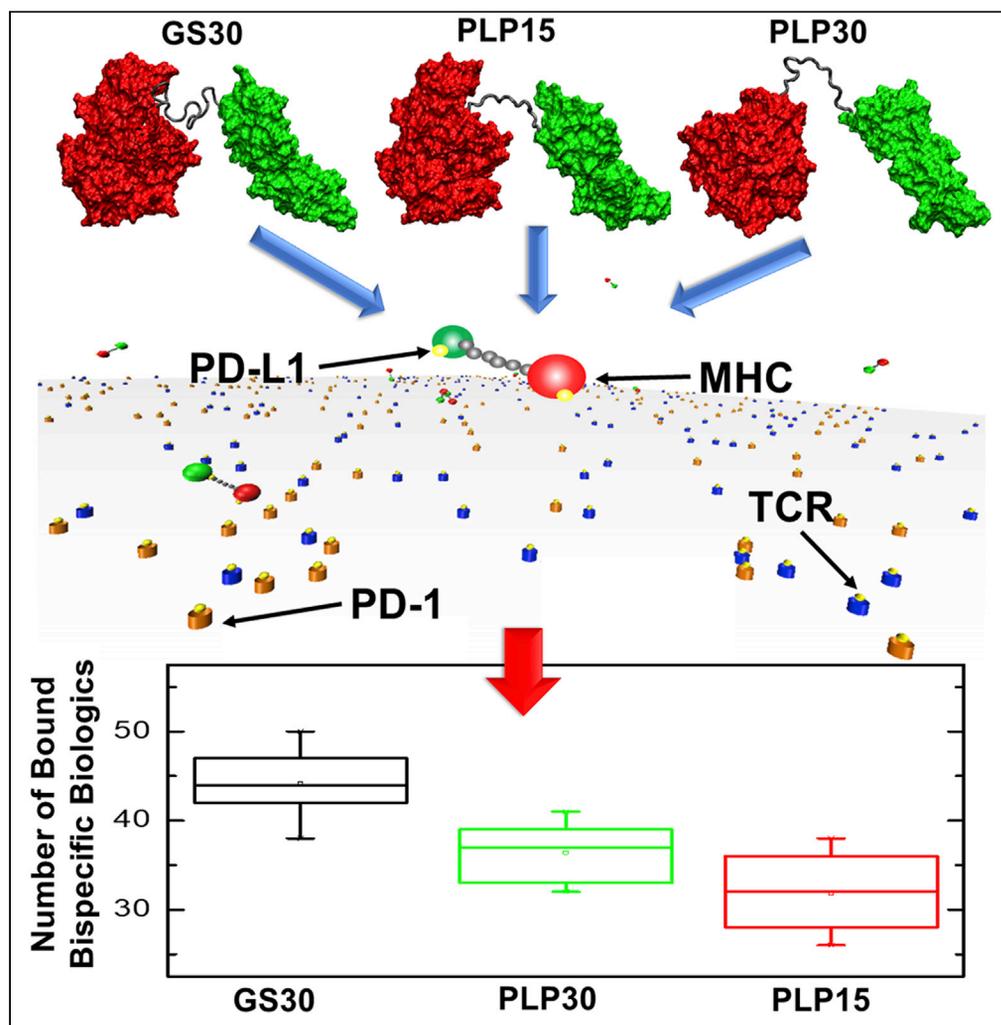# iScience

Article

Understanding the Targeting Mechanisms of Multi-Specific Biologics in Immunotherapy with Multiscale Modeling

Zhaoqian Su, Bo Wang, Steven C. Almo, Yinghao Wu

yinghao.wu@einsteinmed.org

HIGHLIGHTS

Two proteins are connected by different linkers as a model of bispecific biologics

Conformational dynamics of biologics are captured by microsecond MD simulations

Coarse-grained simulations are used to test binding between biologics and receptors

Biologics with long and flexible linkers are more efficient in targeting receptors

Article

Understanding the Targeting Mechanisms of Multi-Specific Biologics in Immunotherapy with Multiscale Modeling

Zhaoqian Su,[1] Bo Wang,[1] Steven C. Almo,[2,3] and Yinghao Wu[1,4,*]

SUMMARY

Immunotherapeutics are frequently associated with adverse side effects due to the elicitation of global immune modulation. To lower the risk of these side effects, recombinant DNA technology is employed to enhance the selectivity of cell targeting by genetically fusing different biomolecules, yielding new species referred to as multi-specific biologics. The design of new multi-specific biologics is a central challenge for the realization of new immunotherapies. To understand the molecular determinants responsible for regulating the binding between multi-specific biologics and surface-bound membrane receptors, we developed a multiscale computational framework that integrates various simulation approaches covering different timescales and spatial resolutions. Our model system of multi-specific biologics contains two natural ligands of immune receptors, which are covalently tethered by a peptide linker. Using this method, a number of interesting features of multi-specific biologics were identified. Our study therefore provides an important strategy to design the next-generation biologics for immunotherapy.

INTRODUCTION

The past decade has witnessed the rise of immunotherapies for the treatment of cancers, autoimmune diseases, and infectious diseases through the strategic modulation of the patient's immune system (Yang, 2015; Gun et al., 2019; Kruger et al., 2019). Traditionally, immunotherapies have relied on high-affinity monoclonal antibodies and Fc fusion proteins to selectively engage presented cell surfaces or soluble signaling molecules, resulting in clinically beneficial immune modulation. In the cases of cancers and infectious diseases, these agents can block inhibitory signals or enhance stimulatory signals, resulting in global immune enhancement, albeit with the risk of adverse autoimmune effects. For example, function-blocking antibodies (e.g., nivolumab and pembrolizumab) that recognize the inhibitory receptor (e.g., programmed cell death-1 [PD-1]) stimulate global immunity by competing with the programmed death-ligand 1 (PD-L1) for PD-1 engagement, so they are cutting-edge treatments for a range of malignancies, including melanoma, renal cell cancer, and non-small cell lung cancer. Analogously, immunotherapeutic strategies that globally reduce immune responsiveness can effectively treat autoimmune diseases; however, in this case, patients are at increased risk from a wide range of opportunistic infections and malignancies that would otherwise be managed by an unperturbed immune response. For example, Orencia is a soluble Fc-fusion of cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) that competes with CD28 for binding the B7 ligands, inhibiting the CD28-associated stimulatory pathways (Pardoll, 2012). Blocking CD28 stimulation causes global immune suppression, making Orencia a leading treatment for autoimmune diseases, including rheumatoid arthritis (Bluestone et al., 2006). However, although these biologics activate receptors on targeted cells, they simultaneously bind to the same type of receptors on the surfaces of other cells, directing a broad immune response (Michot et al., 2016), which can lead to many harmful side effects (Seror and Mariette, 2017).

One strategy to minimize these side effects is developing chimeric constructs in which biomolecules with distinct functions are covalently tethered to yield multi-specific biologics (Baldo, 2015; Padte et al., 2018; Thakur and Lum, 2016). For instance, bispecific T cell engagers (BiTEs) comprise two protein fragments connected by a linker (Huehls et al., 2015): one fragment binds to a T cell-specific molecule, such as CD3, whereas the other binds to an antigen on a tumor cell. The bispecific nature of BiTEs creates specific

[1]Department of Systems and Computational Biology, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Avenue, Bronx, NY 10461, USA

[2]Department of Biochemistry, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Avenue, Bronx, NY 10461, USA

[3]Department of Physiology and Biophysics, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Avenue, Bronx, NY 10461, USA

[4]Lead Contact

*Correspondence:
yinghao.wu@einsteinmed.org

https://doi.org/10.1016/j.isci.2020.101835

localization and physical coupling between cytotoxic T cells and tumor cells, which kills tumor cells. A variety of other bispecific fusion proteins have also been reported. For example, Way et al. designed molecules with enhanced selectivity composed of antibody fragments for cellular targeting tethered via peptides or nucleic acid linkers to signaling molecules, such as cytokines and erythropoietin (Way et al., 2014). Tavernier et al. similarly coupled affinity domains (i.e., single-domain camelid nanobodies) to cytokines (e.g., interferon [IFN] α and tumor necrosis factor [TNF]) via GLY-SER linkers, with promising results in preclinical murine tumor models (Uze and Tavernier, 2015). Finally, Quayle et al. reported a platform on which single-chain peptide-major histocompatibility complex (MHC) constructs are covalently linked to a variety of costimulatory, co-inhibitory, and cytokine molecules (e.g., interleukin [IL]-2) (Quayle et al., 2020). The sc-pMHC domain acts as an "address" to target specific T cell clones for the delivery of a range of co-modulatory domains, resulting in clonal-selective T cell modulation, thus eliminating the side effects of current immunotherapies that elicit global immune modulation. This platform has demonstrated efficacy in a preclinical tumor model and synergy with anti-PD-1 treatment. As illustrated by these examples, the development of new multi-specific biologics is an exceptionally active area for the realization of new immunotherapies that can enhance treatment efficacy and reduce the risk of side effects (Weidle et al., 2013). A remaining challenge, however, is the need for additional strategies to predict, engineer, and control the specificity of biologics in terms of receptor targeting and biological function (Dimasi et al., 2009).

Computational approaches have become a cornerstone of modern drug development (Kontoyianni, 2017; Sliwoski et al., 2014) because, compared with experimental approaches, *in silico* modeling is much less labor-intensive or time-consuming and enables large-scale evaluations of chemical space and design features for subsequent prioritization and experimental validation (Carnero, 2006). Most applications of computational approaches in drug design have focused on and made significant contributions to the development of small chemical compounds (Katsila et al., 2016), peptide-based biomolecules (Farhadi and Hashemian, 2018), and therapeutic antibodies (Choong et al., 2017). In contrast, computational efforts involving multi-specific biologics have been hindered by their complexity and the limitations of the most common computational approaches, such as protein-protein docking (Kaczor et al., 2018) and molecular dynamic (MD) simulations (Karplus and Petsko, 1990). For instance, the benchmark results from computational docking show that the binding interfaces of protein-protein interactions are not only more diverse and more challenging to predict than the binding sites of small molecules on protein surfaces (Villoutreix et al., 2008) but also their binding rates and binding affinities are more difficult to estimate (Brender and Zhang, 2015; Xiong et al., 2017; Qin et al., 2011). Moreover, although MD simulations have become a mature method to study the atomic details associated with protein dynamics (Pan et al., 2019; Plattner et al., 2017), the intense computational requirements generally precludes them from being applied to systems over biologically relevant timescales (Im et al., 2016). Finally, it is critically important but experimentally demanding to trace the binding kinetics between biologics and their protein targets within the cellular environments on the plasma membrane. This requirement is also a significant challenge for most current simulation techniques due to the size of these macromolecules and the heterogeneity of their cellular environments (Ramis-Conde et al., 2008; Chakrabarti et al., 2012; Krobath et al., 2009).

To overcome these limitations, we developed a new multiscale framework to characterize the dynamics of binding between multi-specific biologics and their receptors on the plasma membranes of mammalian cells. Two natural ligands of immune receptors, MHC and PD-L1, are artificially connected by a peptide linker as a test model of multi-specific biologics. Analogous to the natural response, in which the modulation of a naive T cell requires both primary engagement between a T cell receptor (TCR) and MHC and a secondary coregulatory signal (Figure 1A), the bispecific MHC-PD-L1 conjugate will selectively target and modulate only T cells expressing the corresponding TCR on their surfaces (Figure 1B). Based on the hypothesis that the interactions between ligands and receptors can be modulated by the conformational dynamics of linkers joining them (Chen et al., 2013), we further designed virtual fusion proteins with three linkers that represent various lengths and degrees of flexibility. The impacts of these linkers on the dynamics of the biologics have been captured by microsecond-level MD simulations. Moreover, we applied a kinetic Monte Carlo (kMC) algorithm to calculate the rates of TCR/MHC and PD-L1/PD-1 binding and evaluated the impact of protein variants with various affinities on these binding processes. The MD and kMC results were integrated into a low-resolution model to simulate the kinetics of a large system containing hundreds of biologics and cell-surface receptors on a timescale of 10^8 nanoseconds. With this multiscale simulation method, we found that cells expressing high levels of cognate TCR were much more sensitive to ligand concentration than cells with low expression levels of TCR, indicating that surface density is
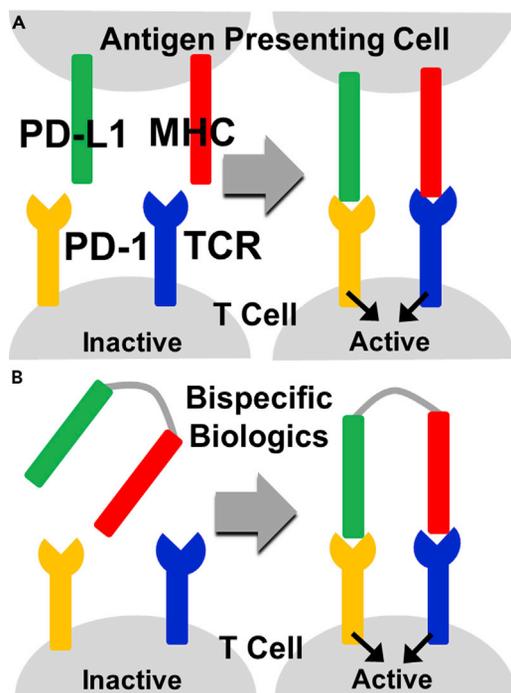
Figure 1. The Biological Background of Binding between Bispecific Biologics and Cell Surface Receptors

(A and B) The traditional two-signal hypothesis for T cell activation (A) requires specific engagement between TCR and MHC-epitope molecules followed by a secondary modulatory message through the engagement between costimulatory molecules, such as PD-L1 and its receptor PD-1. This combinatory signal can be mimicked by an artificial fusion protein containing MHC-epitope and modulatory domains with a peptide linker (B). This fusion protein is used as a specific model of multi-specific biologics to test our multiscale simulations.

an important parameter that can be exploited for selective cell targeting. Interestingly, our simulations also revealed a negative cooperativity between TCR/MHC and PD-L1/PD-1 binding in the biologics. Mutations that strengthen the PD-L1/PD-1 binding can weaken the interaction between TCR and MHC, even if the TCR/MHC binding interface is not changed. Given the topological design of the biologics in which multiple functional modules are covalently linked together, we also observed spatial patterns in which membrane receptors were assembled into small clusters on cell surfaces. Finally, we found that longer, more flexible linkers were more efficient at targeting cell-surface receptors within the system we examined. Taken together, the computational strategy adopted in this study sheds light on the development of new biopharmaceutics and future immunotherapeutic strategies.

RESULTS

The results from the multiscale simulations are organized as follows. In the first section, we describe how we constructed the rigid body (RB)-based model and tested the general properties of binding between bispecific fusion proteins and their receptors. In the second section, we characterize the conformational dynamics of the bispecific fusion protein with all-atom MD simulations. The third section calculates the association rates of binding between wild-type and mutated bispecific fusion proteins and their receptors with residue-based kMC. Finally, we integrate the MD and kMC results into the rigid body-based model in the last section.

Construct a Rigid Body (RB)-Based Model for Cell-Surface Targeting of Multi-Specific Biologics

Our test model of multi-specific biologics is a soluble fusion protein composed of two functional modules, MHC and PD-L1, connected by a genetically encoded peptide linker. As described in the Methods section, we constructed an RB-based model to simulate the binding between the two components of the fusion protein and their cognate cell-surface receptors (i.e., TCR and PD-1, respectively) (Chen et al., 2017). Specifically, the plasma membrane is represented by a square plane at the bottom of a three-dimensional simulation box. The area of the square is 1 μm^2, and the height of the simulation box is 250 nm. Both TCR (blue in Figure 2A) and PD-1 receptors (orange in Figure 2A) are represented as rigid cylinders randomly distributed on the surface. The height and radius of each cylinder are 8 and 4 nm, respectively, which are comparable to the size of real protein structures. Single binding sites (yellow dots in Figure 2A) are placed on top of each receptor, through which it binds to the corresponding ligand in the fusion
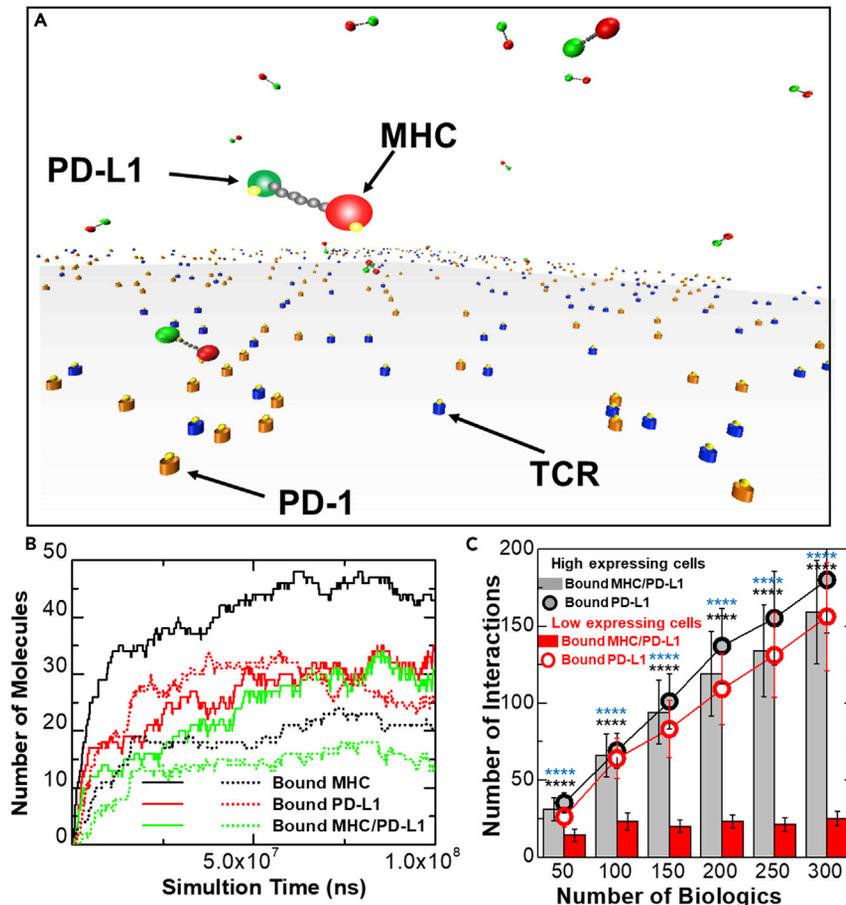
Figure 2. The Rigid Body-Based Model and General Simulation Results

(A–C) An initial configuration of our RB-based model was set up to simulate the binding between cell-surface receptors and multi-specific biologics (A). The plasma membrane is represented by a square plane at the bottom of a three-dimensional simulation box. Receptors are randomly distributed on the membrane, whereas biologics are located in the space above. Each functional module in the biologics is simplified as a spherical RB. Modules are connected by a flexible linker, and receptors are represented by RBs of cylinders. Two types of cells were exposed to the biologics. The specific immune cell that aims to be targeted by the biologics expresses cognate TCR highly, whereas cells in the control model contain low levels of TCR. The simulation results of both cell types are plotted in (B) as a function of time. The results of the high-expressing cells are shown with solid curves, whereas the results of the low-expressing cells are shown with dashed curves. The statistical fluctuations after these systems reached equilibrium were calculated. Specifically, the means and standard errors of the curves within the last 5×10^7 ns are listed as follows: 45.31 $\pm$ 1.55 (black solid), 30.86 $\pm$ 1.69 (red solid), 29.17 $\pm$ 1.92 (green solid), 21.24 $\pm$ 1.60 (black dashed), 28.28 $\pm$ 2.86 (red dashed), 15.82 $\pm$ 1.36 (green dashed). We also changed the numbers of biologics in the extracellular region from 50 to 300. The results are summarized in (C) for both high-expressing (gray) and low-expressing cells (red). The numbers of biologics that formed complexes with both TCR and PD-1 are plotted as bars, whereas the numbers of biologics in which the PD-L1 module is bound to PD-1 are plotted as circles. The means and standard errors of the gray circles from 50 to 300 are listed as follows: 35 $\pm$ 6.5, 69 $\pm$ 11.4, 101 $\pm$ 18, 137 $\pm$ 24.5, 155 $\pm$ 30.6, and 180 $\pm$ 34.7. The means and standard errors of the gray bars from 50 to 300 are listed as follows: 31 $\pm$ 7.3, 66 $\pm$ 14, 94 $\pm$ 20.7, 119 $\pm$ 27.4, 134 $\pm$ 29.7, and 159 $\pm$ 33.7. The means and standard errors of the red circles from 50 to 300 are listed as follows: 26 $\pm$ 6.5, 64 $\pm$ 13, 83 $\pm$ 18.7, 109 $\pm$ 23.1, 131 $\pm$ 27.4, and 156 $\pm$ 35. The means and standard errors of the red bars from 50 to 300 are listed as follows: 14 $\pm$ 4.1, 23 $\pm$ 5.5, 20 $\pm$ 4, 23 $\pm$ 4.3, 21 $\pm$ 4.3, and 25 $\pm$ 4.6. t tests were further performed between bars, as well as between circles at different numbers of biologics. The calculated p values less than 0.0001 are given four asterisks. The p values between bars are marked as black asterisks, whereas the p values between circles are marked as blue asterisks.

protein/biologic. The extracellular region is represented by the space above the plasma membrane; several multi-specific biologics are located in this three-dimensional volume. Each functional module in the biologic is simplified as a rigid spherical with a given radius (4 nm). The two RB modules within a single fusion protein, representing MHC (red in Figure 2A) and PD-L1 (green in Figure 2A), respectively, are

tethered together by peptide linkers (gray dots in Figure 2A) possessing variable degrees of flexibility. On the surfaces of MHC and PD-L1, we also assign single binding sites (yellow dots in Figure 2A) for interaction with their corresponding receptors. Following the initial configuration, the movements of receptors and biologics and their binding were simulated by a diffusion-reaction algorithm until the system reached equilibrium. The Methods section describes the detailed simulation process. A sensitivity analysis on the simulation with different binding affinity values was first carried out, the results of which the Supplemental Information and Figure S3 summarizes.

Based on the sensitivity analysis, we further changed the density of the membrane receptors on the cell surfaces. Considering that membrane receptors can exhibit many expression profiles in different cells, this study specifically compares two models. The first model expressed TCR highly, representing the specific immune cells targeted by the fusion protein; in contrast, the control model hardly expressed TCR. Specifically, we assumed that the expression level in the control model was lower than the expression level of the first model by one order of magnitude. As a result, 300 copies of TCR were distributed on the surface of the high-expressing cells, giving a surface TCR density of ~10^2 molecules per μm^2, which was within the range of experimental observation for T cells (Brameshuber et al., 2018). In contrast, the low-expressing cell had 30 copies of TCR on its surface. The surfaces of both cells also contained 300 copies of PD-1 receptors (Li et al., 2017). In both models, 50 copies of fusion protein were exposed to the receptor-embedded cell surfaces.

Figure 2B shows the simulation results of these two systems. The numbers of molecules that bound to receptors on the surfaces of the high-expressing cells are plotted by solid curves as a function of simulation time, whereas the numbers of molecules that bound to receptors on the surfaces of the low-expressing cells are plotted by dashed curves. The numbers of bound MHC and PD-L1 in both systems are plotted by curves in black and red, respectively. The curves in green show the temporal changes of the fusion proteins' functional modules binding to their corresponding receptors. The figure suggests that MHC modules in almost all biologics-targeted cells expressing high levels of TCR by the end of the simulation. In contrast, less than half the MHC interacted with TCR in the low-expressing cells. We also compared the interaction dynamics between PD-1 and PD-L1 in the high-expressing cells with their interactions in the low-expressing cells, and interestingly, the number of bound PD-L1 in the low-expressing cells increased faster at the beginning of the simulation than in the high-expressing cells. As the simulation continued, we observed more complexes forming between PD-1 and PD-L1 in the high-expressing cells until they had more than the low-expressing ones.

We suggest that this dynamic behavior can be explained as follows. As an MHC module bound to a TCR, the entire biologics diffused with the receptor on the cell surface. This constraint gave the unbound PD-L1 module in the same molecule greater accessibility to search the space proximal to the surface area, raising the probability that it would encounter a vacant PD-1 receptor. In other words, the higher number of bound MHC modules raised the local concentration of the biologics above the membrane surface of the cell that expressed high levels of TCR, promoting the binding of PD-L1. Consequently, more interactions between PD-L1 and PD-1 occurred in the high-expressing cells than in the low-expressing ones, although the expression levels of PD-1 in both were the same. Finally, we counted the number of biologics in which both MHC and PD-L1 bound to receptors. Because the functional response of T cells needs the engagement of both TCR and coregulatory receptors, the number of biologics with both bound MHC and PD-L1 on cells is the variable directly related to the biological outcome of cell signaling. As a result, Figure 2B shows that the number of biologics with both bound MHC and PD-L1 is almost twice as high in the high-expressing cells than the low-expressing cells, indicating that multi-specific biologics can much more effectively bind to the surfaces of their targeted immune cells.

We also tested the concentration dependence of multi-specific biologics in cell-surface targeting. Specifically, the numbers of biologics in the extracellular region rose from 50 to 300 in simulations of both high- and low-expressing cells. We counted how many biologics formed complexes with both TCR and PD-1 at the end of each simulation; Figure 2C summarizes these results as histograms. The number of biologics with both MHC and PD-L1 engaged under different concentrations are plotted as gray bars for the high-expressing cells and red bars for the low-expressing cells. The figure shows that many more biologics bound to both TCR and PD-1 on the surfaces of the high-expressing cells as the extracellular concentration of biologics increased. In contrast, due to the limited number of TCR on the surface of the low-expressing

cells, there were almost no changes in the numbers of biologics with both bound MHC and PD-L1 as the concentration increased, suggesting that the immune cells with a specific type of TCR were much more sensitive to changes in concentration than other cells. In other words, the high specificity of T cell surface targeting can be achieved more easily under high concentrations of multi-specific biologics.

We also plotted the numbers of interactions between PD-L1 and PD-1 in Figure 2C. The figure shows that the number of these interactions increased as the concentrations of extracellular biologics rose. More interestingly, we always found more interactions in the high-expressing cells (gray circles) than the low-expressing ones (red circles), although there was no difference in the surface density of PD-1 or the binding affinity between PD-1 and PD-L1. The only possible reason for this difference is the expression level of TCR. While the high expression of one receptor on the cell surface can recruit more bispecific fusion proteins to the membrane-proximal area, the other vacant module in the biologics is likely to encounter its receptor. Our observations, therefore, suggest that the expression level of one receptor can affect the binding of the other receptor to the bispecific fusion proteins.

In summary, we explored the binding of specifically designed biologics to the surfaces of various cell types by implementing an RB-based model coupled with a diffusion-reaction simulation. We demonstrated the possibility that these biologics can target only a preselected type of immune cells, providing a mechanism-based rationale for using fusion proteins to minimize the potential side effects of immunotherapy. The following studies systematically analyze the detailed binding and dynamic parameters in this RB-based model with higher-resolution simulations.

Characterize the Conformational Dynamics of the Bispecific Fusion Proteins

The polypeptide linker between two contiguous domains controls the conformational flexibility of a multi-domain protein and modulates its functions by restraining inter-domain motions and orientations (Bagowski et al., 2010; Vogel et al., 2004). In biopharmaceutics, multi-domain proteins are synthesized by genetically fusing therapeutic modules using recombinant DNA technology (Berger et al., 2015). The interplay among various protein modules in these biologics is significantly impacted by the compositions, lengths, and structures of their linkers. The direct fusion of protein modules without a linker often impairs bioactivities due to steric incompatibilities that interfere with the binding function (Bai et al., 2005; Bai and Shen, 2006). Therefore, we hypothesize that the linker segment joining MHC and PD-L1 in our system also plays an important role in regulating the function of the bispecific fusion proteins. Specifically, we propose that the linker can control the receptor binding of the fusion protein by modulating its intramolecular conformational dynamics. To evaluate this hypothesis, we designed three types of linkers of different lengths and sequences. The first is called GS30, which has a total length of 30 amino acids. It contains six repeats of the sequence composed of four glycines followed by a serine (i.e., $(Gly_4Ser)_6$). This linker is thought to possess intrinsic disorder due to the flexibility of glycine (Van Rosmalen et al., 2017). The second and third linkers, PLP15 and PLP30, contain 15 and 30 consecutive prolines, respectively. In contrast to the flexible linker GS30, these two linkers are considered more rigid and have well-defined secondary structures, and their initial structures were modeled by following the standard conformation of the left-handed poly-L-proline helix (Adzhubei and Sternberg, 1993). The Methods section describes the specifics of constructing the structural models of the fusion proteins and their linkers, and Figures 3A–3C show their initial structures.

The all-atom MD simulations of these fusion proteins were performed on an Anton 2 supercomputer; the Methods section summarizes the detailed protocols of system preparation and simulation setup. A 2.5-ms-long trajectory was generated to sample the conformational space for each of the three systems. Snapshots taken along the simulation trajectory of the protein with linker GS30 were plotted from Figures 3D–3G as a demonstrating system. These figures show that whereas the individual tertiary structures of the MHC and PD-L1 modules were preserved during the simulation, the GS30 linker explored many configurations, from compact to extended conformations. As a result, there were large changes in the spatial separation between MHC and PD-L1, and their relative orientations also varied considerably. A conformational ensemble was constructed based on the MD simulation results for each fusion protein. Along the simulation trajectory, the structure of each protein was captured every 1 ns. These structures were superimposed onto the initial conformation by RB superposition, and the average position and orientation of MHC and PD-L1 in the fusion protein were derived from the structural ensemble. The distances between the centers of mass in the average positions of MHC and PD-L1 were 7.7, 7.6, and 9.7 nm for proteins with the linkers
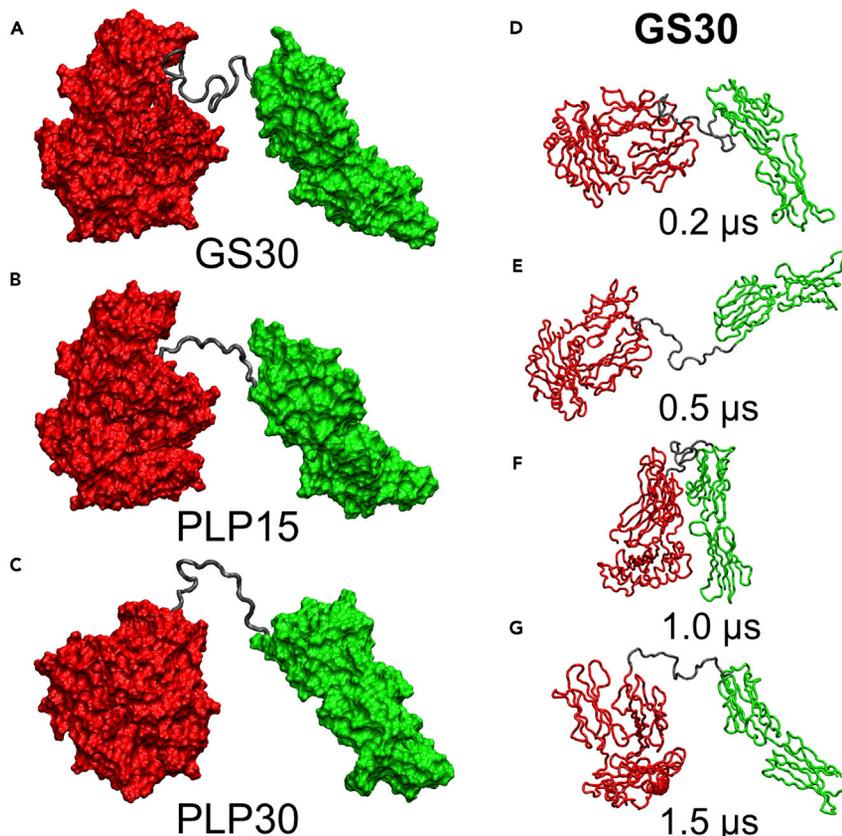
Figure 3. The Initial Structure of Bispecific Biologics with Different Domain Linkers and Selected Snapshots from Molecular Dynamics Simulations

(A–G) In our test model of multi-specific biologics, we computationally fused two functionally independent protein modules with three types of peptide linkers. The structural models of these fusion proteins are shown here for the linkers GS30 (A), PLP15 (B), and PLP30 (C). One module of the fusion proteins is MHC, which is shown in red, whereas the other is PD-L1, which is shown in green. The linkers connecting these two modules are highlighted in gray. A 2.5-ms-long trajectory was generated by the Anton supercomputer to sample the conformational space for each of the three systems. Some representative snapshots are selected from the simulation of linker GS30 at 0.2 μs (D), 0.5 μs (E), 1.0 μs (F), and 1.5 μs (G).

GS30, PLP15, and PLP30, respectively. We attribute the differences in the average lengths between PLP15 and PLP30 to the number of residues in these two linkers, whereas the differences between GS30 and PLP30 are due to their inherent internal flexibility (i.e., the glycine-rich linker can explore more conformations and thus exhibit a shorter end-to-end distance).

For each structure in the ensemble, we also calculated the deviation of each functional module from its average conformation. Figure 4 plots the distributions of these deviations, showing the probability that a given magnitude of deviation is observed in the system. Figure 4A shows the probability distributions of the distance deviation from the average position of MHC, and the distributions were averaged over all three translational degrees of freedom. Similarly, Figure 4B shows the average distance deviation of PD-L1 over three translational degrees of freedom. Both figures suggest that all deviations form extended distributions even for the linkers of poly-prolines, indicating that these linkers are less rigid than expected. This observation is consistent with recent experimental measurements (Ruggiero et al., 2016). More specifically, the distributions of the protein with the linker GS30 (black squares) are wider than linkers PLP15 (red circles) and PLP30 (green triangles), whereas the distributions of the linker PLP15 have the narrowest distribution. Moreover, both linkers PLP15 and PLP30 form normal distribution, whereas the distribution of the GS30 linker is bimodal, probably resulting from the fact that the linker is intrinsically disordered. In addition to the translational deviations, Figures 4C and 4D show the probability distributions of deviation from the average orientation of MHC and PD-L1, respectively. The distributions were averaged over all three
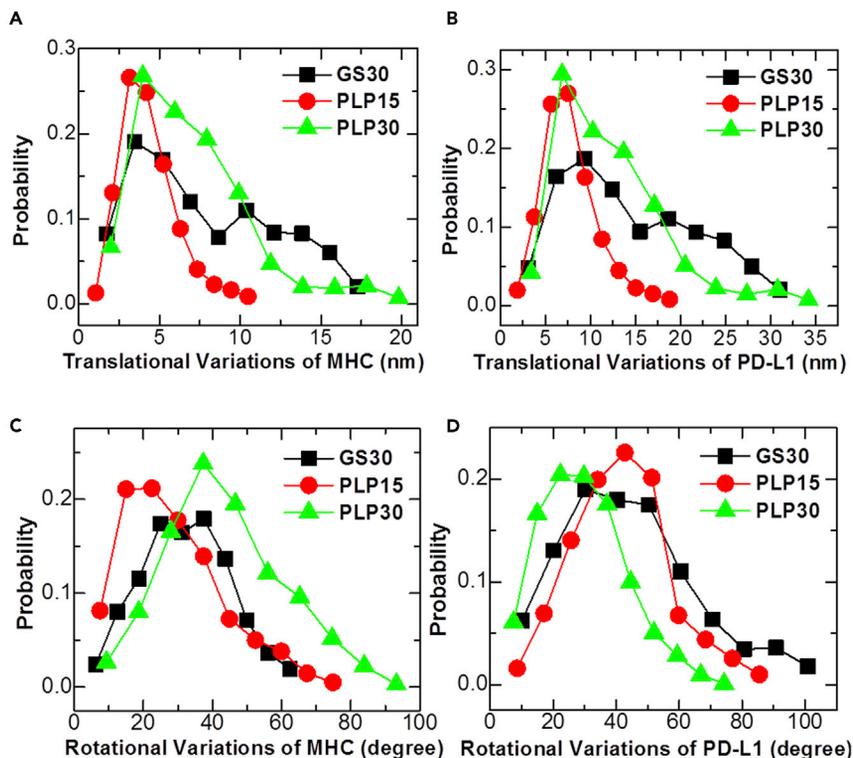
Figure 4. The Conformational Dynamics of the Bispecific Biologics Obtained from Molecular Dynamics Simulations

(A–D) Conformational ensembles were constructed based on the MD simulation results for all three fusion proteins. We calculated the deviations of both functional modules in the fusion proteins from their average conformations and compared the differences in the distributions of these deviations between different linkers. The probability distributions of the distance deviation from the average position of MHC are shown in (A), those of the distance deviation from the average position of PD-L1 in (B), those of the rotational deviation from the average orientation of MHC in (C), and those of the rotational deviation from the average orientation of PD-L1 in (D).

rotational degrees of freedom. Interestingly, the results show that the distributions of GS30 are not wider than those of the other two linkers, different from the distributions of translational deviations. This is due to the fact that the rotations of each functional module in a fusion protein are largely controlled by the covalent bonds that directly connect the functional module to the linkers and are thus less sensitive to linker flexibility. Comparing Figure 4C with Figure 4D, it seems that the rotational variations of PD-L1 are slightly larger than MHC. We speculate that this difference in the distribution of rotational deviation is caused by the molecular weight of MHC being larger than PD-L1. Smaller proteins tend to diffuse more easily than larger proteins and therefore result in higher conformational fluctuations. Importantly, these probability distributions will be used in the RB-based model to adjust the spatial separation and orientation between MHC and PD-L1 within each simulation step, so that the effect of different linkers on receptor binding can be taken into account. The algorithmic details are described in the Methods section.

Estimate the Binding between Functional Modules in Biologics and Their Receptors

Although MHC and PD-L1 are connected together as a single unit, in our test system, they functioned independently by interacting with their corresponding receptors. The association rates between these two ligands and their receptors are estimated by residue-based kMC simulations; the Methods section describes the detailed simulation algorithm. In practice, two pairs of ligand-receptor interactions were simulated separately. The human B7 TCR and MHC class I molecule HLA-A 0201 with a viral peptide TAX (sequence LLFGYPVYV) was used as a specific model system to study the interaction between TCR and MHC (Ding et al., 1998). The atomic coordinates of this complex are taken from the PDB id 1BD2. At the beginning of the simulation, the two proteins were coarse-grained into a residue-based representation and randomly positioned such that the distance between their binding interfaces was kept under a
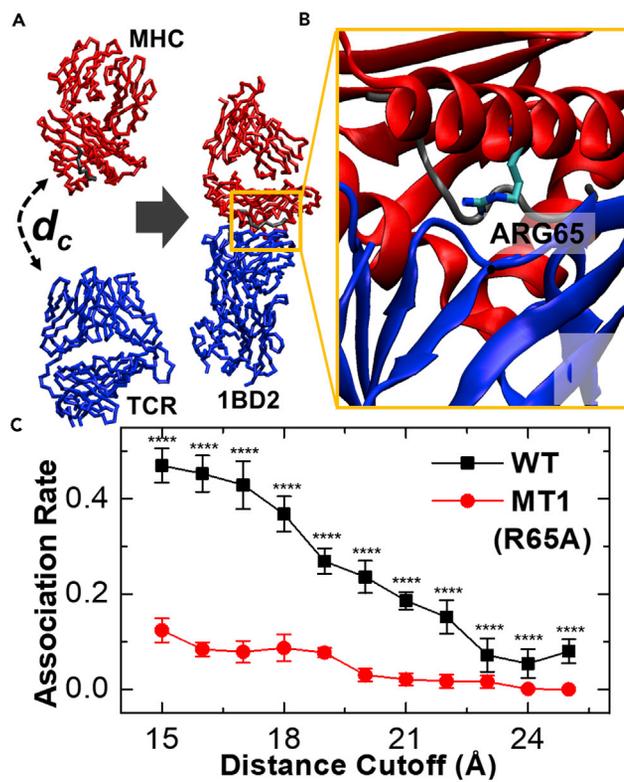
Figure 5. Estimation of Association Rates for Binding between TCR and MHC

(A–C) We simulated the association between TCR and MHC with a residue-based kinetic Monte Carlo algorithm (A). The human TCR B7 and MHC class I molecule HLA-A 0201 with a viral peptide TAX (sequence LLFGYPVYV) was used as a test system (PDB id 1BD2). Multiple simulation trajectories were started from different values of the distance cutoff d_c. We counted how many encounter complexes were formed among these trajectories. We also tested the effect of mutant R65A on the association. The residue ARG65 is located directly on the binding interface between TCR and MHC (B). The calculated frequencies of association at various distance cutoffs were plotted in (C) with error bars for both wild-type (black) and mutant (red). Consistent with the experiment, our simulation results indicate that the mutation of a single residue on the surface of MHC can significantly affect its binding with TCR. The calculated means and standard errors of the association rates for the wild-type from 15 to 25 Å are listed as follows: 0.47 ± 0.036, 0.453 ± 0.038, 0.429 ± 0.05, 0.368 ± 0.037, 0.269 ± 0.027, 0.236 ± 0.034, 0.186 ± 0.018, 0.152 ± 0.035, 0.072 ± 0.035, 0.054 ± 0.03, and 0.08 ± 0.025. The calculated means and standard errors of the association rates for the mutant from 15 to 25 Å are listed as follows: 0.124 ± 0.025, 0.084 ± 0.015, 0.079 ± 0.023, 0.087 ± 0.028, 0.077 ± 0.01, 0.03 ± 0.013, 0.021 ± 0.013, 0.017 ± 0.014, 0.016 ± 0.013, 0.001 ± 0.001, and 0 ± 0. t tests were performed to the two datasets at all values of distance cutoff. p values less than 0.0001 are given four asterisks.

cutoff value d_c, as shown in Figure 5A. After selecting initial conformations, both proteins diffused against each other under the guidance of their inter-molecular interactions, which included hydrophobic and electrostatic effects. The two proteins either diffused away from each other or formed the complex before the end of simulations (Figure 5A). To obtain statistically meaningful results, 10^3 simulation trajectories were generated with different initial conformations but constrained by the same cutoff value. From these trajectories, we counted how frequently the TCR/MHC complexes were formed and systematically scanned different cutoff values from 15 to 25 Å. Figure 5C plots the relation between the distance cutoff d_c and the frequency of complexes formation as black squares, showing that the highest frequency of association was over 0.45 when the value of distance cutoff equaled 15 Å. However, the association rate dropped as the distance cutoff increased, indicating that complexes are more difficult to form between TCR and MHC if they are initially separated farther from each other.

Previous experiments showed that mutations at the interface of TCR and MHC strongly affected their interaction. For example, the R65A mutation in MHC reduced the binding affinity between TCR and MHC (Blevins et al., 2016). As shown in Figure 5B, ARG65 is directly involved in the binding interface of the complex.
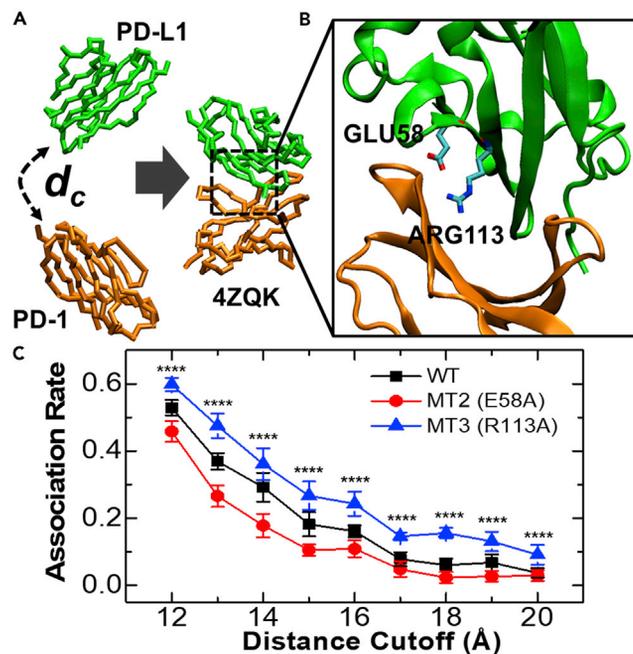
Figure 6. Estimation of Association Rates for Binding between PD-1 and PD-L1

(A–C) The association between PD-L1 and PD-1 was also simulated with the same residue-based kinetic Monte Carlo algorithm (A). The complex formed by human PD-L1 and PD-1 was used as a test system (PDB id 4ZQK). Multiple simulation trajectories were started from different values of the distance cutoff d_c. We also counted how many encounter complexes were formed among these trajectories. Two polar residues at the binding interface, GLU58 and ARG113, were specifically selected (B). We tested the effect of these two mutants on the association. The calculated frequencies of association at various distance cutoffs were plotted in (C) with error bars for both wild-type and mutants. Interestingly, the association is weakened in mutant E58A (red circles) but strengthened in mutant R113A (blue triangles) relative to the wild-type (black squares). The calculated means and standard errors of the association rates for the wild-type from 12 to 20 Å are listed as follows: 0.529 ± 0.023, 0.369 ± 0.025, 0.292 ± 0.043, 0.182 ± 0.036, 0.161 ± 0.018, 0.078 ± 0.021, 0.06 ± 0.02, 0.068 ± 0.024, and 0.037 ± 0.014. The calculated means and standard errors of the association rates for the mutant E58A from 12 to 20 Å are listed as follows: 0.458 ± 0.031, 0.266 ± 0.032, 0.178 ± 0.035, 0.105 ± 0.017, 0.109 ± 0.025, 0.048 ± 0.023, 0.023 ± 0.017, 0.027 ± 0.016, and 0.029 ± 0.016. The calculated means and standard errors of the association rates for the mutant R113A from 12 to 20 Å are listed as follows: 0.598 ± 0.02, 0.475 ± 0.036, 0.361 ± 0.047, 0.267 ± 0.042, 0.243 ± 0.036, 0.146 ± 0.012, 0.156 ± 0.015, 0.131 ± 0.029, and 0.091 ± 0.029. The one-way analysis of variance (ANOVA) was performed to the three datasets at all values of distance cutoff. p values less than 0.0001 are given four asterisks.

To test the effect of this mutant, we replaced the side chain of ARG65 in the wild-type MHC molecule with ALA. The kMC simulation was then applied to generate 10^3 trajectories under different distance cutoff values. Based on the statistical analysis of these trajectories, the frequencies of association at various distance cutoffs were calculated (red circles in Figure 5C). Comparing with the wild-type, the figure shows that the mutation dramatically weakened the association between TCR and MHC. The highest frequency of association dropped from 0.45 in the wild-type to lower than 0.2 in the mutated system. Consistent with the experimental data, our simulations indicated that the mutation of a single residue on the surface of MHC could significantly impact its binding with TCR.

Parallel to the TCR/MHC interaction, the complex formed between human PD-1 and its ligand PD-L1 was evaluated with the same simulation approach. The atomic coordinates of this complex were taken from the PDB id 4ZQK (Zak et al., 2015). Following the same simulation procedure, 10^3 trajectories were carried out with different initial conformations in which the distances between the binding interfaces of PD-L1 and PD-1 fell under a specific cutoff value d_c (Figure 6A). Different distance cutoffs from 12 to 20Å were tested, and for each cutoff, we calculated the frequency of association by counting how many complexes were successfully formed among all the trajectories. Figure 6C plots the relation between d_c and the frequency of complex formation as black squares. We further examined mutations that affect the association between PD-L1 and PD-1—specifically, two surface residues on PD-L1, GLU58, and ARG113, which are directly involved in

the binding interface (Figure 6B). We predicted that altered electrostatic interactions at the binding interface would lead to differences in binding between two proteins, and we evaluated the behavior of the E58A and R113A mutants relative to wild-type PD-L1. The kinetic Monte Carlo simulation was then applied to these mutants to generate 10^3 trajectories within the same range of distance cutoff values, and based on the statistical analysis of these trajectories, the frequencies of association were calculated. As shown in Figure 6C, the association is weakened in the E58A mutant (red circles) compared with the wild-type. In contrast, interestingly, the association is strengthened in the R113A mutant (blue triangles) relative to the wild-type. The effects of these two mutants on the association between PD-L1 and PD-1 can be validated using experimental methods, such as surface plasmon resonance (Daghestani and Day, 2010).

In summary, using computational simulations, we could calculate the rate of association between ligands in our designed biologics and their targeted receptors. We further demonstrated that we could estimate how mutations of specific residues at the binding interfaces between ligands and receptors affected the association rates. Finally, these association rates for both wild-types and mutants will be integrated into the RB model as input parameters to realistically simulate the binding between biologics and receptors.

Integrate the Dynamic and Binding Parameters into the RB-Based Simulations

The association rates of TCR/MHC and PD-1/PD-L1 binding were calculated by the statistical analysis of the kMC simulation results and provided parameters to guide the association between biologics and receptors in RB-based simulations. Specifically, the association rate between wild-type MHC and TCR is 3.5×10^{-2} ns^{-1}, and the association rate between wild-type PD-1 and PD-L1 is 3.0×10^{-2} ns^{-1}. Moreover, the binding affinities for TCR/MHC and PD-1/PD-L1 interactions were used to guide the dissociation of biologics and receptors. The values of binding affinities for the interactions between wild-type ligands and receptors were derived from previous experimental measurements: the dissociation constant for the interaction between wild-type MHC (HLA-A 0201) and TCR (B7) is 1.3×10^{-6}M (Davis-Harrison et al., 2005), corresponding to a binding energy of -8.01 kcal/mol, whereas the dissociation constant for interaction between wild-type PD-1 and PD-L1 is 6.36×10^{-6}M (Lazar-Molnar et al., 2017), corresponding to a binding energy of -7.11 kcal/mol. In addition to these binding parameters, dynamic parameters were also obtained from the Anton all-atom MD simulations to model the relative positions and orientations of MHC and PD-L1 in biologics; these dynamic parameters are specific to each of the three linkers. With these parameters, the RB-based simulations incorporated realistic information about the energetic features at the ligand/receptor binding interfaces, the sequence and structural patterns of different linkers, and their impacts on receptor targeting. For instance, Figure 7A plots the final configuration from a specific simulation scenario in which biologics contained wild-type MHC and PD-L1 and were connected by the GS30 linker. The figure shows that most biologics were bound to their targeted receptors on cell surfaces. Interestingly, these complexes formed by biologics and receptors were spatially localized in small clusters on the plasma membrane. We speculate that the complicated topology of large oligomers formed between the bispecific fusion proteins and their multiple receptors causes local crowding on cell surfaces, leading to this aggregation. The spatial organization of membrane receptors based on the binding of biologics can be validated using experimental methods, such as super-resolution microscopy.

The effects of different mutations in the biologics on receptor targeting were further estimated and compared with the wild-type. Specifically, four types of biologics mutants were tested. The first (MT1) contained the mutated residue R65A in the MHC functional module. The association rate and binding affinity between MHC and TCR in MT1 are 8.0×10^{-3} ns^{-1} and -5.48 kcal/mol, respectively. The second and third mutants (MT2 and MT3) contained the mutated residues E58A and R113A in the functional module of PD-L1, respectively. The association rate and binding affinity between PD-L1 and PD-1 in MT2 are 1.5×10^{-2} ns^{-1} and -5.3 kcal/mol, whereas the association rate and binding affinity in MT3 are 4.5×10^{-2} ns^{-1} and -9.0 kcal/mol, respectively. The last mutant (MT4) contained double mutations with both R65A in MHC and E58A in PD-L1. As described above, the biologics with these mutations were positioned above a 1×1 μm surface area containing 300 TCR and 300 PD-1. Starting with the same initial conditions, Figure 7B summarizes the simulation results of all these mutants and compares them with the wild-type system, as shown by the black curve. The numbers of biologics that bound to both TCR and PD-1 on the surfaces are plotted as a function of simulation time. The figure indicates that all mutants weakened the targeting effect of the biologics. In particular, the double mutant (MT4) almost totally abolished the receptor binding, as shown by the green curve in Figure 7B.

More interestingly, Figure 7B suggests that MT3 (orange curve) also reduced the number of biologics that bound to both TCR and PD-1, even though the binding between PD-1 and PD-L1 was strengthened in the
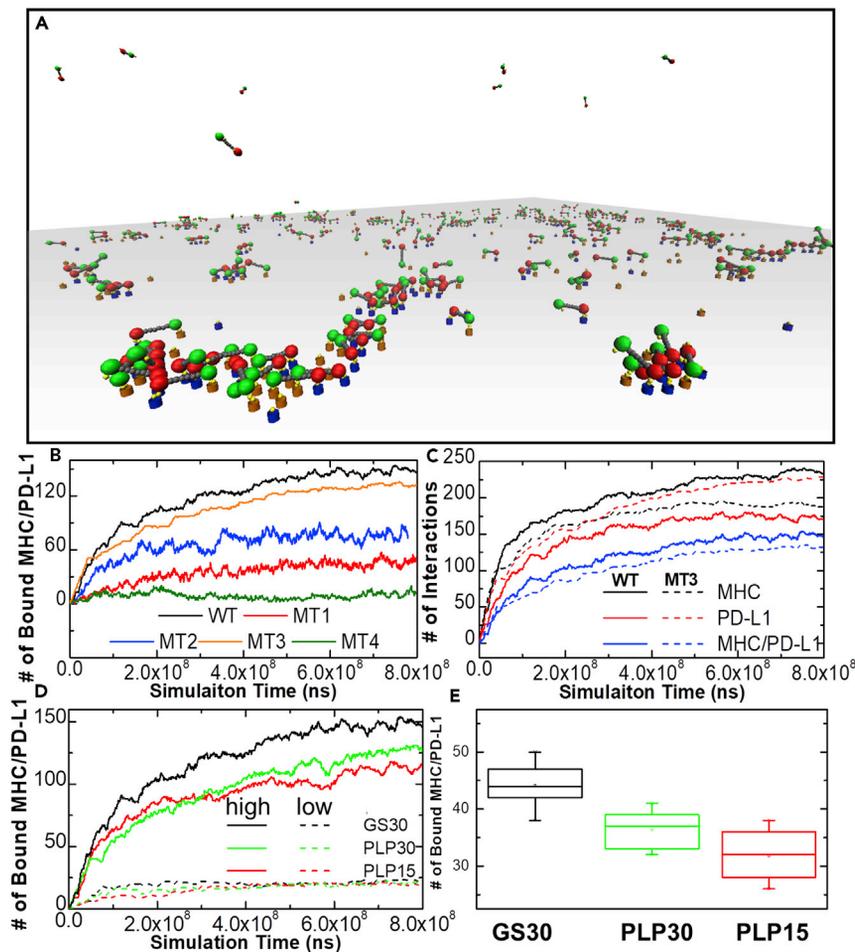
Figure 7. The Rigid Body Simulation Results After the Integration of Conformational Dynamics and Binding Constants from the High-Resolution Models

(A–E) The final configuration of the RB simulation is shown in (A), in which biologics containing wild-type MHC and PD-L1 with the GS30 linker were exposed to cells expressing high levels of TCR. In (B), we tested the effect of mutations in different functional modules on receptor targeting. The means and standard errors of the curves in this plot within the last 2×10^8 ns are listed as follows: 142.5 ± 6.61 (black), 43.32 ± 4.28 (red), 75.57 ± 5.06 (blue), 126.45 ± 5.69 (orange), 7.65 ± 2.99 (green). In (C), we show the mutation that strengthens the PD-L1/PD-1 binding (MT3) can weaken the interaction between TCR and MHC, even if the TCR/MHC binding interface is not changed. The means and standard errors of the curves in this plot within the last 2×10^8 ns are listed as follows: 233.27 ± 4.51 (black solid), 173.79 ± 3.22 (red solid), 147.39 ± 3.15 (blue solid), 190.21 ± 2.02 (black dashed), 224.26 ± 2.52 (red dashed), 130.87 ± 2.41 (blue dashed). In (D), we also tested how different types of linkers affect the binding between biologics and receptors on the surface of both high- and low-expressing cells. The means and standard errors of the curves in this plot within the last 2×10^8 ns are listed as follows: 147.39 ± 3.15 (black solid), 111.71 ± 3.34 (red solid), 124.91 ± 3.96 (green solid), 22.02 ± 1.30 (black dashed), 19.93 ± 0.57 (red dashed), 20.21 ± 1.31 (green dashed). To achieve both statistical significance and computational accessibility, we carried out multiple trajectories with a relatively smaller simulation setup for all three linker systems, the statistical distributions of which are shown in (E) as a box-whisker plot. The box of each distribution in the plot includes the 25th–75th percentile range for the number of biologics bound to both TCR and PD-1, while their average number is marked in the middle of each box. The whisker indicates the outlier of the distribution with the coefficient equal to 1.5.

mutant. To understand how this positive mutant resulted in a negative effect on surface targeting, we performed a more detailed analysis in Figure 7C. In addition to comparing the number of biologics in which both functional modules bound to their targeted receptors between wild-type (blue solid curve) and MT3 (blue dashed curve) mutants, we further compared the number of biologics with one of two functional modules bound to its corresponding receptor. The number of bound PD-L1 in wild-type biologics is plotted by a red solid curve, the number of bound MHC in wild-type biologics is plotted by a black solid curve, and the number of bound PD-L1 and MHC in MT4 is plotted by red and black dashed curves, respectively. The

figure shows more bound PD-L1 in wild-type biologics than MT3, which can be explained by the strengthened interaction between PD-1 and PD-L1 due to the mutation of residue ARG113. However, the figure also shows that the number of bound MHC decreased in MT3, although the interaction between MHC and TCR in the mutant remained unchanged. In other words, if the binding of one functional module is too strong, it prevents the other functional module in the biologics from reaching its receptor. Our results thus suggest that MHC and PD-L1 in the bispecific fusion proteins are negatively coupled with one another while seeking their respective receptors. This observation could also explain why the number of biologics that bound both TCR and PD-1 is lower in the mutant. This negative cooperativity of binding between different functional modules should be considered during the design of new multi-specific biologics in the future.

We further examined the effects of linker type on the interactions between multi-functional biologics and receptors. As described earlier, the parameters of translation and rotational deviations from each functional module were derived from MD simulations for biologics with GS30, PLP15, and PLP30 linkers, which controlled the distribution of distances and orientations between MHC and PD-L1. For the purpose of simplicity, only wild-type MHC and PD-L1 were considered. As a result, three types of biologics were exposed to two models of cells: those with high expression levels of TCR and those with low expression levels of TCR. Figure 7D summarizes the simulation results of these six combinations of scenarios, with the numbers of biologics that bind to both TCR and PD-1 plotted as a function of simulation time. The figure shows no difference when biologics with all three linkers were targeted to low-expressing cells, as plotted by the dashed curves. However, when they interacted with receptors on the surfaces of the high-expressing cells, they behaved very differently. Specifically, the GS30 linker resulted in the highest number of biologics targeting both TCR and PD-1 (black curve), whereas the linker PLP15 led to the lowest number of biologics with both ligands MHC and PD-L1 bound to their receptors (red curve).

To achieve both statistical significance and computational accessibility, we performed multiple trajectories on a relatively smaller simulation setup for all three linker systems. In detail, 10 trajectories were generated for each system, with each trajectory extending for 5×10^7 ns and starting from a random configuration with 100 biologics distributed in the extracellular region plus 100 TCR and 100 PD-1 on the bottom of the simulation box with a volume of $500 \times 500 \times 200$ nm^3. We counted the numbers of biologics that bound to both TCR and PD-1 at the ends of all trajectories. Figure 7E plots the distributions of all three linkers. Consistent with Figure 7D, the figure indicates that the GS30 linker systematically resulted in more receptor targeting than the other two linkers, whereas the system with the PLP15 linker yielded the least receptor targeting. We applied one-way analysis of variance (ANOVA) to test the statistical significance of our observation (Welch, 1951), which yielded a calculated F-statistic score of 29.5 with a p value of 0.0001, suggesting that variations of receptor binding caused by different linkers are statistically significant. Based on this statistical result, we suggest that longer, more flexible linkers in multi-specific biologics can play a positive role in cell-surface targeting by enhancing the local conformational sampling of receptor binding.

This observation can also be explained by analyzing torsional angles between two functional modules in the biologics, which the Supplemental Information and Figure S2A also describe. We calculated the values of these angles for all snapshots from the MD simulations and compared their distributions for biologics containing different linkers. Figure S2B plots the results of the comparison as histograms, showing that linker GS30 forms a uniform distribution within the range −180° to +180° (red bars). In contrast, multiple peaks were observed in the distribution of linker PLP15 (black bars), indicating the packing preference between two functional modules. Intuitively, a linker with a more uniform torsional-angle distribution may have a lower bias in seeing binding partners for the second module once the first is bound. This entails that biologics with this type of linker can bind to more receptors on cell surfaces. Moreover, while biologics with longer, more flexible linkers can target more receptors, we can further assume that they are more sensitive to mediate the T cell coregulatory pathways. This prediction may be validated by T cell stimulation assays. For instance, when we exposed biologics with different linkers to CD8+ 8.3 T cells, we expected that the high conformational dynamics in linkers such as GS30 would enhance T cell activation so that stronger signals of T cell proliferation would be detected than the biologics, which carry shorter, less flexible linkers, such as PLP15.

CONCLUDING DISCUSSIONS

Increasingly, biological medications derived from living organisms are garnering approval from the Food and Drug Administration for treating many cancers, such as melanoma (Kinch, 2015). However, administering these immunotherapeutic biologics often has extensive side effects (Ott et al., 2017; Wanchoo

et al., 2017) because these molecules can simultaneously activate many off-target immune cells and elicit global immune modulation. To lower these risks, recombinant DNA technology is frequently implemented to genetically fuse naturally occurring molecules to form multi-specific biologics with enhanced specificity for binding to their targeted cells (Baldo, 2015). Although computational methods provide opportunities to understand the functional characteristics of these multi-specific biologics on a quantitative level, the feasibility and practicality of these approaches are limited by the trade-off between the timescales that can be evaluated and the accuracy that can be achieved. To address these challenges, we constructed a computational framework that integrates simulation algorithms on different levels to understand the binding between multi-specific biologics and surface-bound membrane receptors. We applied the method to an artificially designed biologic, in which two functional modules were covalently tethered with a peptide linker. The conformational dynamics of the biologics were modulated by modifying the length and composition of the linker, and computational mutagenesis was performed to tune the binding between each module in the biologics with their respective receptors. This multiscale framework has revealed molecular and cellular factors that regulate the binding specificity of biologics. For instance, cells expressing higher levels of targeted receptors are more sensitive to variations in ligand concentrations. On the molecular level, we demonstrated that the linker is an important, underappreciated contributor to modulating ligand-receptor interactions. Moreover, the discovery that the mutations that impact affinity in one module can affect the function of the other module suggests important considerations regarding the binding mechanisms of multivalent multi-specific ligands. In summary, our method highlights the importance of computational modeling in the development of next-generation biologics for immunotherapy.

One observation from our simulations is the negative cooperativity between the bindings of different functional modules with their corresponding receptors. Our previous computational model (Chen et al., 2017) indicated the mechanism underlying this negative coupling effect. With that model, we found that if the lifetime of a ligand-receptor bond is much shorter than the average time of diffusion a ligand spends to find its receptor, tethering different functional modules creates a negligible coupling effect. When the binding affinities increase to a range in which the lifetime of a ligand-receptor bond is comparable to the average diffusion time a ligand needs to find its receptor, one functional module in a biologics is highly probable to find its receptor after the dissociation of other modules in the biologics. As a result, binding one functional module in a multi-specific biologics positively enhances the binding of other modules. When the binding affinities further increase to a range in which the lifetime of a ligand-receptor bond is much longer than the average diffusion time a ligand needs to find its receptor; however, the binding of one functional module in a multi-specific biologics could kinetically trap other functional modules from approaching their receptors when they diffuse. Consequently, negative cooperativity was obtained among bindings of different functional modules. Consistent with this model, the simulation results of our study suggest that the negative coupling between the binding of different functional modules is a result of the enhanced association rate and binding affinity of the mutant MT3.

With our RB-based simulation, we could study the binding kinetics between cell surface-bound receptors and multi-domain flexible ligands with both spatial information and a long timescale. Other mesoscopic simulation techniques on a similar level can also be applied to study the dynamics of cellular systems, of course. For instance, SpringSaLaD has recently been developed to study the function of receptor clustering in cell signaling (Michalski and Loew, 2016), adopting a similar molecular representation and diffusion-reaction algorithm. Unlike our method, however, the diffusions in SpringSaLaD are implemented by a Langevin dynamic integrator. While applying this method to a system with multivalent interactions among membrane-anchored receptors and adaptor, and effector proteins containing multiple binding motifs, Chattarraj et al. found that receptor clustering was diminished by the increased flexibility of the interacting molecules (Chattaraj et al., 2019). It would be interesting to apply SpringSaLaD to the molecular system in this article and compare their output with ours. We expect both simulations to obtain qualitatively similar results, that is, that biologics containing long, flexible linkers capture more receptors due to enhanced local conformational searching.

The test system in this study, the multi-specific biologic that connects MHC with PD-L1, was inspired by prior work on the single-chain peptide MHC (Samanta et al., 2011). That work fused a peptide called IGRP$_{206-214}$, the β2m, and the heavy chain of the class I MHC molecule H-2K^d into one molecule. X-ray diffraction analysis confirmed that the peptide was presented in the canonical binding groove of the heavy chain. Further experiments showed that the protein reagent could induce the apoptosis of naive CD8$^+$ 8.3

T cells but had no effect on the non-cognate AI4 T cells. This work demonstrated the ability of single-chain peptide MHC to achieve the specificity of T cell targeting. Its conjugation with the coregulatory protein PD-L1 would also allow manipulation of the signaling response of these targeted cells, potentially eliciting a therapeutic effect. This type of fusion protein also possesses unique advantages over other targeted biological therapies. For instance, they do not need to be internalized to carry out their functions, unlike antibody-drug conjugates, which suffer from a poor internalization rate (Sievers and Senter, 2013). For the future generalization of our method, the coregulatory module in the multi-specific biologics can be replaced by any molecular category, such as an antibody, cytokine, or antibody fragment. For example, instead of targeting the inhibitory pathway with PD-L1, TNF can be incorporated to initiate a stimulatory pathway. With the antigen-specific MHC, the integration of a new regulatory module can shape the function of a specially selected T cell subgroup and avoid a global immune response. Alternatively, the current model could be adapted to study other categories of multi-specific biologics. For instance, a single layer of plasma membrane was modeled in this study, but to study the binding properties of BiTE at the interface of a T cell and a tumor, as mentioned in the Introduction, it could easily be extended to a system with two layers of plasma membrane. The CD3 receptors and antigens would be placed on the surfaces of the T cell and tumor cell, respectively, and BiTE would be distributed at the intercellular region so that one fragment of this bispecific fusion protein could bind to CD3 and the other could bind to the antigen on the opposite cell surface. This would quantitatively evaluate the functional impacts of BiTE on the coupling between cytotoxic T cells and tumor cells. Using the multiscale simulation framework developed in this study and its future extensions, we will be able to fine-tune the molecular features of different biologics, such as the connectivity between different modules and their binding avidity. We will also be able to tune the binding properties of each module through large-scale computational mutagenesis studies. In summary, our study adds a powerful platform to a suite of existing approaches to engineer new biomolecules for immunotherapy.

Limitations of the Study

Further improvements to our simulation method will be made to overcome limitations associated with the simplifying assumptions in the current model. For instance, in living cells, biomolecules occupy 20%–40% of cytoplasmic volume, significantly affecting the thermodynamics and kinetics of molecular binding (Kozer and Schreiber, 2004; Minton, 2001). Similarly, cell surfaces and interfaces are crowded by many membrane receptors, adhesion molecules, and channels. Our simulation ignored the effect of cell-surface crowding, which could affect the binding kinetics of multi-specific biologics with their targeted receptors. The spatial heterogeneity and diversity of molecular composition on cell membranes can easily be introduced into the RB-based model. Specifically, in addition to multi-specific biologics and their targeted receptors, we can place many other cell-surface proteins on cell surfaces, which will collide repulsively with the biologics and form nonspecific interactions with each other, thus allowing the effect of crowding on both diffusion and binding to be studied. Additionally, the plasma membrane was modeled as a flat planar surface in this study, and the effects of membrane fluctuations were not considered. In the future, a more realistic representation of the plasma membrane will be constructed to model its collective motions as an elastic medium. For instance, Voronoi tessellation (Zaninetti, 1989) might be applied to discretize a plasma membrane into a 2D meshwork, and the conformation of the meshwork can be changed by the elastic network model (Atilgan et al., 2001). We believe that the binding rates between receptors and multi-specific biologics will be sensitive to the collective motions of the plasma membrane. Conversely, the ligand-receptor interactions on cell surfaces will affect the fluctuations of the plasma membrane. We expect this cooperativity between receptor binding and membrane fluctuations to play an important role in regulating the cell-surface targeting of multi-specific biologics. Finally, a hybrid simulation technique that combines the approach in this article with the mathematic modeling of intracellular signaling pathways would allow us to explore the impacts of certain biologics on the modulation of specific immune responses.

Resource Availability

Lead Contact

Yinghao Wu, Associate Professor, Albert Einstein College of Medicine, Bronx, NY 10461, Email: yinghao.wu@einsteinmed.org.

Materials Availability

This study does not generate any new unique reagents.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101835.

AUTHOR CONTRIBUTIONS

B.W., Z.S., S.C.A., and Y.W. designed the research; Z. S., B.W., and Y.W. performed the research; Z.S., B.W., and Y.W. analyzed the data, and S.C.A. and Y.W. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adzhubei, A.A., and Sternberg, M.J. (1993). Left-handed polyproline II helices commonly occur in globular proteins. J. Mol. Biol. *229*, 472–493.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. *80*, 505–515.

Bagowski, C.P., Bruins, W., and Te Velthuis, A.J. (2010). The nature of protein domain evolution: shaping the interaction network. Curr. Genomics *11*, 368–376.

Bai, Y., Ann, D.K., and Shen, W.C. (2005). Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. Proc. Natl. Acad. Sci. U. S. A. *102*, 7292–7296.

Bai, Y., and Shen, W.C. (2006). Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. Pharm. Res. *23*, 2116–2121.

Baldo, B.A. (2015). Chimeric fusion proteins used for therapy: indications, mechanisms, and safety. Drug Saf. *38*, 455–479.

Berger, S., Lowe, P., and Tesar, M. (2015). Fusion protein technologies for biopharmaceuticals: applications and challenges: Editor Stefan R Schmidt. mAbs *7*, 456–460.

Blevins, S.J., Pierce, B.G., Singh, N.K., Riley, T.P., Wang, Y., Spear, T.T., Nishimura, M.I., Weng, Z., and Baker, B.M. (2016). How structural adaptability exists alongside HLA-A2 bias in the

human alphabeta TCR repertoire. Proc. Natl. Acad. Sci. U. S. A. *113*, E1276–E1285.

Bluestone, J.A., St Clair, E.W., and Turka, L.A. (2006). CTLA4Ig: bridging the basic immunology with clinical application. Immunity *24*, 233–238.

Brameshuber, M., Kellner, F., Rossboth, B.K., Ta, H., Alge, K., Sevcsik, E., Gohring, J., Axmann, M., Baumgart, F., Gascoigne, N.R.J., et al. (2018). Monomeric TCRs drive T cell antigen recognition. Nat. Immunol. *19*, 487–496.

Brender, J.R., and Zhang, Y. (2015). Predicting the effect of mutations on protein-protein binding interactions through structure-based interface profiles. PLoS Comput. Biol. *11*, e1004494.

Carnero, A. (2006). High throughput screening in drug discovery. Clin. Transl. Oncol. *8*, 482–490.

Chakrabarti, A., Verbridge, S., Stroock, A.D., Fischbach, C., and Varner, J.D. (2012). Multiscale models of breast cancer progression. Ann. Biomed. Eng. 40, 2488–2500.

Chattaraj, A., Youngstrom, M., and Loew, L.M. (2019). The interplay of structural and cellular biophysics controls clustering of multivalent molecules. Biophys. J. 116, 560–572.

Chen, J., Almo, S.C., and Wu, Y. (2017). General principles of binding between cell surface receptors and multi-specific ligands: a computational study. PLoS Comput. Biol. 13, e1005805.

Chen, X., Zaro, J.L., and Shen, W.C. (2013). Fusion protein linkers: property, design and functionality. Adv. Drug Deliv. Rev. 65, 1357–1369.

Choong, Y.S., Lee, Y.V., Soong, J.X., Law, C.T., and Lim, Y.Y. (2017). Computer-aided antibody design: an overview. Adv. Exp. Med. Biol. 1053, 221–243.

Daghestani, H.N., and Day, B.W. (2010). Theory and applications of surface plasmon resonance, resonant mirror, resonant waveguide grating, and dual polarization interferometry biosensors. Sensors 10, 9630–9646.

Davis-Harrison, R.L., Armstrong, K.M., and Baker, B.M. (2005). Two different T cell receptors use different thermodynamic strategies to recognize the same peptide/MHC ligand. J. Mol. Biol. 346, 533–550.

Dimasi, N., Gao, C., Fleming, R., Woods, R.M., Yao, X.T., Shirinian, L., Kiener, P.A., and Wu, H. (2009). The design and characterization of oligospecific antibodies for simultaneous targeting of multiple disease mediators. J. Mol. Biol. 393, 672–692.

Ding, Y.H., Smith, K.J., Garboczi, D.N., Utz, U., Biddison, W.E., and Wiley, D.C. (1998). Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity 8, 403–411.

Farhadi, T., and Hashemian, S.M. (2018). Computer-aided design of amino acid-based therapeutics: a review. Drug Des. Devel. Ther. 12, 1239–1254.

Gun, S.Y., Lee, S.W.L., Sieow, J.L., and Wong, S.C. (2019). Targeting immune cells for cancer therapy. Redox Biol. 25, 101174.

Huehls, A.M., Coupet, T.A., and Sentman, C.L. (2015). Bispecific T-cell engagers for cancer immunotherapy. Immunol. Cell Biol. 93, 290–296.

Im, W., Liang, J., Olson, A., Zhou, H.X., Vajda, S., and Vakser, I.A. (2016). Challenges in structural approaches to cell modeling. J. Mol. Biol. 428, 2943–2964.

Kaczor, A.A., Bartuzi, D., Stepniewski, T.M., Matosiuk, D., and Selent, J. (2018). Protein-protein docking in drug design and discovery. Methods Mol. Biol. 1762, 285–305.

Karplus, M., and Petsko, G.A. (1990). Molecular dynamics simulations in biology. Nature 347, 631–639.

Katsila, T., Spyroulias, G.A., Patrinos, G.P., and Matsoukas, M.T. (2016). Computational approaches in target identification and drug discovery. Comput. Struct. Biotechnol. J. 14, 177–184.

Kinch, M.S. (2015). An overview of FDA-approved biologics medicines. Drug Discov. Today 20, 393–398.

Kontoyianni, M. (2017). Docking and virtual screening in drug discovery. Methods Mol. Biol. 1647, 255–266.

Kozer, N., and Schreiber, G. (2004). Effect of crowding on protein - protein association rates: fundamental differences between low and high mass crowding agents. J. Mol. Biol. 336, 763–774.

Krobath, H., Rozycki, B., Lipowsky, R., and Weikl, T.R. (2009). Binding cooperativity of membrane adhesion receptors. Soft Matter 5, 3354–3361.

Kruger, S., Ilmer, M., Kobold, S., Cadilha, B.L., Endres, S., Ormanns, S., Schuebbe, G., Renz, B.W., D'haese, J.G., Schloesser, H., et al. (2019). Advances in cancer immunotherapy 2019 - latest trends. J. Exp. Clin. Cancer Res. 38, 268.

Lazar-Molnar, E., Scandiuzzi, L., Basu, I., Quinn, T., Sylvestre, E., Palmieri, E., Ramagopal, U.A., Nathenson, S.G., Guha, C., and Almo, S.C. (2017). Structure-guided development of a high-affinity human programmed cell death-1: implications for tumor immunotherapy. EBioMedicine 17, 30–44.

Li, K., Cheng, X., Tilevik, A., Davis, S.J., and Zhu, C. (2017). In situ and in silico kinetic analyses of programmed cell death-1 (PD-1) receptor, programmed cell death ligands, and B7-1 protein interaction network. J. Biol. Chem. 292, 6799–6809.

Michalski, P.J., and Loew, L.M. (2016). SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume. Biophys. J. 110, 523–529.

Michot, J.M., Bigenwald, C., Champiat, S., Collins, M., Carbonnel, F., Postel-Vinay, S., Berdelou, A., Varga, A., Bahleda, R., Hollebecque, A., et al. (2016). Immune-related adverse events with immune checkpoint blockade: a comprehensive review. Eur. J. Cancer 54, 139–148.

Minton, A.P. (2001). The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J. Biol. Chem. 276, 10577–10580.

Ott, P.A., Hodi, F.S., Kaufman, H.L., Wigginton, J.M., and Wolchok, J.D. (2017). Combination immunotherapy: a road map. J. Immunother. Cancer 5, 16.

Padte, N.N., Yu, J., Huang, Y., and Ho, D.D. (2018). Engineering multi-specific antibodies against HIV-1. Retrovirology 15, 60.

Pan, A.C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T.M., and Shaw, D.E. (2019). Atomic-level characterization of protein-protein association. Proc. Natl. Acad. Sci. U. S. A. 116, 4244–4249.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat. Rev. Cancer 12, 252–264.

Plattner, N., Doerr, S., De Fabritiis, G., and Noe, F. (2017). Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nat. Chem. 9, 1005–1011.

Qin, S., Pang, X.D., and Zhou, H.X. (2011). Automated prediction of protein association rate constants. Structure 19, 1744–1751.

Quayle, S.N., Girgis, N., Thapa, D.R., Merazga, Z., Kemp, M.M., Histed, A., Zhao, F., Moreta, M., Ruthardt, P., Hulot, S., et al. (2020). CUE-101, a Novel E7-pHLA-IL2-Fc fusion protein, enhances tumor antigen-specific T-cell activation for the treatment of HPV16-driven malignancies. Clin. Cancer Res. 26, 1953–1964.

Ramis-Conde, I., Drasdo, D., Anderson, A.R.A., and Chaplain, M.A.J. (2008). Modeling the influence of the E-cadherin-beta-catenin pathway in cancer cell invasion: a multiscale approach. Biophys. J. 95, 155–165.

Ruggiero, M.T., Sibik, J., Orlando, R., Zeitler, J.A., and Korter, T.M. (2016). Measuring the elasticity of poly-l-proline helices with terahertz spectroscopy. Angew. Chem. Int. Ed. Engl. 55, 6877–6881.

Samanta, D., Mukherjee, G., Ramagopal, U.A., Chaparro, R.J., Nathenson, S.G., Dilorenzo, T.P., and Almo, S.C. (2011). Structural and functional characterization of a single-chain peptide-MHC molecule that modulates both naive and activated CD8+ T cells. Proc. Natl. Acad. Sci. U. S. A. 108, 13682–13687.

Seror, R., and Mariette, X. (2017). Malignancy and the risks of biologic therapies: current status. Rheum. Dis. Clin. North Am. 43, 43–64.

Sievers, E.L., and Senter, P.D. (2013). Antibody-drug conjugates in cancer therapy. Annu. Rev. Med. 64, 15–29.

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E.W., Jr. (2014). Computational methods in drug discovery. Pharmacol. Rev. 66, 334–395.

Thakur, A., and Lum, L.G. (2016). "NextGen" biologics: bispecific antibodies and emerging clinical results. Expert Opin. Biol. Ther. 16, 675–688.

Uze, G., and Tavernier, J. (2015). High efficiency targeting of IFN-alpha activity: possible applications in fighting tumours and infections. Cytokine Growth Factor Rev. 26, 179–182.

Van Rosmalen, M., Krom, M., and Merkx, M. (2017). Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. Biochemistry 56, 6565–6574.

Villoutreix, B.O., Bastard, K., Sperandio, O., Fahraeus, R., Poyet, J.L., Calvo, F., Deprez, B., and Miteva, M.A. (2008). In silico-in vitro screening of protein-protein interactions: towards the next generation of therapeutics. Curr. Pharm. Biotechnol. 9, 103–122.

Vogel, C., Bashton, M., Kerrison, N.D., Chothia, C., and Teichmann, S.A. (2004). Structure,

function and evolution of multidomain proteins. Curr. Opin. Struct. Biol. *14*, 208–216.

Wanchoo, R., Karam, S., Uppal, N.N., Barta, V.S., Deray, G., Devoe, C., Launay-Vacher, V., and Jhaveri, K.D. (2017). Adverse renal effects of immune checkpoint inhibitors: a narrative review. Am. J. Nephrol. *45*, 160–169.

Way, J.C., Collins, J.J., Keasling, J.D., and Silver, P.A. (2014). Integrating biological redesign: where synthetic biology came from and where it needs to go. Cell *157*, 151–161.

Weidle, U.H., Tiefenthaler, G., Weiss, E.H., Georges, G., and Brinkmann, U. (2013). The intriguing options of multispecific antibody formats for treatment of cancer. Cancer Genomics Proteomics *10*, 1–18.

Welch, B.L. (1951). On the comparison of several mean values: an alternative approach. Biometrika *38*, 330–336.

Xiong, P., Zhang, C., Zheng, W., and Zhang, Y. (2017). BindProfX: assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. J. Mol. Biol. *429*, 426–434.

Yang, Y. (2015). Cancer immunotherapy: harnessing the immune system to battle cancer. J. Clin. Invest. *125*, 3335–3337.

Zak, K.M., Kitel, R., Przetocka, S., Golik, P., Guzik, K., Musielak, B., Domling, A., Dubin, G., and Holak, T.A. (2015). Structure of the complex of human programmed death 1, PD-1, and its ligand PD-L1. Structure *23*, 2341–2348.

Zaninetti, L. (1989). Dynamical voronoi tessellation.1. The two-dimensional case. Astron. Astrophys. *224*, 345–350.

Supplemental Information

Understanding the Targeting

Mechanisms of Multi-Specific Biologics

in Immunotherapy with Multiscale Modeling

Zhaoqian Su, Bo Wang, Steven C. Almo, and Yinghao Wu

Understand the Targeting Mechanisms of Multi-specific Biologics in Immunotherapy by Multiscale Modeling

Zhaoqian Su, Bo Wang, Steven C. Almo and Yinghao Wu

Supplemental Information

Transparent Methods

Simulate cell targeting of multi-specific biologics with a diffusion-reaction algorithm

As described in the **Results**, an RB-based model was constructed to simulate the binding between cell-surface receptors and multi-specific biologics (Xie et al., 2014). In brief, the plasma membrane was represented by a flat surface below the extracellular region, each receptor was represented by an RB of cylinders, and each functional module in the biologics was simplified as a spherical RB. The sizes of these RBs were comparable to their real protein structures. Binding sites were assigned on the tops of receptors and on the surfaces of functional modules in biologics. Two functional modules were further tethered together. Considering the flexibility of peptide linkers, we allowed translational and rotational fluctuations between these two modules around their mean positions and orientations, as described later. Given the model representation and a randomly generated initial configuration (**Figure 2a**), the dynamics of the system evolved following a diffusion-reaction algorithm. Molecules were selected in random order for stochastic diffusion as the first scenario at each simulation time step. The biologics were free to diffuse throughout the simulation box, while diffusions of membrane-bound receptors were confined to the plasma membrane. A periodic boundary condition was applied for membrane receptors along both the x and y directions of the surface. The same 2D periodic boundary conditions were imposed in the extracellular region for biologics, and they were not allowed to move below the plasma membrane. If any biologics moved beyond the top of the simulation box, it bounced back. The amplitude and probability of the translational and rotational movements of each molecule were determined by its corresponding diffusion constant. To capture the internal conformational variations between MHC and PD-L1, an additional operation for each biologics module was implemented after diffusion to generate a small, random perturbation along their three translational and three rotational degrees of freedom.

Specifically, the probability and corresponding amplitudes of these perturbations were derived from the Anton MD simulation results, as shown in **Figure 4**. The histograms in the figure indicate the probability distributions of translational and rotational variations around the average conformation of each functional module based on the MD sampling. They were first converted to a cumulative distribution by adding up all these probabilities from 0 to 1 so that the probability of conformational variations with specific amplitude could be related to the corresponding range in this cumulative distribution. A random number was then generated from a

uniform distribution between 0 and 1. The value of the random number was used to determine the amplitude of variations by checking the cumulative distribution range into which it fell. At each simulation time step, this procedure was carried out for each functional module to select the amplitudes of variations along its translational and rotational degrees of freedom. After various amplitudes were applied to change the positions and orientations of both functional modules in all biologics, their internal conformations were effectively changed.

Reactions between biologics and receptors followed diffusions as the second scenario of the algorithm. Association is triggered if the distance between binding sites of the specific functional modules in a biologics and its corresponding receptor is below a predetermined cutoff value, and the probability of triggering association is determined by the association rate. In contrast, the dissociation between a pair of molecules is triggered by a probability that is calculated by the given values of their association rate and binding affinity. At the next time step, a pair of previously dissociated molecules has two possible scenarios: if their distance is still below the cutoff, reassociate as a geminate recombination, and if not, diffuse farther away from each other. The parameters of association rates and binding affinities are either calculated by a residue-based kinetic Monte Carlo simulation, as described below, or adopted from the experimental measurements in the literature. If one functional module in a biologics binds to a receptor, the entire biologics will move with the receptor as a single unit on a plasma membrane. While the bound module is attached to its receptor, its position and orientation change with the receptor while maintaining its internal conformation changes. In short, the probability distributions derived from the MD simulation will randomly perturb the position and orientation of the other vacant module, as described above, giving it the flexibility and kinetic freedom to seek its local conformational space and thereby be accessible for binding to its corresponding receptors on the plasma membrane. As this diffusion-reaction process iterates in both Cartesian and compositional spaces, the system will finally reach equilibrium. Finally, the reliability of the model to simulate the binding of soluble ligands and cell-surface receptors has been validated by a simpler testing system. Details about the model validation can be found in the following section and **Figure S1**.

Calculation of association rates using residue-based kinetic Monte Carlo simulations

The kinetic Monte Carlo algorithm (Xie et al., 2017) was used to simulate the associations of TCR and MHC and PD-1 and PD-L1 with a coarse-grained model of protein structures. Each residue in the model was represented by a Cα atom plus the representative center of its side-chain, which was selected based on the specific properties of the amino acid. The simulation started with an initial conformation in which two molecules were placed randomly, whereas their corresponding binding interfaces were separated under the range of a given distance cutoff d_c. Following the initial conformation, each protein diffused randomly within one simulation step. A physics-based scoring function guided the diffusions of proteins during these simulations. The scoring function contained terms that evaluated the electrostatic interaction and hydrophobic effects between proteins. Based on the calculated energy, the Metropolis criterion was applied to determine the probability of accepting the corresponding diffusional movements. The simulation trajectory was terminated if an encounter complex formed at the end of each simulation step through the corresponding interface. Otherwise, this simulation procedure was repeated until it reached the maximal time duration.

Practically, this simulation algorithm was performed in parallel under different distance cutoffs to effectively estimate the association rate. Given a specific distance cutoff value, 10^3 trajectories were carried out. Each trajectory started from a relatively different initial conformation but with the initial distances between the binding interfaces of ligands and receptors in all trajectories below this cutoff value. Encounter complexes were successfully formed in some of these 10^3 trajectories, while proteins diffused away from each other at the end of other trajectories. By counting how many encounter complexes formed among these trajectories, the value of the association rate was effectively calculated, as described by the statistical analysis in our previous study (Wang et al., 2018). By integrating these rate parameters into the RB-based model, we could realistically simulate the dynamics of how multi-specific biologics target their corresponding receptors on cell surfaces. To link the residue-based and the RB-based simulations under the same timescale, the maximal time duration to terminate each residue-based simulation trajectory is fixed to equal the time step in the RB simulation. In detail, each trajectory of residue-based simulation consists of 10^3 steps, and each step is 0.01 ns so that the total simulation time for each trajectory is 10 ns, which is a single time step in the RB simulation.

System setup and detailed protocol of Anton all-atom MD simulations

The structural models of the bispecific fusion protein were constructed by computationally fusing two protein X-ray structures with a peptide linker. The MHC module, consisting of 375 amino acids, was adopted from the PDB ID 3NWM (Samanta et al., 2011), while the other module PD-L1 contained two immunoglobulin domains and was adopted from the PDB ID 4Z18 (Pascolutti et al., 2016). The target peptide in the groove of the original MHC was not modeled in the system to avoid instability during simulations. The C-terminus of the MHC light chain was connected to the N-terminal domain of PD-L1 by a peptide linker. Specifically, three types of linker with various lengths and sequences were constructed to provide a comprehensive test. As specified in the **Results**, the linker GS30 contained six copies of GGGGS fragments. Because there is no side-chain in glycine, this type of linker is assumed to have a high degree of flexibility. As a result, the initial structures of the linker were built by ModLoop (Fiser and Sali, 2003). The other two linkers, PLP15 and PLP30, were poly-prolines, which can form well-defined secondary structures as either more compact right-handed α-helices or more extended left-handed α-helices (Kumar and Bansal, 2016). Traditionally, the left-handed poly-proline is relatively more rigid and is thus more common in structural biology studies (Doose et al., 2007). As a result, the initial structures of linkers PLP15 and PLP30 in this study were built following the standard configuration of a left-handed α-helix.

All the equilibrium simulations of constructed fusion proteins were run on the Anton 2 supercomputer at the Pittsburgh Supercomputing Center (Shaw et al., 2009). Each system contained an average number of 2.52×10^5 atoms. Proteins were solvated with water molecules and neutralized by adding Na+ and Cl– ions. **Figures 2a**–**2c** show their initial structures. Systems were embedded into orthorhombic cells with approximate dimensions of 125×120×160 Å with periodic boundary conditions. All production runs took the NPT ensemble with constant pressure (1 atm) and physiological temperature (310 K) using the Nose-Hoover thermostat. We chose the CHARMM36m force field for proteins and the TIP4P-D water model (Piana et al., 2015). This water model enables more accurate simulation of the dynamics of proteins with intrinsic flexible linkers due to its increased water-dispersion interactions (Henriques et al., 2015). The Gaussian-split Ewald algorithm (Shan et al., 2005) was used to compute the long-range electrostatic interactions with a 64×64×64 Å mesh. The cutoff for short-range non-bonded interactions was at least 11 Å for all boxes. All the other system-optimized simulation parameters

in the Anton software were chosen by default. Finally, a 2.5 µs trajectory was collected from the Anton 2 supercomputer for each system.

Model validation for the rigid-body-based simulation of binding between ligands and cell surface receptors

In order to evaluate the reliability of our rigid-body-based diffusion-reaction simulation method, we applied it to simulate a simple system in which only one type of receptors is expressed on membrane surface. They can form interactions with extracellular ligands which appear as monomeric form. We assume that this simple binding process between membrane receptors and soluble ligands can be described by the chemistry of surface adsorption, which has been well formulated by Langmuir Isotherm (Eckl and Gruler, 1988) as $\Theta = \frac{\alpha[M]}{1+\alpha[M]}$. In the Langmuir equation, Θ is the fraction of bound receptors on the 2D surface, [M] is the concentration of ligands in solvent. The Langmuir adsorption constant α is related to the binding equilibrium constant K_D as $\alpha = 1/K_D$. As a result, the Langmuir equation indicates that there is a linear relation between the reciprocal of the surface fraction versus the reciprocal of the solvent concentration. In order to test if this linear relationship can be reflected by our model. We carried out simulations under different concentrations of soluble ligands. The size of the simulation box was fixed with a volume of $500 \times 500 \times 200$ nm^3 and number of surface receptors was fixed as 100. The binding affinity of ligand-receptor interaction was fixed as -10kT. For each tested concentration, the average value of surface fraction was calculated, as defined by the percentage of surface receptors that form complexes with soluble ligands from simulations. **Figure S1** presents the plots between the reciprocal of solvent concentration and the surface fraction. The figure shows that the reciprocal of the surface fraction and the reciprocal of the solvent concentration clearly form a linear relation. The Pearson correlation coefficient (PCC) from the linear regression is 0.995. Moreover, the slope of this linear curve indicates the equilibrium constant of binding. By fitting the data in the curve, we derived the equilibrium constant as 110 µM, corresponding to an effective affinity of −9.1kT. This effective affinity from fitting the simulation results is very close to the original testing affinity of -10kT. Altogether, the linear relation between the reciprocal of the surface fraction and solvent concentration plus the successful reproduction of the equilibrium constant indicating the reliability of using the

diffusion-reaction method to simulate the binding between membrane-bound receptors and soluble ligands.

Characterize the impacts of different linkers on conformational coupling between two functional modules

In order to characterize the impacts of different linkers on the relative bending between MHC and PD-L1, the torsional angle between these two functional modules were derived from the MD simulation trajectories. The torsional angle can be determined as follow. Firstly, four representative points were selected from the biologics: the center of mass of MHC; the starting residue of the linker; the ending residue of the linker; and the center of mass of PD-L1. After the selection of these points, three vectors were built by connecting these points with each other. The torsional angle Φ is then defined as the dihedral formed by these three vectors, as shown in **Figure S2a**. The range of torsional angle is between -180° and +180°. Based on above definition, we calculated the value of angle Φ for all snapshots in the MD simulations. This calculation was carried out for all biologics that contain one of the three different linkers, and the distributions of calculated Φ values were plotted as histograms. The difference in these distributions among linkers are compared with each other in **Figure S2b**. The figure shows that linker GS30 forms a uniform distribution within the range from -180° to +180° (red bars). In contrast, multiple peaks were observed in the distribution of linker PLP15 (black bars), indicating the packing preference between two functional modules. Additionally, the linker PLP30 (blue bars) form a relatively less uniform distribution than GS30. Based on this figure, we speculate that a linker with more uniform distribution in torsional angle may have a lower bias in searching binding partners for the second module once the first one is bound. Consequently, we can further assume that the biologics with this type of linker can bind to more receptors on cell surface. Interestingly, this assumption can be validated by our rigid-body-based simulation results, in which we showed that the GS30 linker resulted in the highest number of biologics targeted to both TCR and PD-1, while the linker PLP15 led to the lowest number of biologics with both ligand MHC and PD-L1 bound to their receptors

Sensitivity analysis of how changes of binding parameters affect the results of rigid-body-based
diffusion-reaction simulations

In order to explore the sensitivity of our model to different binding parameters, we simultaneously changed both binding affinities between MHC and TCR, and between PD-L1 and PD-1. Both affinities were varied from -6kT to -14kT with an interval of 2kT. Therefore, a total number of 5×5=25 combinations were tested. For each combination, simulation was carried out from a random configuration with 100 biologics distributed in extracellular region, plus 100 TCR and 100 PD-1 on the bottom of the simulation box with a volume of $500×500×200$ nm^3. The length of each simulation trajectory is $1×10^8$ ns. The overall results are summarized in **Figure S3** as two-dimensional contour plots. The x axis and y axis in the figure stand for the binding affinities of TCR/MHC and PD-1/PD-L1 interactions, respectively. The number of interactions formed between cell surface TCR and MHC in the biologics is represented by the color index of the contour in **Figure S3a**, with red stands for the highest number and blue for the lowest number. Similarly, the color index of the contour in **Figure S3b** gives the number of interactions formed between cell surface PD-1 and PD-L1 in the biologics, while the color index of the contour in **Figure S3c** gives number of biologics that bound to both TCR and PD-1 on cell surface. The contours show that the TCR/HMC and PD-1/PD-L1 interactions are mutually affected due the spatial tethering of two ligands in the bispecific biologics. More specifically, the TCR/HMC and PD-1/PD-L1 interactions can be positively enhanced with each other when their individual affinities are not too strong (shown by the upper-left corner in **Figure S3a** and lower-right corner in **Figure S3b**). On the other hand, when both affinities of TCR/HMC and PD-1/PD-L1 interactions become strong (the upper-right corners in **Figure S3c**), a negative coupling effect was observed between these two interactions. This sensitivity analysis is consistent with the results in which a mutation strengthening the PD-L1/PD-1 binding can weaken the interaction between TCR and MHC. Detailed explanation about the possible mechanisms underlying our observation is provided in the main text.
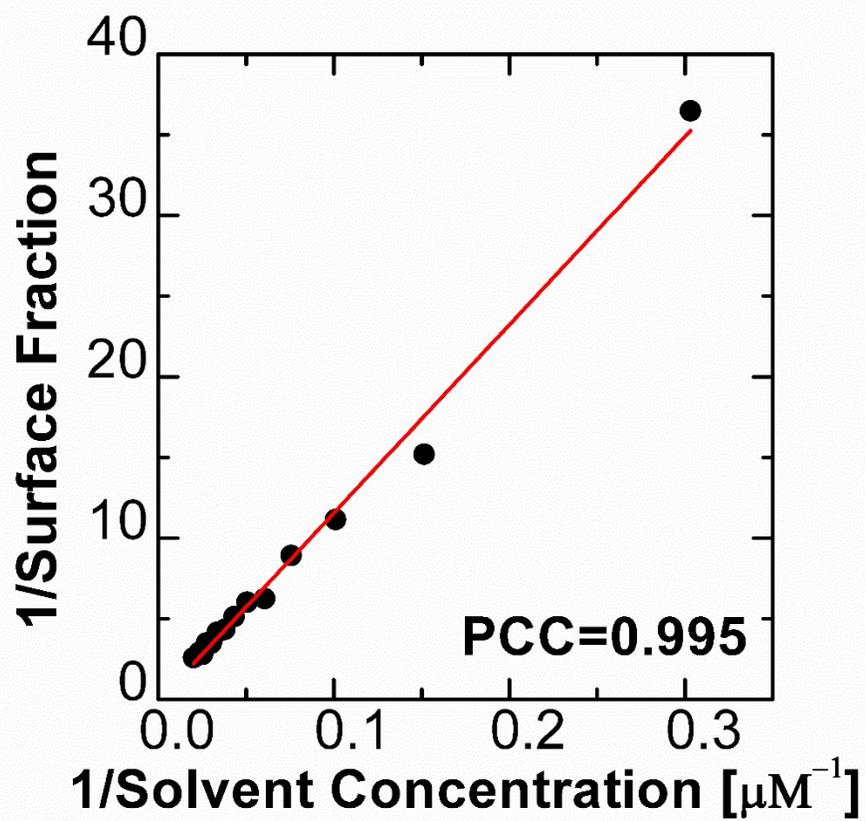
Figure S1: Model validation for the rigid-body-based simulation of binding between ligands and cell surface receptors, related to Figure 2.
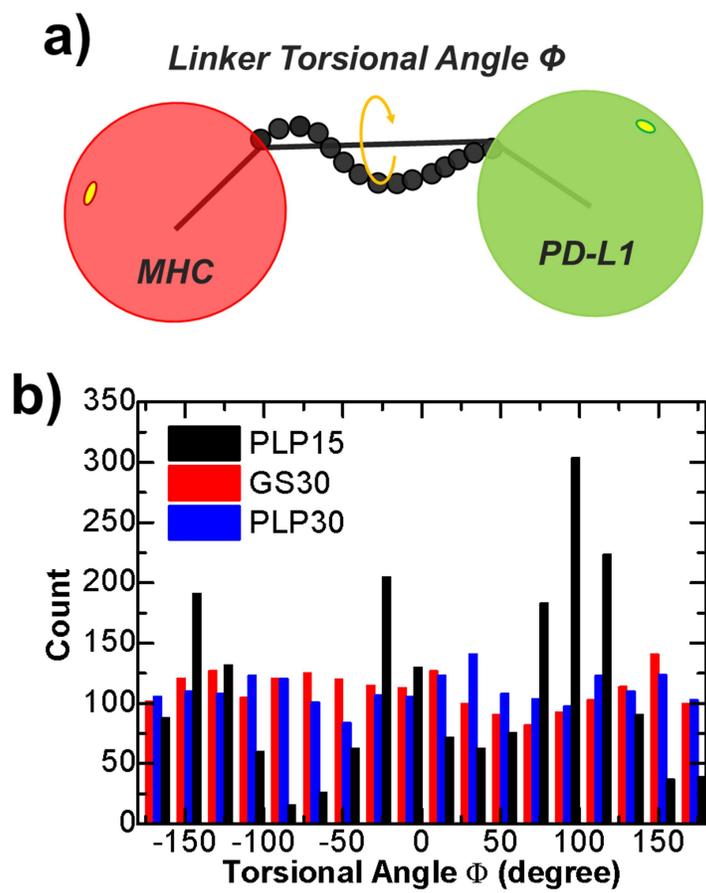
Figure S2: Characterize the impacts of different linkers on conformational coupling between two functional modules, related to Figure 3.
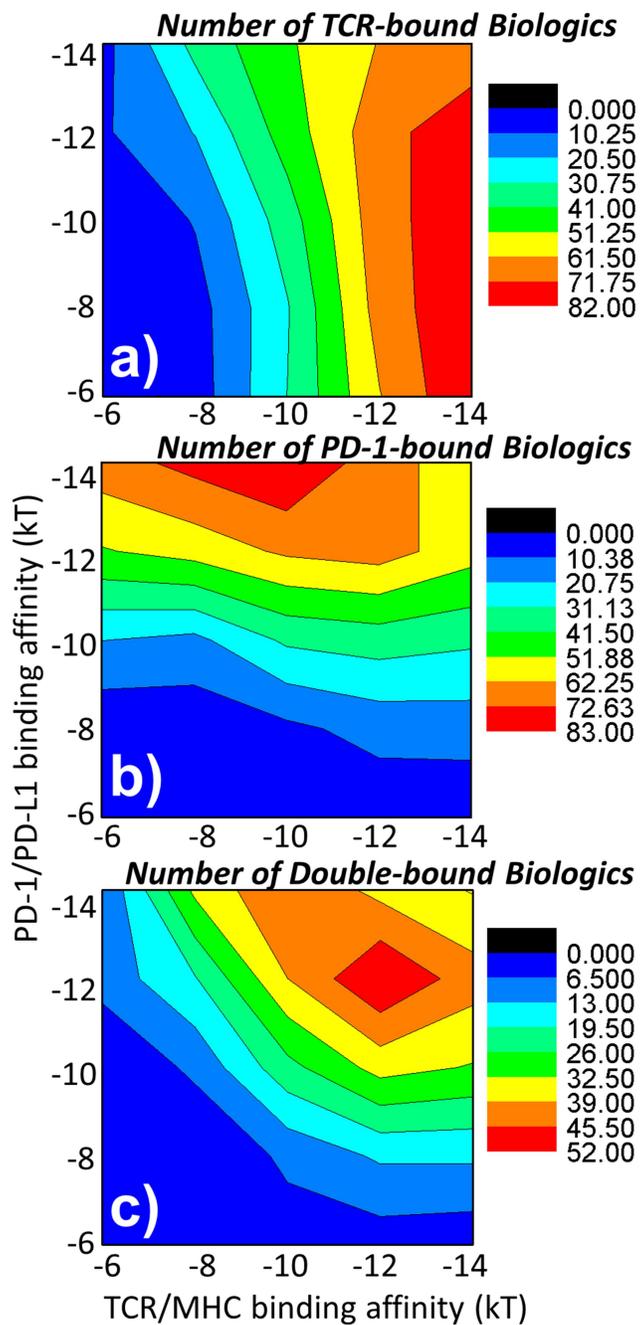
Figure S3: **Sensitivity analysis of our model to the changes of binding affinities**, related to Figure 7.

Supplemental References

Doose, S., Neuweiler, H., Barsch, H. & Sauer, M. (2007). Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc Natl Acad Sci U S A,* 104**,** 17400-5.

Eckl, K. & Gruler, H. (1988). DESCRIPTION OF CELL-ADHESION BY THE LANGMUIR ADSORPTION-ISOTHERM. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences,* 43**,** 769-776.

Fiser, A. & Sali, A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics,* 19**,** 2500-1.

Henriques, J., Cragnell, C. & Skepo, M. (2015). Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J Chem Theory Comput,* 11**,** 3420-31.

Kumar, P. & Bansal, M. (2016). Structural and functional analyses of PolyProline-II helices in globular proteins. *J Struct Biol,* 196**,** 414-425.

Pascolutti, R., Sun, X., Kao, J., Maute, R. L., Ring, A. M., Bowman, G. R. & Kruse, A. C. (2016). Structure and Dynamics of PD-L1 and an Ultra-High-Affinity PD-1 Receptor Mutant. *Structure,* 24**,** 1719-1728.

Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. (2015). Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B,* 119**,** 5113-23.

Samanta, D., Mukherjee, G., Ramagopal, U. A., Chaparro, R. J., Nathenson, S. G., Dilorenzo, T. P. & Almo, S. C. (2011). Structural and functional characterization of a single-chain peptide-MHC molecule that modulates both naive and activated CD8+ T cells. *Proc Natl Acad Sci U S A,* 108**,** 13682-7.

Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O. & Shaw, D. E. (2005). Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys,* 122**,** 54101.

Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Bank, J. A., Young, C., Deneroff, M. M., Batson, B., Bowers, K. J., Chow, E., Eastwood, M. P., Ierardi, D. J., Klepeis, J. L., Kuskin, J. S., Larson, R. H., Lindorff-Larsen, K., Maragakis, P., Moraes, M. A., Piana, S., Shan, Y. & Towles, B. (2009). Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis.* Portland, Oregon: ACM.

Wang, B., Xie, Z. R., Chen, J. & Wu, Y. (2018). Integrating Structural Information to Study the Dynamics of Protein-Protein Interactions in Cells. *Structure*.

Xie, Z.-R., Chen, J. & Wu, Y. (2014). A coarse-grained model for the simulations of biomolecular interactions in cellular environments. *Journal of Chemical Physics,* 140**,** 054112.

Xie, Z. R., Chen, J. & Wu, Y. (2017). Predicting Protein-protein Association Rates using Coarse-grained Simulation and Machine Learning. *Sci Rep,* 7**,** 46622.