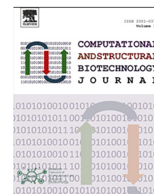
Computational and Structural Biotechnology Journal 20 (2022) 3473–3481

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www elsevier com/locate/csbj



A computational algorithm to assess the physiochemical determinants of T cell receptor dissociation kinetics

Zachary A. Rollins [b], Jun Huang [d], Ilias Tagkopoulos [c], Roland Faller [b], Steven C. George [a,]

[a] Department of Biomedical Engineering
[b] Department of Chemical Engineering
[c] Department of Computer Science
[d] University of California, Davis, Davis, California, Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL

RTICLE INFO

Article history:
Received 27 April 2022
Received in revised form 20 June 2022
Accepted 21 June 2022
Available online 25 June 2022

Keywords:
T cell receptor
Peptide major histocompatibility complex
Immunogenicity
Steered molecular dynamics
Machine learning

BSTR CT

The rational design of T Cell Receptors (TCRs) for immunotherapy has stagnated due to a limited understanding of the dynamic physiochemical features of the TCR that elicit an immunogenic response. The physiochemical features of the TCR-peptide major histocompatibility complex (pMHC) bond dictate bond lifetime which, in turn, correlates with immunogenicity. Here, we: i) characterize the force-dependent dissociation kinetics of the bond between a TCR and a set of pMHC ligands using Steered Molecular Dynamics (SMD); and ii) implement a machine learning algorithm to identify which physiochemical features of the TCR govern dissociation kinetics. Our results demonstrate that the total number of hydrogen bonds between the CDR2β-MHC (β), CDR1 -Peptide, and CDR3β-Peptide are critical features that determine bond lifetime.

2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Main

T cell-based immunotherapies (e.g., chimeric antigen receptor-T, or CAR-T; and TCR-engineered-T, or TCR-T) have provided transformative therapeutic responses in a small subset of cancers and patients (1–5); however, progress in solid tumors has been agonizingly slow. For example, CAR-T cells require an antigen on the tumor cell surface, but the majority (85%) of identified neoantigens are intracellular (6) and thus are immunogenic only when a representative fragment is presented on the cell surface in a peptide-major histocompatibility complex (i.e., pMHC). Although TCR-T therapy is MHC-restricted, this approach can target intracellular antigens, and the remarkable sensitivity of a TCR to recognize a single pMHC molecule (7) provides an additional strategic advantage. Nonetheless, identifying neoepitopes, matching these with immunogenic TCRs, and minimizing off-target effects remain significant challenges to implementation of these therapies (8).

Recent reports demonstrate that single-cell sequencing and machine learning technologies can identify patient- and tumor-specific neoepitopes (9,10). However, identification of partner TCRs remains challenging, despite the fact that tumor-specific T cells can be found in the peripheral blood (11,12). The human immune system generates tumor-specific T cells in a process that begins with random V(D)J recombination to create the hypervariable regions of the TCR and β chains. While this process generates a stunningly large number of *possible* TCRs (>10^{20}-10^{61}) (13,14), including 10^{6}-10^{8} in the peripheral blood, it is inherently inefficient and does not necessarily produce a TCR with appropriate immunogenicity for a given tumor (15). Alternate strategies of TCR identification have also fallen short; for example, TCR affinity enhancement can lead to a loss of TCR specificity (16,17) and does not always determine immunogenicity (18).

Computational techniques such as steered molecular dynamics (SMD) and machine learning may enable the creation of highly immunogenic, tumor-specific TCRs through rapid and efficient screening of the vast number of possible TCRs. The success of these techniques depends on accurate *in vitro* predictions of T cell immunogenicity, a goal that remains elusive. Quantitative descriptors of the TCR-pMHC bond identified in previous studies do not consistently correlate with immunogenicity (18–21). The majority of these studies measured equilibrium parameters of the TCR-pMHC bond (e.g., affinity), which do not account for the non-equilibrium mechanical forces on the TCR-pMHC bond present *in vivo*. Recent studies using DNA-based tension probes have estimated this force at 10–20 pN (22,23), and subsequent studies demonstrate that dissociation kinetics (i.e., bond lifetime) of the

Corresponding author at: Department of Biomedical Engineering, 451 E. Health Sciences Drive, room 2315, University of California, Davis, Davis, CA 95616.
E-mail address: scgeorge@ucdavis.edu (S.C. George).

https://doi.org/10.1016/j.csbj.2022.06.048
2001-0370/ 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

TCR-pMHC bond at this physiologic force can predict immunogenicity (24–31). These correlations are consistent across species, TCR-pMHC pairs, and experimental systems (24–31). Importantly, force-dependent bond lifetime represents an alternative hypothesis to affinity with no straightforward integration.

Here, we seek to discern the atomic-level physiochemical features that determine the TCR-pMHC bond lifetime under force (i.e., characterize the TCR-pMHC's force-dependent dissociation kinetics). As a first attempt to manipulate the bond lifetime of the TCR-pMHC over a wide range and to develop a novel computational methodology, we characterized the force-dependent dissociation kinetics of a single TCR (with a known crystal structure) to 17 possible pMHCs using steered molecular dynamics (SMD). Then, we used several machine learning algorithms, including linear regression, to identify the physiochemical features and the specific regions of the TCR regulating bond lifetime. The dataset for this initial study is limited due to the computational cost of atomistic molecular dynamic simulations (i.e., we utilized 350,000 core-hours used to accrue this dataset). Simulations were performed on a high perforamce computing cluster with two 8-core CPUs running at 2.4 GHz. Although this modest dataset is limited to a single TCR, this methodology sets precedence for an encompassing study of a multitude of known TCR-pMHC structures which will require a significant allocation on one of the world's largest supercomputers. Nonetheless, our results provide intriguing insight into the determinants of the TCR-pMHC bond strength and demonstrate that the total number of hydrogen bonds (H-bonds) between the CDR2β-MHC (β), CDR1 -Peptide, and CDR3β-Peptide are critical features that determine bond lifetime for the DMF5 TCR. This finding may inform the rational design of TCRs for TCR-T cell therapy, and provides a path forward to create more advanced and predictive machine learning algorithms.

2. Methods

2.1. Molecular Dynamics setup

The crystal structure of the human DMF5 TCR complexed with agonist pMHC MART1-HLA-A2 (PDB code: 3QDJ) (32) was the initial structure for all simulations (Fig. 1**A**). To generate the 17 TCR-pMHC pairs, amino acid substitutions were made to the MART1 peptide (AAGIGILTV) using the Mutagenesis plugin on Pymol Molecular Graphics System (Schrödinger, New York, New York). A property distance index (PD) was calculated to determine peptide amino acid sequence similarity to MART1 (SDAP, https://fermi.utmb.edu/SDAP/) (33) (**Table S1**). Interfacial substructures (Fig. 1**B**) were defined by sequential residues from the corresponding chains: TCR (CDR1 : 24–32, CDR2 : 50–55, CDR3 : 89–99), TCRβ (CDR1β: 25–31, CDR2β: 51–58, CDR3β: 92–103), MHC (MHC (β): 50–85, MHC ( ): 138–179), and peptide (1–9). To determine protonation states, pKa values were calculated using propka3.1 (34,35) and residues were considered deprotonated in Gromacs (36) if pKa values were below the physiological pH 7.4. The resulting systems were solvated using the TIP3P water model (37) in rectangular water boxes large enough to satisfy the minimum image convention. Na^+ and Cl^- ions were added to neutralize protein charge and reach physiologic salt concentration of 150 mM. All simulations were performed with Gromacs 2019.1 (36) using the CHARMM 22 plus CMAP force field for proteins (sometimes referred to as CHARMM 27) (38) and orthorhombic periodic boundary conditions. All simulations were in full atomistic detail.

2.2. Energy minimization and equilibration

Generating equilibrated starting structures for the Steered Molecular Dynamics simulations required four steps: (1) Steepest descent energy minimization to ensure correct geometry and the absence of steric clashes; (2) 100 ps simulation in the constant volume (NVT) ensemble to bring atoms to correct kinetic energies, while maintaining temperature at 310 K by coupling all protein and non-protein atoms to separate baths using the velocity rescale thermostat with a 0.1 ps time constant (39), (3) 100 ps simulation in the constant pressure (NPT) ensemble using Berendsen pressure coupling (39) and a 2.0 ps time constant to maintain isotropic pressure at 1.0 bar; and (4) Production MD simulations conducted for 50–150 ns with no restraints. The protein structures were evaluated every 50 ns to determine if all protein chains were equilibrated by root mean square deviation. To ensure true NPT ensemble sampling during 100 ns production runs, the Nose-Hoover thermostat (40) and Parrinello-Rahman barostat (41) were used to maintain temperature and pressure, respectively. Time constants were 2.0 and 1.0 ps for pressure and temperature coupling, respectively, utilizing the isothermal compressibility of water, 4.5^{-5} bar^{-1}. Box size for equilibration was 10.627 × 7.973 x.10.685 nm^3 with 48,000 water molecules, 300 ions, and 157,000 total atoms. All simulation steps used the Particle Ewald Mesh algorithm (42,43) for long-range electrostatic calculations with cubic interpolation and 0.12 nm maximum grid spacing. Short-range non-bonded interactions were cut off at 1.2 nm using the Verlet cutoff-scheme and all bond lengths were constrained using LINCS algorithm (44). The leap-frog algorithm was used for integrating equations of motion with 2 fs time steps. After the preparation runs, three independent MD configurations for each peptide mutant were extracted and used as the three starting points for steered molecular dynamics simulations.

2.3. Steered Molecular Dynamics SMD

The full TCR-pMHC complex structure was extracted from the preparation run for each peptide mutant to generate three SMD starting configurations. The main axis of these protein complexes was aligned along the *x*-axis of the box and solvated in rectangular water boxes with dimensions 30 × 9.972 × 12.685 nm^3. Solvent was again represented by the TIP3P water model and Na^+ and Cl^- ions were added to neutralize protein charge and reach physiologic salt concentration of 150 mM. This resulted in 120,000 water molecules, 700 ions, and 370,000 total atoms. All Gromacs structure files are uploaded to a Dryad repository (https://doi.org/10.25338/B8R33G) for the exact atomic specifications. Before pulling, all systems underwent (1) energy minimization; (2) 100 ps NVT; and (3) and 100 ps NPT to remove high energy contacts without disturbing the configurations. During pull, the Nose-Hoover thermostat and Parrinello-Rahman barostat were used to maintain temperature and pressure. 500 pN linear potential was applied to the center of mass (COM) of the TCR and pMHC in the *x*-direction and simulations continued until distance between COMs reached 0.49 times the box size in *x*-direction (Fig. 1**A**). The COM was chosen as the site of applied force because pulling from the TCR and MHC termini resulted in artificial unfolding (45). The 500 pN pulling force was chosen because no substantial differences in root mean square fluctuations of the interfacial substructures were found between pulling at 10 pN and 500 pN (45). All simulation trajectories and selected frames were visualized using the Pymol Molecular Graphics System (Schrödinger, New York, New York).