure 1B) are the primary binding site of PDZ5 domain in the INAD PDZ5-Kon peptide fusion protein[15,45] (PDB: 2LA8). The Kon peptide binds to PDZ5 domain by adding it as an antiparallel β-strand to PDZ5 β-sheet structure αB/βB (Figure 9A) and inserting its Val106 residue into the hydrophobic pocket of PDZ5 active site; this binding manner is consistent in both the intact system and split complex of Kon peptide−PDZ5 interaction. However, the polypeptide linker between Kon peptide and PDZ5 domain would lose its 3_{10}-helix structure due to it broken (Figure 6), which disrupts the established interactions of adjacent Arg100 residue with PDZ5 domain, causing a marked structural degradation in the N-terminus of Kon peptide (Figure 10).

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

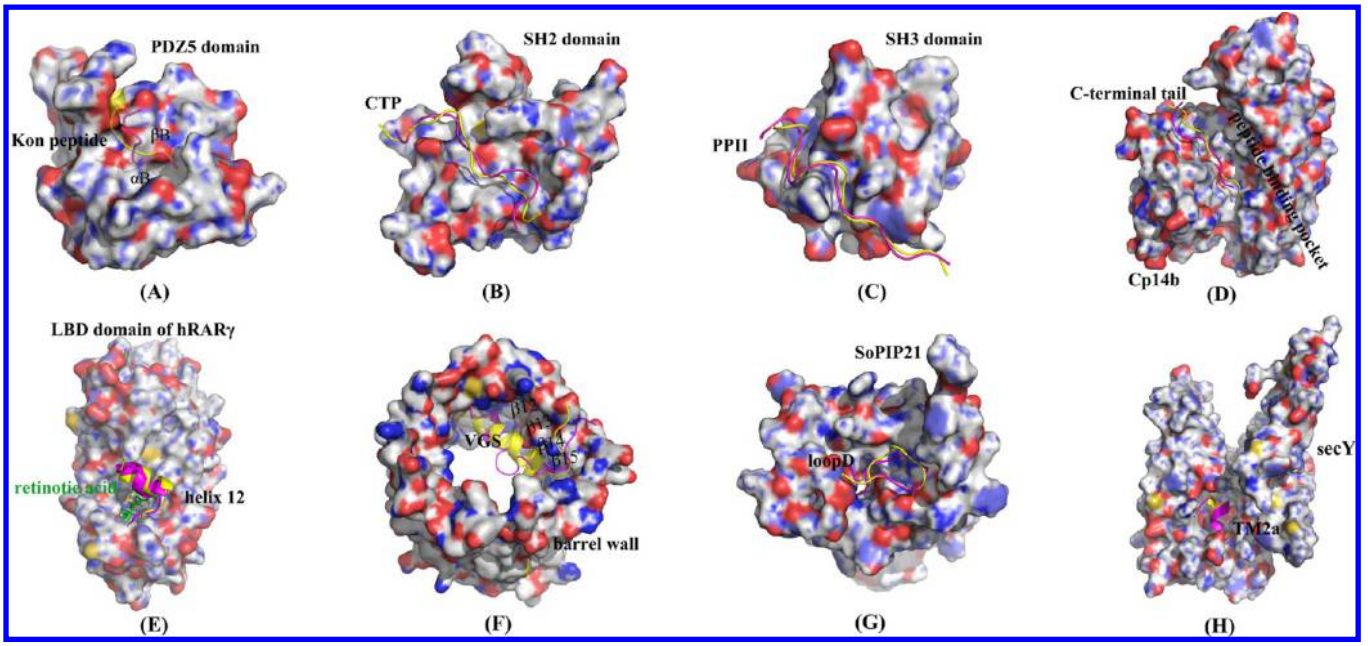


Figure 9. Comparison of the MD-equilibrated conformations of SBPs complexed with their target domains when the SBPs are connected to (yellow) and split from (magentas) the targets in sequence: (A) INAD PDZ5-Kon peptide fusion protein (PDB: 2LA8), (B) *c*-Src kinase CTP (PDB: 1FMK), (C) *c*-Src kinase PPII (PDB: 1FMK), (D) 14−3−3 protein Cp14b (PDB: 3EFZ), (E) human retinoic acid receptor hRARγ (PDB: 2LBD), (F) voltage-dependent anion channel (PDB: 2JK4), (G) spinach aquaporin SoPIP21 (PDB: 1Z98), and (H) SecYEβ protein-conducting channel (PDB: 1RHZ).

Table 2. Total Binding Free Energies and Decomposed Energetic Components of Eight SBP−Target Systems (Tagged with "no") and Their Peptide−Protein Complex Counterparts (Tagged with "split") (kcal·mol^{-1})

	ΔE_{nbd}		ΔE_{int}										
system (PDB_tag)	ΔE_{elc}	ΔE_{vdw}	ΔE_{bond}	ΔE_{angle}	ΔE_{tors}	ΔE_{gas}	ΔG_{sol_npol}	ΔG_{sol_pol}	ΔG_{dslv}	ΔG_{polar}	ΔH	$-T\Delta S$	ΔG_{total}
2LA8_no	−90.59	−71.48	0.35	1.04	7.41	−153.27	−6.90	100.58	93.68	9.99	−50.59	34.41	−16.18
2LA8_split	−64.62	−59.56	0.00	0.00	0.00	−124.18	−5.11	81.97	76.86	17.35	−47.32	34.23	−13.09
1FMK_CTP_no	−636.36	−40.23	0.33	1.30	5.32	−669.64	−5.34	596.33	590.99	−40.03	−78.65	54.31	−24.34
1FMK_CTP_split	−667.79	−55.85	0.00	0.00	0.00	−723.64	−6.94	642.38	635.44	−25.41	−88.20	59.52	−28.68
1FMK_PPII_no	−501.99	−119.88	0.74	2.14	20.08	−608.92	−13.66	530.15	516.49	18.16	−92.43	50.45	−41.98
1FMK_PPII_split	−551.55	−110.63	0.00	0.00	0.00	−662.18	−13.20	572.97	559.77	21.42	−102.41	53.77	−48.64
3EFZ_no	−358.68	−64.75	0.30	1.08	5.47	−414.84	−7.36	343.16	335.80	−15.52	−79.04	48.96	−30.08
3EFZ_split	−416.85	−55.25	0.00	0.00	0.00	−472.10	−6.22	400.51	394.29	−16.34	−72.81	45.31	−27.50
2LBD_no	−103.16	−57.06	0.42	1.04	11.19	−146.06	−6.41	117.05	110.64	13.89	−35.42	30.23	−5.19
2LBD_split	−146.07	−61.54	0.00	0.00	0.00	−207.61	−6.72	174.36	167.64	28.29	−39.97	32.57	−7.34
2JK4_no	−157.78	−97.04	0.29	1.04	7.10	−246.39	−13.39	205.45	192.06	47.67	−54.33	50.36	−3.97
2JK4_split	−176.28	−71.51	0.00	0.00	0.00	−247.79	−9.80	209.37	199.57	33.09	−48.22	46.58	−1.64
1Z98_no	−233.65	−76.63	0.86	2.07	13.93	−293.41	−8.71	237.37	228.66	3.72	−64.75	40.14	−24.61
1Z98_split	−244.69	−62.75	0.00	0.00	0.00	−307.44	−7.64	253.41	245.77	8.72	−57.68	37.21	−20.47
1RHZ_no	−123.74	−64.58	0.67	2.01	12.82	−172.83	−6.47	126.15	119.68	2.41	−53.15	35.07	−18.08
1RHZ_split	−165.28	−57.88	0.00	0.00	0.00	−223.16	−6.21	178.51	172.30	13.23	−50.86	33.65	−17.21

Although the SBP peptide VGS of voltage-dependent anion channel (PDB: 2JK4) locates horizontally inside the barrel-shaped pore of the transmembrane protein in both intact system and split complex, breaking polypeptide linker would impair the interaction strength between VGS and barrel wall (Figure 9F), leading to a slight motion of VGS from the barrel wall, during which several hot-spot residues that originally present in SBP−target system are lost (Figure 10). Binding energy analysis can also well reflect this point; the binding affinity of SBP−target system was estimated to be −54.33 kcal·mol^{-1}, which is larger than that −48.22 kcal·mol^{-1} of corresponding complex counterpart (see Table 2), confirming that the linker breaking can modestly influence the packing behavior of VGS within the transmembrane pore of voltage-dependent anion channel.

3.7. Overall Understanding of SBP-Mediated Biological Mechanism. The results obtained from above intensive computational investigations suggest that the intact SBP−target systems and their split counterparts share a consistent profile of structural property, energy landscape and dynamics behavior, where specific sequence pattern and few hot-spot residues of peptide segments are essential to the recognition and interaction of SBPs with their target domains, while the polypeptide linkers between the SBPs and domains seem also to have marginal effect on the interaction. Since SBP is a portion of its parent monomer and connected to its target via a flexible linker, it is supposed that SBP adopts a so-called

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

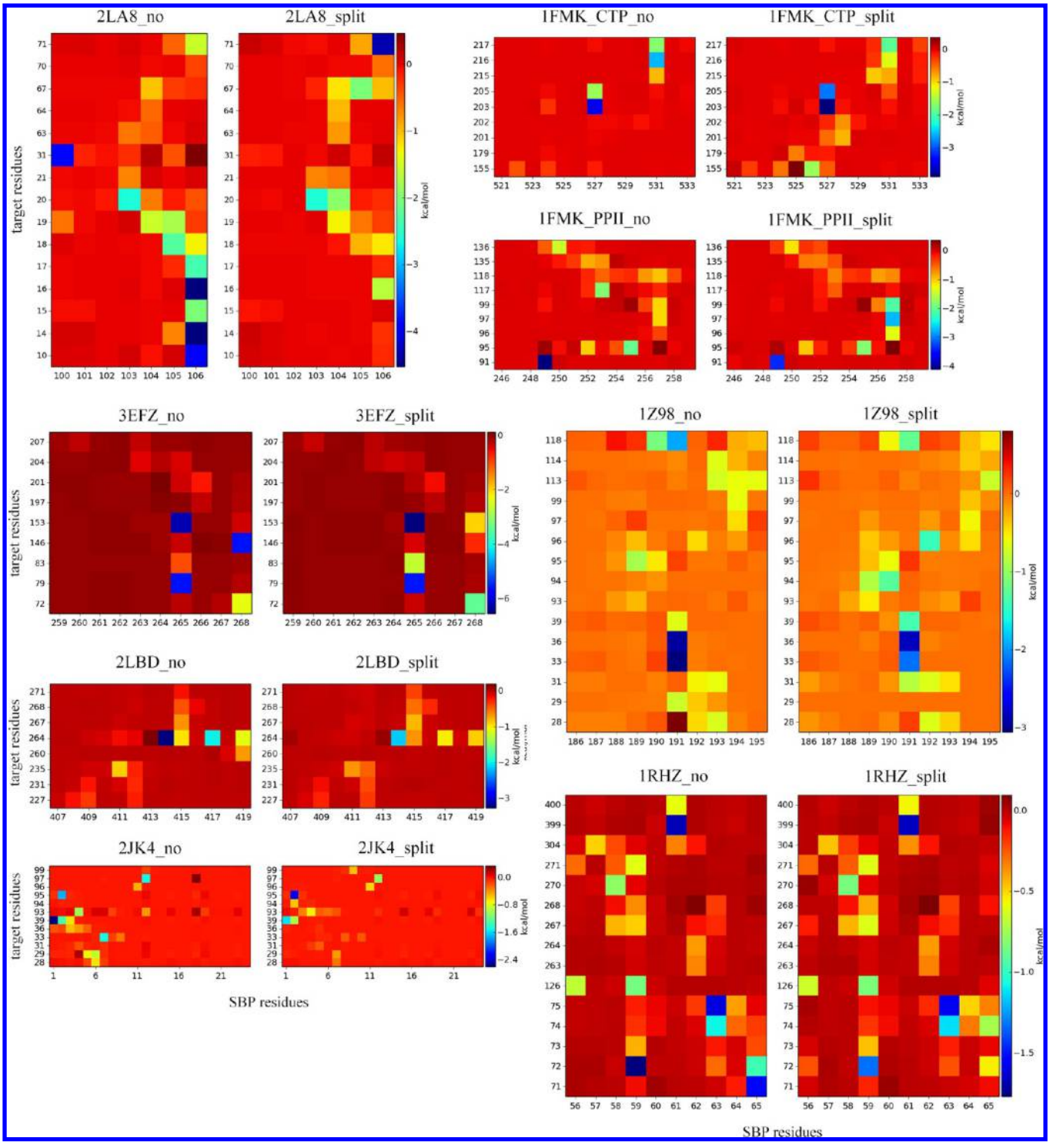


Figure 10. Residue pair interaction energy maps between the SBPs and their target domains for the eight intact SBP−target systems (tagged with "no") and their split peptide−protein complex counterparts (tagged with "split"). The maps were derived from the last 10 ns MD simulations.

"binding-on-swinging" manner to interact with the target weakly and reversibly; this feature makes SBP ideal for mediating biological mechanisms that are required as fast response to stimuli, and which need to be specific while at the same time easily activated and inactivated.

SBP can function as self-regulatory switcher to control the biological behavior of its parent protein. For example, the SecYEβ protein-conducting channel has a small helical "plug" that conducts SBP role and locates at the bottom of extracellular funnel (Figure 11AB).[46] The SecYEβ is commonly dimerized in plasma membrane. However, previous theoretical study demonstrated that only monomeric form of the protein is also biologically active that manipulates the translocation of peptide substances and small proteins from cytoplasm to periplasm by controllably opening and closing the plug (Figure 11C).[47] MD simulations revealed that the plug, no matter it is connected to or split from rest of the protein in sequence, can

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

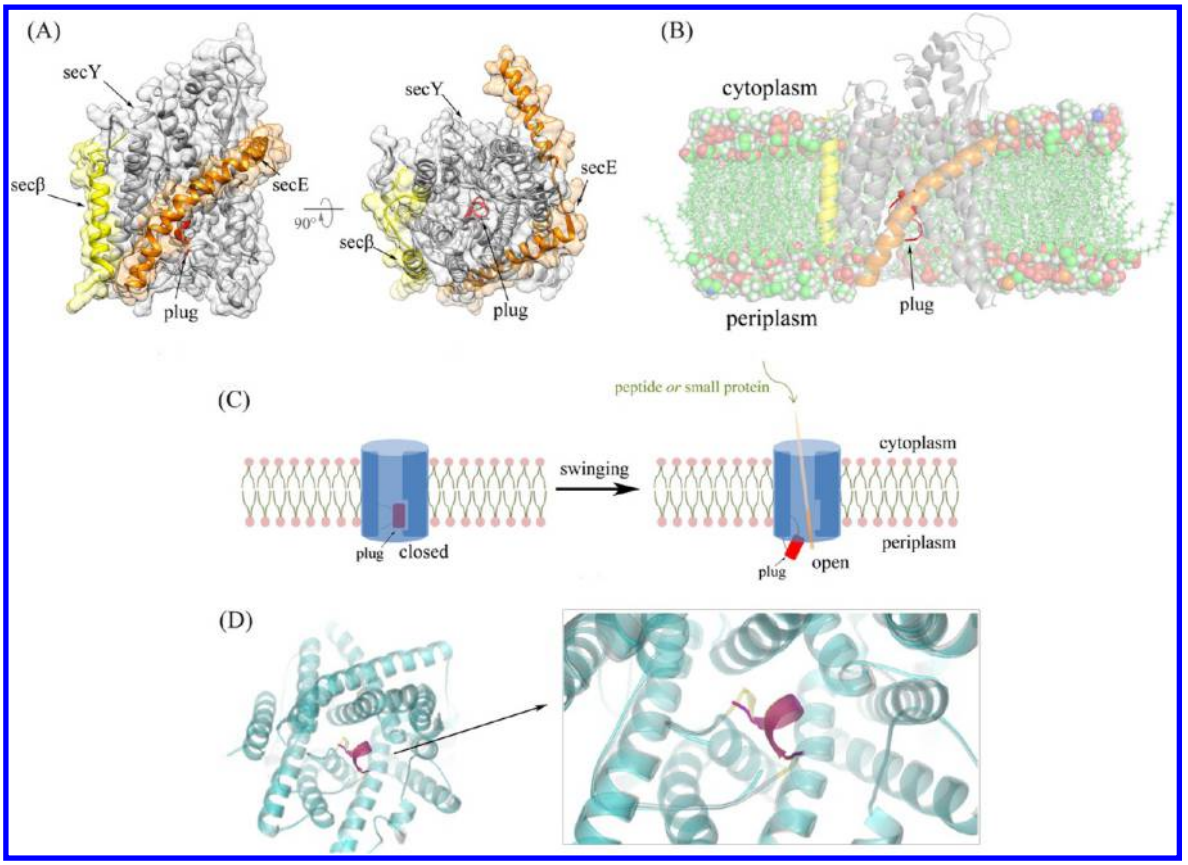


Figure 11. SecYEβ protein-conducting channel and its SBP-mediated biological mechanism. (A) Side (left) and top (right) views of SecYEβ (PDB: 1RHZ). (B) Side view of SecYEβ embedded in plasma membrane and water (not shown) environment. (C) Schematic representation of SBP-mediated SecYEβ switching between closed and open states by swinging a small helical plug. (D) Superposition of intact SBP–target system (cyan) on corresponding split peptide–protein complex (gray), in which the plugs of the two models (shown in "yellow" and "magentas", respectively) are supposed to each other fair well.

interact tightly with a pore ring of SecYEβ to lock the channel in closed state (Figure 11D).

The SBP–target binding/unbinding can be triggered by various events such as protein modification (i.e., phosphonation, methylation, and acetylation), physical condition change (i.e., pH value, temperature, and pressure), and ligand regulation (i.e., binding of small-molecule ligand to the monomeric protein). For instance, the autoinhibition of nonreceptor tyrosine kinase *c*-Src is triggered by the phosophorylation of Tyr527 residue,[13,37] while Ser115 dephosphorylation promotes the interaction of loop D with N-terminal residues in spinach aquaporin SoPIP21 to close the flux of the water channel.[38,39,48]

Some other SBPs like Kon peptide and PPII adopt a competitive inhibition strategy to regulate their parent monomers. Kon peptide fulfills its autoinhibitory function by occupying reversibly in the peptide-binding pocket of PDZ5 domain, thus preventing natural peptide ligands from the pocket and disrupting relevant signal transduction.[15] Moreover, conformational change in target domain can also alter SBP binding affinity. For example, the ligand binding domain (LBD) of human retinoic acid receptor hRARγ is the binding site of diverse hormones;[27] the LBD conformation is changed when bound with all-trans retinoic acid to facilitate SBP helix12 binding (Figure 9E), thus inhibiting hRARγ–DNA recognition and interaction.

4. CONCLUSIONS

In the current work, the SBP concept was for the first time introduced to describe a new biomolecular phenomenon that a peptide segment in monomeric protein fulfills its biological function by dynamically binding to/unbinding from its cognate target in the same monomer. In order to elucidate the molecular mechanism and biological implication underlying SBP–target recognition and association, we systematically performed atomistic MD simulations and post binding energy analysis for eight typical SBP–target systems and their peptide–protein complex counterparts in order to compare between them comprehensively. It was found that the intact SBP–target systems share a high similarity with corresponding split counterparts in structure, energy and dynamics aspects, suggesting that SBP–target interaction is almost a binding behavior. In addition, the polypeptide linker between SBP and target seems to have a marginal effect on the recognition and interaction, implying that folding cannot be completely ignored in the process. We therefore hope the biologically functional SBP described in this study can be established as a new and special biomolecular phenomenon spanning between folding and binding, and their biological function and medicinal significance would be further exploited in the future.

AUTHOR INFORMATION

Corresponding Author

*E-mail: p_zhou@uestc.edu.cn. Tel.: +862883202351.

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329–342

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 31200993), the Young Teacher Doctoral Discipline Fund of Ministry of Education of China (No. 20120185120025), and the New Academic Researcher Award of UESTC.

■ REFERENCES

(1) Petsalaki, E.; Stark, A.; García-Urdiales, E.; Russell, R. B. Accurate Prediction of Peptide Binding Sites on Protein Surfaces. *PLoS Comput. Biol.* **2009**, *5*, e1000335.
(2) Akiva, E.; Friedlander, G.; Itzhaki, Z.; Margalit, H. A Dynamic View of Domain-Motif Interactions. *PLoS Comput. Biol.* **2012**, *8*, e1002341.
(3) Neduva, V.; Russell, R. B. Peptides Mediating Interaction Networks: New Leads at Last. *Curr. Opin. Biotechnol.* **2006**, *17*, 465−471.
(4) Petsalaki, E.; Russell, R. B. Peptide-Mediated Interactions in Biological Systems: New Discoveries and Applications. *Curr. Opin. Biotechnol.* **2008**, *19*, 344−350.
(5) Verschueren, E.; Vanhee, P.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Protein-Peptide Complex Prediction through Fragment Interaction Patterns. *Structure* **2013**, *21*, 789−797.
(6) Zhou, P.; Huang, J.; Tian, F. Specific Noncovalent Interactions at Protein-Ligand Interface: Implications for Rational Drug Design. *Curr. Med. Chem.* **2012**, *19*, 226−238.
(7) Zhou, P.; Wang, C.; Ren, Y.; Yang, C.; Tian, F. Computational Peptidology: A New and Promising Approach to Therapeutic Peptide Design. *Curr. Med. Chem.* **2013**, *20*, 1985−1996.
(8) Von Heijne, G. The Signal Peptide. *J. Membr. Biol.* **1990**, *115*, 195−201.
(9) Suokas, M.; Myllylä, R.; Kellokumpu, S. A Single C-Terminal Peptide Segment Mediates both Membrane Association and Localization of Lysyl Hydroxylase in the Endoplasmic Reticulum. *J. Biol. Chem.* **2000**, *275*, 17863−17868.
(10) Moldaver, D.; Larché, M. Immunotherapy with Peptides. *Allergy* **2011**, *66*, 784−791.
(11) Nguyen, J. T.; Lim, W. A. How Src Exercises Self-Restraint. *Nat. Struct. Mol. Biol.* **1997**, *4*, 256−260.
(12) Gonfloni, S.; Williams, J. C.; Hattula, K.; Weijland, A.; Wierenga, R. K.; Superti-Furga, G. The Role of the Linker between the SH2 Domain and Catalytic Domain in the Regulation and Function of Src. *EMBO J.* **1997**, *16*, 7261−7271.
(13) Boggon, T. J.; Eck, M. J. Structure and Regulation of Src Family Kinases. *Oncogene* **2004**, *23*, 7918−7927.
(14) Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y. C-Terminal Peptide of Fusarium Heterosporum Lipase is Necessary for its Increasing Thermostability. *J. Biochem.* **1998**, *124*, 1124−1129.
(15) Liu, W.; Wen, W.; Wei, Z.; Yu, J.; Ye, F.; Liu, C.-H.; Hardie, R. C.; Zhang, M. The INAD Scaffold is a Dynamic, Redox-Regulated Modulator of Signaling in the Drosophila Eye. *Cell* **2011**, *145*, 1088−1101.
(16) Bayrhuber, M.; Meins, T.; Habeck, M.; Becker, S.; Giller, K.; Villinger, S.; Vonrhein, C.; Griesinger, C.; Zweckstetter, M.; Zeth, K. Structure of the Human Voltage-Dependent Anion Channel. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15370−15375.
(17) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* **2007**, *447*, 1021−1025.
(18) Dogan, J.; Mu, X.; Engström, Å.; Jemth, P. The Transition State Structure for Coupled Binding and Folding of Disordered Protein Domains. *Sci. Rep.−UK* **2013**, *3*, No. 2017.
(19) Rose, P. W.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dimitropoulos, D.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Prlić, A.; Quesada, M. The RCSB Protein Data Bank: New Resources for Research and Education. *Nucleic Acids Res.* **2013**, *41*, D475−D482.
(20) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. Visualizing and Quantifying Molecular Goodness-of-Fit: Small-Probe Contact Dots with Explicit Hydrogen Atoms. *J. Mol. Biol.* **1999**, *285*, 1711−1733.
(21) Zhou, P.; Tian, F.; Shang, Z. 2D Depiction of Nonbonding Interactions for Protein Complexes. *J. Comput. Chem.* **2009**, *30*, 940−951.
(22) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. *AMBER 12*; University of California, San Francisco, 2012; Vol. *1*, p 3.
(23) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side - Chain Torsion Potentials for the Amber FF99sb Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(24) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(25) Andersen, H. C. RATTLE: A "Velocity" Version of the SHAKE Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.
(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(27) Renaud, J.-P.; Rochel, N.; Ruff, M.; Vivat, V.; Chambon, P.; Gronemeyer, H.; Moras, D. Crystal Structure of the RAR-Γ Ligand-Binding Domain Bound to All-Trans Retinoic Acid. *Nature* **1995**, *378*, 681−689.
(28) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(29) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes Database. *Bioinformatics* **2006**, *22*, 623−625.
(30) Skjevik, Å. A.; Madej, B. D.; Walker, R. C.; Teigen, K. LIPID11: A Modular Framework for Lipid Simulations Using Amber. *J. Phys. Chem. B* **2012**, *116*, 11124−11136.
(31) Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.
(32) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(33) Giorgino, T. Computing 1-D Atomic Densities in Macromolecular Simulations: The Density Profile Tool for VMD. *Comput. Phys. Commun.* **2014**, *185*, 317−322.
(34) Kučerka, N.; Tristram-Nagle, S.; Nagle, J. F. Structure of Fully Hydrated Fluid Phase Lipid Bilayers with Monounsaturated Chains. *J. Membr. Biol.* **2006**, *208*, 193−202.
(35) Kučerka, N.; Nieh, M.-P.; Katsaras, J. Fluid Phase Lipid Areas and Bilayer Thicknesses of Commonly Used Phosphatidylcholines as a Function of Temperature. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2011**, *1808*, 2761−2771.
(36) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen - Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(37) Wenqing, X.; Harrison, S.; Eck, M. Three-Dimensional Structure of the Tyrosine Kinase c-Src. *Nature* **1997**, *385*, 595−602.
(38) Törnroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. Structural Mechanism of Plant Aquaporin Gating. *Nature* **2005**, *439*, 688−694.
(39) Nyblom, M.; Frick, A.; Wang, Y.; Ekvall, M.; Hallgren, K.; Hedfalk, K.; Neutze, R.; Tajkhorshid, E.; Törnroth-Horsefield, S.

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342

Structural and Functional Analysis of SoPIP2; 1 Mutants Adds Insight into Plant Aquaporin Gating. *J. Mol. Biol.* **2009**, *387*, 653−668.
(40) Li, W.; Schulman, S.; Boyd, D.; Erlandson, K.; Beckwith, J.; Rapoport, T. A. The Plug Domain of the Secy Protein Stabilizes the Closed State of the Translocation Channel and Maintains a Membrane Seal. *Mol. Cell* **2007**, *26*, 511−521.
(41) Miller, M. L.; Jensen, L. J.; Diella, F.; Jorgensen, C.; Tinti, M.; Li, L.; Hsiung, M.; Parker, S. A.; Bordeaux, J.; Sicheritz-Ponten, T. Linear Motif Atlas for Phosphorylation-Dependent Signaling. *Sci. signal.* **2008**, *1*, ra2.
(42) Stein, A.; Aloy, P. Contextual Specificity in Peptide-Mediated Protein Interactions. *PLoS One* **2008**, *3*, e2524.
(43) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **2010**, *18*, 188−199.
(44) Yu, H.; Zhou, P.; Deng, M.; Shang, Z. Indirect Readout in Protein-Peptide Recognition: A Different Story from Classical Biomolecular Recognition. *J. Chem. Inf. Model.* **2014**, *54*, 2022−2032.
(45) Staneva, I.; Wallin, S. Binding Free Energy Landscape of Domain-Peptide Interactions. *PLoS Comput. Biol.* **2011**, *7*, e1002131.
(46) van den Berg, B.; Clemons, W. M.; Collinson, I.; Modis, Y.; Hartmann, E.; Harrison, S. C.; Rapoport, T. A. X-Ray Structure of a Protein-Conducting Channel. *Nature* **2003**, *427*, 36−44.
(47) Gumbart, J.; Schulten, K. Molecular Dynamics Studies of the Archaeal Translocon. *Biophys. J.* **2006**, *90*, 2356−2367.
(48) Johansson, I.; Karlsson, M.; Shukla, V. K.; Chrispeels, M. J.; Larsson, C.; Kjellbom, P. Water Transport Activity of the Plasma Membrane Aquaporin PM28A is Regulated by Phosphorylation. *Plant Cell Online* **1998**, *10*, 451−459.
(49) Brokx, S. J.; Wernimont, A. K.; Dong, A.; Wasney, G. A.; Lin, Y.-H.; Lew, J.; Vedadi, M.; Lee, W. H.; Hui, R. Characterization of 14−3-3 Proteins from Cryptosporidium Parvum. *PLoS One* **2011**, *6*, e14827.

DOI: 10.1021/ci500522v
J. Chem. Inf. Model. 2015, 55, 329−342