Title

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

Authors

Rollins, Zachary A
Huang, Jun
Tagkopoulos, Ilias
et al.

Publication Date

2022

DOI

10.1016/j.csbj.2022.06.048

Peer reviewed

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

ARTICLE 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

ABSTRACT

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-MHCα(β), CDR1α-Peptide, and CDR3β-Peptide are critical features that determine bond lifetime.

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 α 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/

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-MHCα(β), CDR1α-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: TCRα (CDR1α: 24–32, CDR2α: 50–55, CDR3α: 89–99), TCRβ (CDR1β: 25–31, CDR2β: 51–58, CDR3β: 92–103), MHCα (MHCα(β): 50–85, MHCα(α): 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).

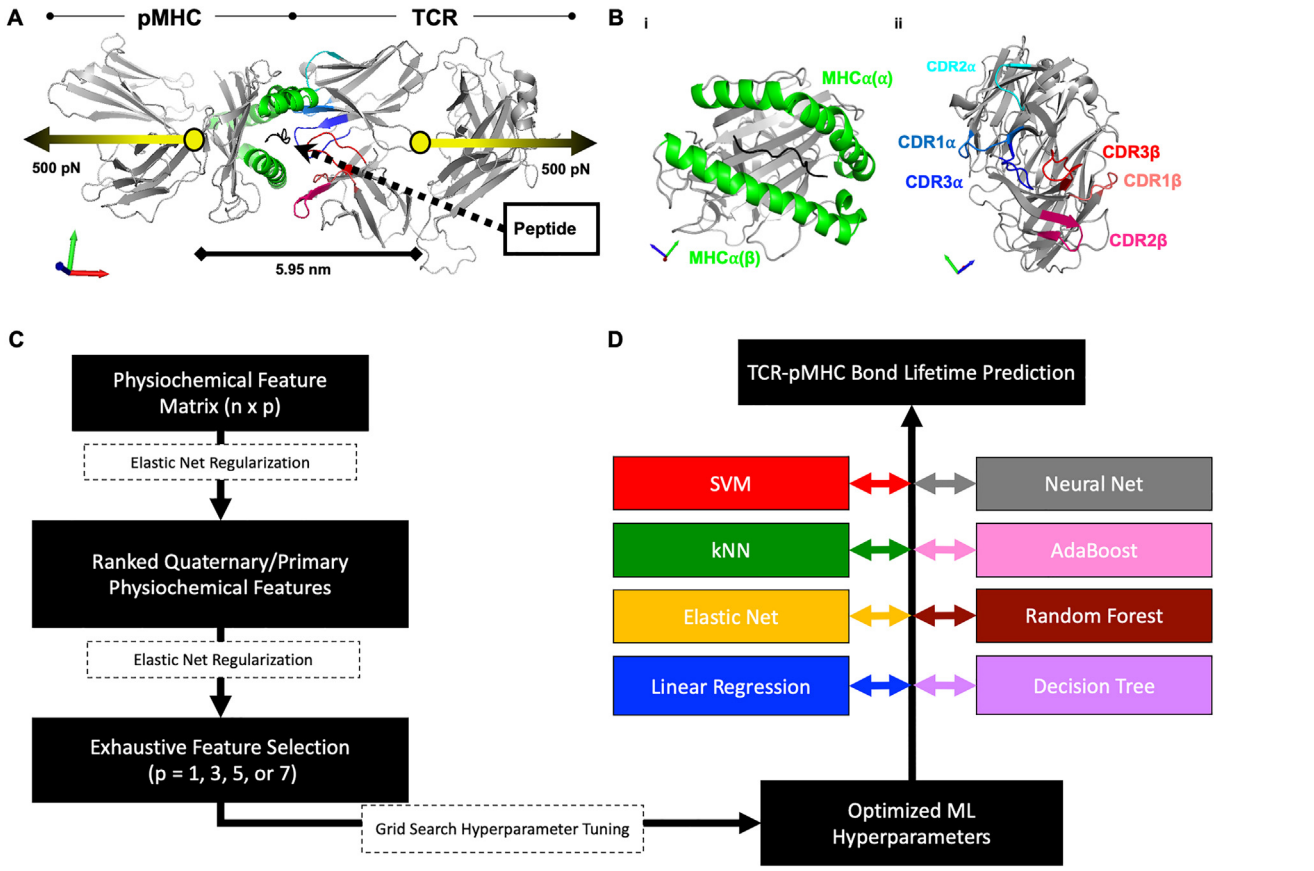

Fig. 1. Steered Molecular Dynamics (SMD) simulations and machine learning algorithms were used to identify the physiochemical features that predict TCR-pMHC bond lifetime. **(A)** Starting structure for SMD of TCR and pMHC (shown at the top with black lines and circle arrowheads). The location/direction of pulling are depicted with yellow circles/arrows, respectively; the black scale bar with diamond arrowheads denotes the definition of distance between centers of mass. The non-interacting bodies of the TCR and pMHC are colored in gray. Axes directions are indicated in the left corner (red: +x-direction, blue: +y-direction, and green: +z-direction). **(B)** The primary interfacial substructures: **(i)** MHCα(α) & MHCα(β) = green, Epitope = black; and **(ii)** TCR CDR1α = light blue, TCR CDR2α = cyan, TCR CDR2α = dark blue, TCR CDR1β = salmon, TCR CDR2β = light red, and TCR CDR3β = red. **(C)** A two-layer Elastic Net-Exhaustive Feature Selection algorithm (dashed boxes) was used to obtain ranked and reduced feature sets. **(D)** Selected features were used to tune hyperparameters (dashed box) for each machine learning model (Linear Regression = blue, Elastic Net = orange, k-Nearest Neighbors = green, Support Vector Machines = red, Decision Tree = purple, Random Forest = brown, AdaBoost = pink, Neural Net = gray). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Physiochemical descriptors and data analysis

Physiochemical descriptors were evaluated by defining Gromacs index groups (gmx make_ndx) and using Gromacs-suite analysis tools (i.e., gmx hbond, gmx rms, gmx rmsf, gmx sasa, gmx gyrate, gmx distance). The physiochemical features were calculated for the specified index groups and detailed descriptions of these calculations can be found in the Gromacs manual (36). The feature averages (e.g., SASA, RMSF, etc.) are the arithemetic average of triplicate SMD trajectories. Instananeuos features are averaged over all snapshots (instants) of the SMD trajectories at 10 picosecond time resolution. Data analyses were performed by standard python packages for data handling and visualization (i.e., numpy (46), pandas (47), seaborn (48), matplotlib (49), statistics (50), and GromacsWrapper (51), and custom python scripts. Random mutants were generated with a custom python script compatible with Pymol using the random python package and selecting a random location and amino acid to mutate the peptide. The machine learning algorithms were developed using the sklearn package (52,53) and exhaustive feature selection was performed using mlxtend package (54). The geometry of a Lennard-Jones contact (LJ-contact) is defined as a distance<0.35 nm between atoms. The L1 peptide bond lifetime was an outlier (z-score = 3.65 > 3). To reduce the effects of the outlier on the dataset, median absolute error was selected as the scoring criterion and L1 was excluded from correlation coefficient calculations. The mean absolute error represents the arithmetic average of median absolute error from repeated three-fold cross validation. The Pearson correlation coefficient (r_p) and Spearman rank correlation coefficients (r_s) were calculated using the correlation method in the pandas python package. Akaike and Bayesian Information Criterion (AIC and BIC) were calculated from the standard deviation of repeated three-fold cross validation of the best machine learning algorithm selected from the hyperparameter grid search. Statistical significance was determined by performing a one-tailed student's *t*-test ($p < 0.05$) for each machine learning algorithm across feature sets. Custom scripts relevant to mutant generation, feature selection, machine learning, and the production of figures have been made available in a GitHub repository: https://github.com/zrollins/TCR.ai.git.

2.5. Feature selection and Machine learning algorithms

Pearson and Spearmen correlation coefficients between all features are on the Dryad repository. Features were ranked and reduced utilizing a two-layer Elastic Net – Exhaustive Search algorithm (Fig. 1**C**). First, Elastic Net Regularization (55) was used with all physiochemical features and a grid search was performed to optimize hyperparameters. The optimized hyperparameters were

implemented into the Exhaustive Feature Selector (54) and the best individual features were ranked by repeated (n_repeats = 3) threefold cross-validation. The top ten features were ranked by mean absolute error and feature combinations were exhaustively searched, utilizing Elastic Net Regularization, to determine the best combinations of 3, 5, and 7 features (Fig. 1**C**). The best feature combinations were selected by mean absolute error arithmetically averaged over the cross-validation. These feature combinations were then implemented into several machine learning algorithms to determine the most predictive model of bond lifetime (Fig. 1**D**) (52,53). The machine learning algorithm hyperparameter optimization was performed on a high performance compute cluster at the University of California, Davis College of Engineering and the best model for each feature set was scored on absolute error and ranked by the arithmetic average of repeated threefold cross-validation (i.e., n_splits = 3, n_repeats = 3, random_state = 1). Detailed documentation regarding the cross validation and hyperparameter optimization of two-layer Elastic Net – Exhaustive Search feature selection and machine learning predictions are provided in the supporting information. In addition, this dataset has been made freely available on the Dryad repository (https://doi.org/10.25338/B8R33G).

3. Results

3.1. Bond lifetime

As the starting point to simulate the force-dependent dissociation kinetics of 17 TCR-pMHC pairs using SMD, we used the previously reported crystal structure (PDB ID: 3QDJ) (32) of the DMF5 TCR (from a melanoma patient) bound to the MART1 peptide (AAGIGILTV)-MHC complex (Fig. 1**A**). We then replaced the MART1 peptide with 16 different peptides (**Table S1**) for a total of 17 TCR-pMHC pairs. Ten peptides were chosen from a set of known pMHCs (56,57) and 7 were generated through random point mutation of the MART1 peptide. The similarity of peptides to the MART1 peptide is evaluated by a property distance index (PD) (33) (**Table S1**). The ten known and 7 unknown pMHCs have no known standard measure of T cell activation, therefore mean computational bond lifetime is used as a proxy. For these 17 TCR-pMHC pairs, the mean bond lifetime in the SMD simulations was 5400 ± 1700 picoseconds (Fig. 2).

3.2. Physiochemical features of the TCR-pMHC

Next, we identified two sets of physiochemical features which, at distinct resolution levels, describe the TCR-pMHC bond during the SMD simulation. The first set characterizes physiochemical features of the entire TCR-pMHC interaction (e.g., total H-bonds between the TCR and pMHC). This characterization provides an overall assessment of the physiochemical features that might impact bond lifetime and is consistent with the quaternary structure of globular proteins. We considered features likely to impact dissociation kinetics and thus included H-bonds (58), LJ- contacts (59), distance between the TCR and pMHC (60,61), solvent accessible surface area (SASA) (62), root mean square fluctuations (RMSF) (63), and the gyration tensor of the TCR and pMHC. This approach resulted in 18 features for the first set, and we dubbed these quaternary features (**Table S2**).

An understanding of the physiochemical features that regulate dissociation kinetics of the global TCR-pMHC bond provides an overall assessment of which physiochemical features regulate bond lifetime. However, this approach does not identify the subregions of the TCR-pMHC bond that regulate bond lifetime and thus are suitable targets for rational design of TCRs. The hypervariable regions of the TCR can be divided into 3 complementarity determining regions (CDRs) on the α and β chain, respectively. Within the MHC, the peptide is surrounded by α-helices which also interact with the nearby chains of the TCR (Fig. 1**B**). These MHC α-helices are located on the MHCα chain and these substructures are defined by their interaction with the TCR α and β chain, respectively (i.e., MHCα(α) and MHCα(β)). These TCR CDRs and MHC α-helices form an interface with the peptide antigen – the variable in this study – and based on their physical location are likely to influence TCR-pMHC bond lifetime. Hence, we also identified a second set of features focused on the interface between the TCR and the pMHC (e.g., CDR3α loop of the TCR and the MHCα(β) chain, Fig. 1**B**). This higher level of resolution is consistent with the secondary structures (e.g., α-helices) of a protein. Again, we considered features that are likely to affect dissociation kinetics and thus included H-bonds, LJ-contacts, distance between the subregions, SASA, RMSF, and the gyration tensor of the sub-regions. From these considerations, we identified 79 secondary features (**Table S3**) that could potentially impact dissociation kinetics. The quaternary and secondary features were further categorized into chemical – such as H-bonds and LJ-Contacts – and physical –

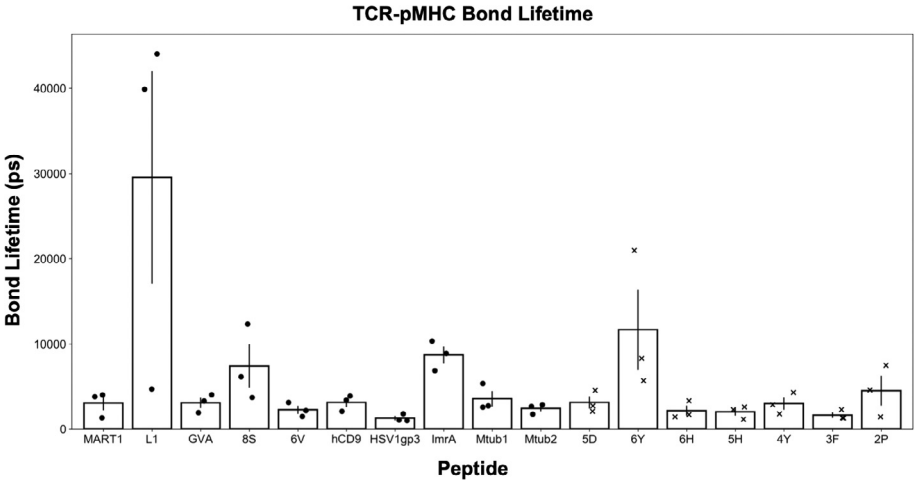

Fig. 2. TCR-pMHC bond lifetime for 17 different peptides. Using Steered Molecular Dynamics (SMD), we applied a constant force of 500 pN at the center of mass for the TCR and pMHC and estimated the bond lifetime for 17 different peptides. Known peptides and those with random point mutations are denoted with circles and and crosses, respectively. Each TCR-pMHC was pulled apart 3 times using different equilibrated structures. The bar height represents the mean bond lifetime and the error represents the standard error of measurement.

including RMSF, SASA, and the gyration tensor – interaction parameters.

3.3. TCR-pMHC bond lifetime prediction using quaternary physiochemical features

To examine how quaternary physiochemical features influence TCR-pMHC bond dissociation kinetics, we ranked the top ten quaternary features after an Elastic Net grid search for each individual feature (Fig. 3**A**). The scoring criterion was mean absolute error of bond lifetime in picoseconds. After Elastic Net grid search, chemical interaction features, in particular Total LJ-contacts and Total H-bonds, were the most predictive (Fig. 3**A**); in particular, the total number of unique LJ-Contacts between TCR and pMHC had the smallest mean absolute error. In addition, the total LJ-Contacts had the highest Pearson and Spearman correlation coefficients (**Figure S1**, **Table S4**).

We next explored whether a combination of quaternary physiochemical features would improve predictions of bond lifetime. To accomplish this, we applied a regularized regression method (Elastic Net; see **Methods**) as a filter to identify predictive feature sets. To avoid overfitting (64–66), feature sets were reduced utilizing an Elastic Net (55) – Exhaustive Search (54) algorithm (Fig. 1**C**) to determine the best combinations of 3, 5, and 7 features. Using these combinations, we then trained and tested 8 different machine learning algorithms to estimate TCR-pMHC bond lifetime (Fig. 1**D**) (52,53). Several algorithms, including simple linear regression (limited dataset), were hyperparameter searched and predictive performance was evaluated. Although physical quaternary features were selected in this exhaustive search (Fig. 3**B**), these did not significantly improve the predictive power of the machine learning models (Fig. 3**C**). This finding holds for all machine learning algorithms, as determined by the lack of statistically significant increase in mean accuracy or decrease in information criteria scores (Akaike and Bayesian Information Criteria) with increasing model complexity (**Figure S2**, **Table S5**).

The best feature combination and machine learning model was chosen based on the lowest error and standard deviation from

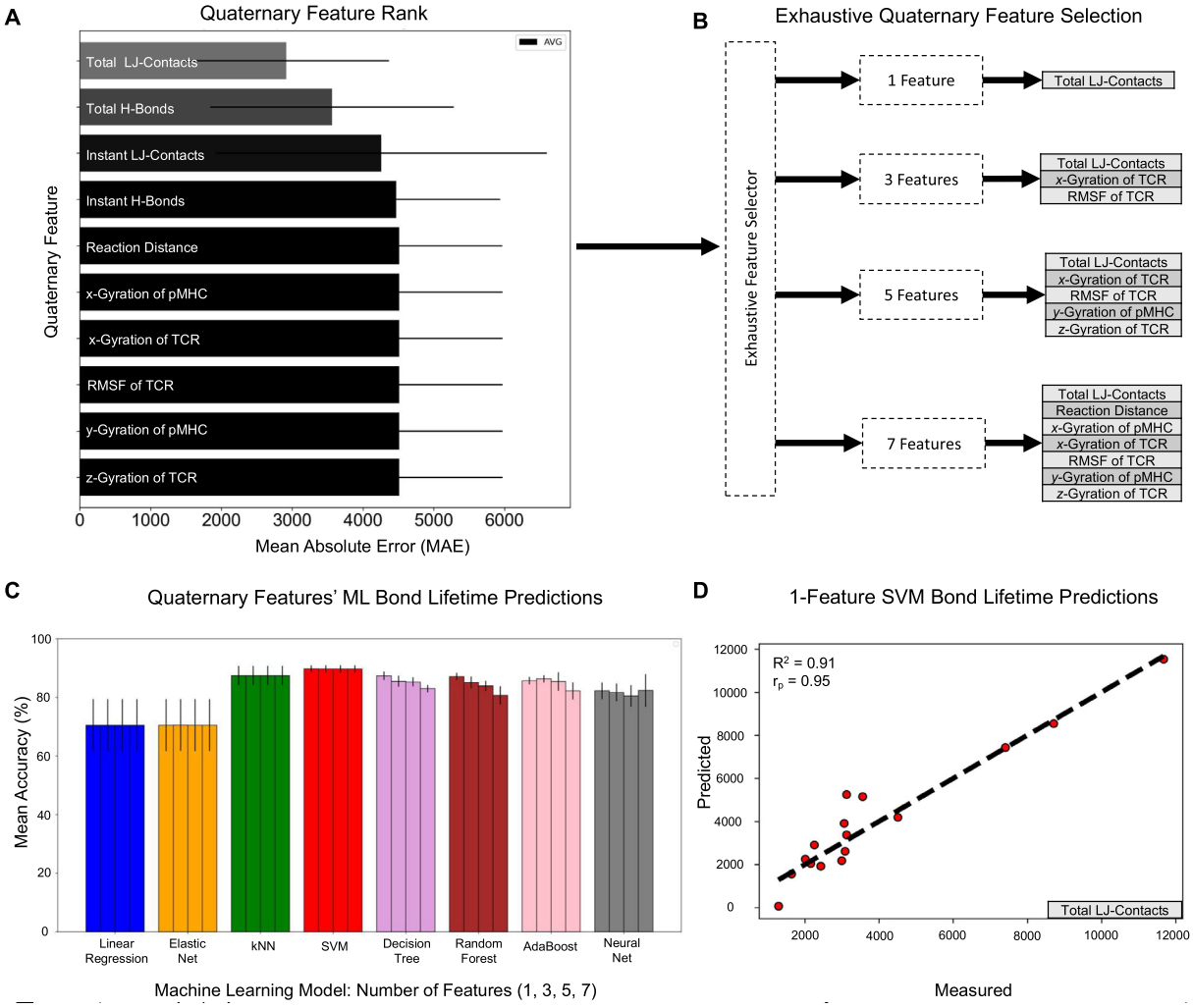

Fig. 3. Quaternary Feature Selection and Bond Lifetime Predictions. (A) Mean absolute test error from elastic net regularization was used to select the top ten quaternary features. Errors represent the best test set standard deviation from repeated threefold cross-validation. **(B)** According to an exhaustive search, the best feature sets (i.e., p = 1, 3, 5, and 7) to predict bond lifetime. **(C)** The mean accuracies of bond lifetime prediction for all feature sets in (B) and machine learning models after hyperparameter tuning (Linear Regression = blue, Elastic Net = orange, k-Nearest Neighbors = green, Support Vector Machines = red, Decision Tree = purple, Random Forest = brown, AdaBoost = pink, Neural Net = gray). The grouped bars represent the number of quaternary features included in increasing order (i.e., 1, 3, 5, and 7 features) for the respective machine learning model. Errors represent the best test set standard error from repeated threefold cross-validation. The machine learning model standard error from cross-validation (n = 9) was statistically compared for increasing feature sets by a one-tailed student's *t*-test: #p < 0.10, *p < 0.05, **p < 0.01. **(D)** The scatter plot of predicted and measured bond lifetimes from the selected one-feature Support Vector Machines algorithm with the coefficient of determination (top left), the Pearson correlation coefficient (top left), and the feature set (bottom right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

repeated three-fold cross-validation. Our results demonstrated that a feature set of only LJ-Contacts combined with a Support Vector Machines is best at predicting bond lifetime (Fig. 3**D**). The mean absolute error using Support Vector Machines was 560 ± 200 picoseconds producing an accuracy of 90.0 ± 3.7% (i.e., 1–560/5400).

3.4. TCR-pMHC bond lifetime prediction using secondary physiochemical features

Analogous to our strategy to assess quaternary features of the TCR-pMHC, we examined secondary features. We ranked the top ten secondary features after an Elastic Net grid search for each individual feature (Fig. 4**A**). The total number of unique H-bonds between CDR2β -MHCα(β) generated the smallest mean absolute error (Fig. 4**A**). In addition, the top three features had the highest Pearson and Spearman correlation coefficients (**Figure S3**, **Table S4**).

We explored whether a combination of secondary physiochemical features would improve the prediction of bond lifetime. Following the same algorithm as for quaternary features, we applied an Elastic Net (55) – Exhaustive Search (54) algorithm (Fig. 1**D** to identify the best combinations of 3, 5, and 7 secondary features; cross-validated 8 machine learning models with these feature combinations; and selected the best feature combination and machine learning model based on error, standard deviation, and information criteria. Interestingly, the best 3 feature combination (CDR2β -MHCα(β), CDR1α-Peptide, and CDR3β-Peptide) selected by exhaustive search (Fig. 4**B**) did not correspond to the top three individual features selected by Elastic Net rank (Fig. 4**A**) or correlation coefficients (**Table S4**). Compared to the single best feature, the best 3-feature combination statistically improved bond lifetime predictions for Linear Regression, k-Nearest Neighbors, Decision Tree, and Random Forest machine learning algorithms (Fig. 4**C**). Increases in mean accuracy were not statistically significant beyond 3 features (Fig. 4**C**, **Table S6**). Moreover, these algorithms reduced information criteria scores (Akaike and Bayesian

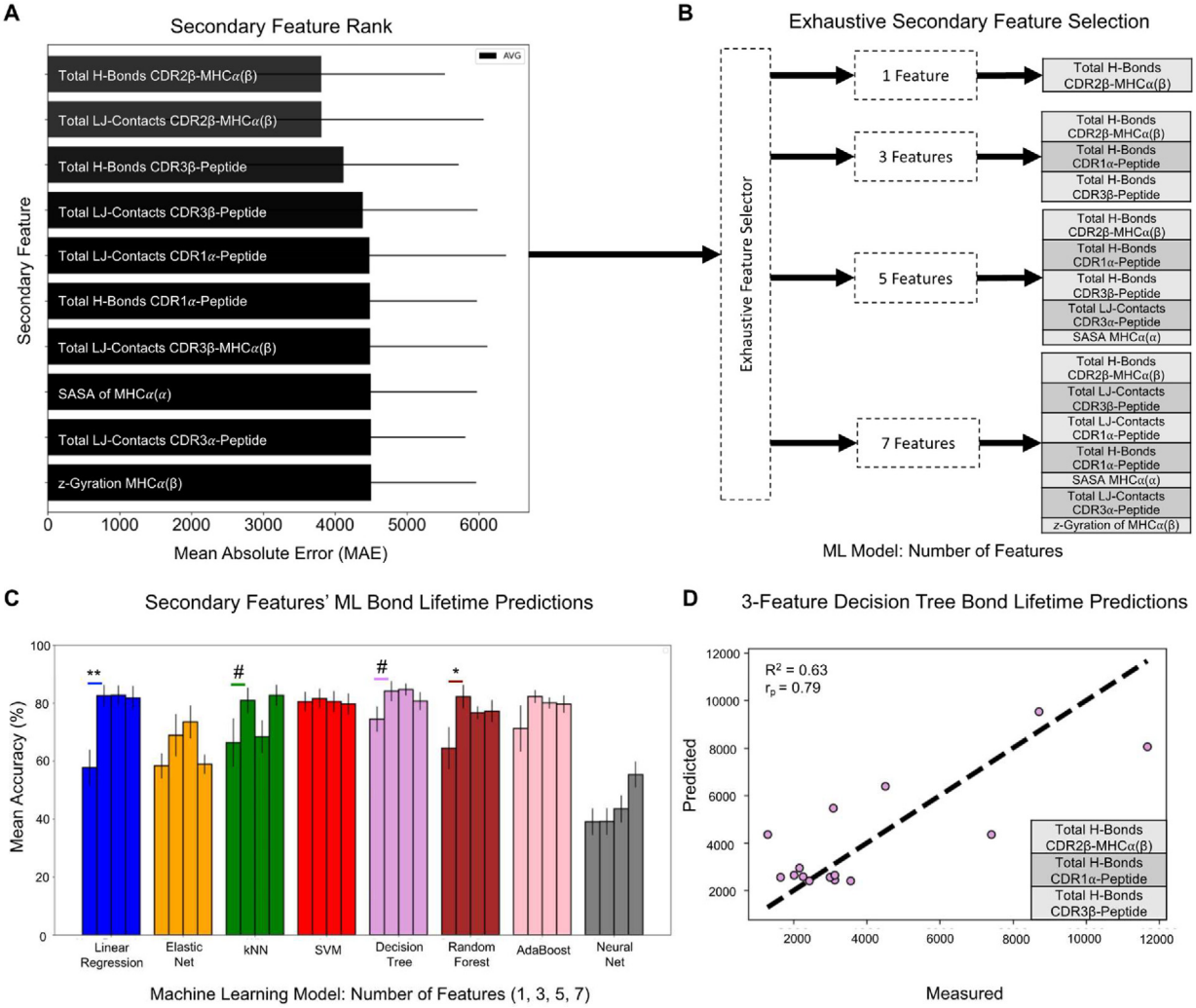

Fig. 4. Secondary Feature Selection and Bond Lifetime Predictions. (**A**) Mean absolute test error from elastic net regularization was used to select the top ten secondary features. Errors represent the best test set standard deviation from repeated threefold cross-validation. (**B**) According to an exhaustive search, the best feature sets (i.e., p = 1, 3, 5, and 7) to predict bond lifetime. **(C)** The mean accuracies of bond lifetime prediction for all feature sets in (B) and machine learning models after hyperparameter tuning (Linear Regression = blue, Elastic Net = orange, k-Nearest Neighbors = green, Support Vector Machines = red, Decision Tree = purple, Random Forest = brown, AdaBoost = pink, Neural Net = gray). The grouped bars represent the number of secondary features included in increasing order (i.e., 1, 3, 5, and 7 features) for the respective machine learning model. Errors represent the best test set standard error from repeated threefold cross-validation. The machine learning model standard error from cross-validation (n = 9) was statistically compared for increasing feature sets by a one-tailed student's *t*-test: #p < 0.10, *p < 0.05, **p < 0.01. **(D)** The scatter plot of predicted and measured bond lifetimes from the selected 3-feature Decision Tree algorithm with the coefficient of determination (top left), the Pearson correlation coefficient (top left), and the feature set (bottom right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Information Criteria) when increasing from 1 to 3 features, whereas the Elastic Net, Support Vector Machines, and Neural Net algorithms increased both AIC & BIC (**Figure S4**). These results indicate that, among the secondary features and machine learning algorithms tested, a 3-feature combination utilizing a Decision Tree provides the most accurate prediction of bond lifetime (Fig. 4**D**). The absolute error using the Decision Tree was 870 ± 570 picoseconds (**Table S6**), or an accuracy of 84 ± 10%. In addition, this Decision Tree prediction by the best 3 feature combination exceeded the Pearson correlation coefficient of the individual features (Fig. 4**D, Figure S4**).

4. Discussion

T cell-based immunotherapies, such as TCR-engineered-T cells, provide exciting potential to treat a wide range of cancers, including solid tumors. However, this potential has not been reached, due, in part, to the inability to rapidly and efficiently explore the vast TCR space to identify optimal tumor-specific TCRs. Experimental methods to design and test potential TCRs are expensive and slow, thus hindering throughput. In contrast, computational algorithms that utilize machine learning have enormous potential to rapidly interrogate the TCR space and identify a small number of candidates for more efficient experimental testing. We initiate this premise using SMD to create a small database of TCR-pMHC bond lifetimes, then created machine learning algorithms to predict bond lifetime based on quaternary and secondary features of the TCR-pMHC bond. Using the quaternary features, we found that total LJ-contacts could predict bond lifetime with 90% accuracy. More importantly, we also found that we could predict bond lifetime with an accuracy of 84% using only the total H-bonds between three subregions of the TCR-pMHC: CDR2β-MHCα(β), CDR1α-Peptide, and CDR3β-Peptide. Although these subregions may only apply to the DMF5 TCR, these results validate the methodology and identify new, unanticipated regions of the TCR to target in the rational design of TCRs for immunotherapy.

4.1. Quaternary features of the TCR-pMHC

Upon quaternary feature investigation, the LJ-Contacts between the TCR and pMHC dominated bond lifetime prediction. In fact, for all machine learning algorithms investigated, there was no statistically significant i) increase in mean accuracy when expanding to larger feature sets (Fig. 3**C**) or ii) decrease in information criteria scores (**Figure S4**). Moreover, although physical features (e.g., *x*-Gyration of TCR) were selected in the exhaustive feature selection process (Fig. 3**B**), these did not significantly increase mean accuracy. This demonstrates that no selected physical features improve predictive performance and thus the atomic motion of the TCR or pMHC is unlikely to regulate dissociation kinetics.

4.2. Secondary features of the TCR-pMHC

To identify the specific subregions of the TCR that determine the TCR-pMHC bond lifetime, we investigated the TCR-pMHC interface and included substructures, or secondary protein features, that defined the interaction (Fig. 1**B**). Physiochemical features within each substructure and between adjacent substructures (**Table S3**) were then evaluated to determine the best predictors of bond lifetime. Among the features and machine learning algorithms selected, a 3-feature combination of secondary features (CDR2β-MHCα(β), CDR1α-Peptide, and CDR3β-Peptide) was selected as the most accurate predictor of TCR-pMHC bond lifetime. This was based on: i) a decrease in information criteria score for 5 of 8 machine learning algorithms; and ii) a statistically significant increase in mean accuracy for 4 of 8 machine learning algorithms when increasing the feature set size from 1 to 3. We found that the combination of total H-bonds between these subregions could predict bond lifetime with the highest accuracy.

The finding that both the total number of unique H-bonds between CDR3β-Peptide and CDR1α-Peptide predict TCR-pMHC bond lifetime is consistent with known DMF5-MART1 structural immunology. There are a reported three H-bonds between the CDR3β-Peptide compared to zero H-bonds between the CDR3α-Peptide. Similarly, a reported two H-bonds between the CDR1α-Peptide and one H-bond between the CDR1β-Peptide (32). Our finding that the number of unique H-bonds between CDR3β-Peptide and CDR1α-Peptide are predictive of bond lifetime is consistent with known structural immunology and serves as validation of the methodology.

The finding that the unique H-bonds between the CDR2β-MHCα(β) predict TCR-pMHC bond lifetime is unanticipated. For example, the DMF5-MART1 crystal structure reports one H-bond between both the CDR2β-MHCα(β) and CDR2α-MHCα(α). However, the H-bonds between CDR2β-MHCα(β) remained in all exhaustive search feature sets (Fig. 3**B**) whereas the H-bonds between CDR2α-MHCα(α) was not selected as a predictive feature of bond lifetime. This insight demonstrates the utility of identifying interfacial substructures that may be manipulated to effect TCR-pMHC bond lifetime. Most attention has been focused on the heralded CDR3 domains (67) given the proximity to the peptide (Fig. 1**A, B**). In contrast, CDR2 flanks the MHCα(α) and MHCα(β) domains. It is perhaps not surprising, given the significantly larger number of residues (MHCα(β) = 42 residues) compared to the peptide (peptide = 9 residues), that interactions between the CDR2β and the MHCα(β) could potentially be the most energetically significant physiochemical features to impact bond lifetime. This is consistent with our previous study demonstrating that mutations to the MART1 peptide alter the conformation of the MHCα(α) and MHCα(β) resulting in increased coulombic potential between the TCR and MHC (45).

Overall, these results suggest that mutagenesis strategies to increase hydrogen bonding between CDR2β-MHCα(β), CDR1α-Peptide, and CDR3β-Peptide may enhance TCR-pMHC force-dependent bond lifetime. It is important to acknowledge that the interactions between these interfacial substructures may be specific to the DMF5 TCR and will require further investigation to generalize. Nonetheless, these results increase attention to the CDR2 regions in the future of TCR design. Finally, in contrast to previous reports (28,59), peptide radius of gyration and CDR3α-CDR3β distance were not selected in the top ten predictive features. This is likely due to the artificial pMHC unfolding by pulling from TCR-pMHC termini [27] and the lack of diversity in TCR-pMHC pairs evaluated (28,59), respectively.

4.3. Computational methods

One of the limiting factors of this study is the computational constraint of generating a SMD dataset; here, we examined 17 TCR-pMHC pairs. Larger datasets would likely provide more useful insight into feature combinations that predict TCR-pMHC bond lifetime, but come at a significant additional computational cost. Similarly, although the two-layer Elastic Net – Exhaustive Search feature selection methodology provided a rapid filtering of physiochemical features, this biases the machine learning predictor towards features selected by Elastic Net. At the cost of computation, exhaustive or recursive feature selection for each machine learning predictor may improve predictive performance. However, the focus of this work is to provide an architecture for identifying physiochemical features that dictate TCR-pMHC dissociation kinetics.

4.4. Bond lifetime

The force dependent bond lifetime (at ~ 10–20 pN) has been reported to correlate with TCR-pMHC immunogenicity. These findings highlight the importance of TCR-pMHC bond lifetime and suggest that the TCR needs to sustain and form transient bonds under load for sufficient time to initiate biochemical signaling. Thus, we utilized force-dependent bond lifetime as an objective function to uncover the physiochemical determinants of this biomolecular design feature. It is important to note that this biomolecular design feature does not necessarily conflict with catch-slip bond behavior (24), and we recognize that our approach may be expanded in the future to include other physiochemical characteristics of the TCR-pMHC bond.

4.5. Conclusions

We have demonstrated the utility of combining two computational methods – steered molecular dynamics and machine learning – to create a methodology that can be used to examine the physiochemical features of the TCR-pMHC bond that predict force-dependent bond lifetime. Future applications of this work may inform TCR mutagenesis strategies to target neopeptide-MHCs in solid tumors. Our initial results suggest that the physiochemical features of three subregions of the TCR-pMHC are of particular importance in determining bond lifetime (CDR2β-MHCα(β), CDR1α-Peptide, and CDR3β-Peptide) for the DMF5 TCR and provide new and unanticipated regions of the TCR to manipulate in the rational design of TCR-engineered T cells.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Simulations were performed on the hpc1/hpc2 clusters in the UC Davis, College of Engineering. This work was supported in part by startup funding to SCG from the Department of Biomedical Engineering.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.06.048.

References

[1] Johnson LA, Morgan RA, Dudley ME, Cassard L, Yang JC, Hughes MS, et al. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. Blood 2009;114(3):535–46.
[2] Linette GP, Stadtmauer EA, Maus MV, Rapoport AP, Levine BL, Emery L, et al. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. Blood 2013;122(6):863–71.
[3] Moore T, Wagner CR, Scurti GM, Hutchens KA, Godellas C, Clark AL, et al. Clinical and immunologic evaluation of three metastatic melanoma patients treated with autologous melanoma-reactive TCR-transduced T cells. Cancer Immunol Immunother 2018;67(2):311–25.
[4] Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. Science 2006;314(5796):126–9.
[5] Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, et al. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. J Clin Oncol 2011;29(7):917–24.
[6] Weekes MP, Antrobus R, Lill JR, Duncan LM, Hör S, Lehner PJ. Comparative analysis of techniques to purify plasma membrane proteins. Journal of biomolecular techniques : JBT 2010;21(3):108–15.
[7] Sykulev Y, Joo M, Vturina I, Tsomides TJ, Eisen HN. Evidence that a Single Peptide–MHC Complex on a Target Cell Can Elicit a Cytolytic T Cell Response. Immunity 1996;4(6):565–71.
[8] He Q, Jiang X, Zhou X, Weng J. Targeting cancers through TCR-peptide/MHC interactions. J Hematol Oncol 2019;12(1):139.
[9] Gartner JJ, Parkhurst MR, Gros A, Tran E, Jafferji MS, Copeland A, et al. A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types. Nature Cancer 2021;2(5):563–74.
[10] Kosaloglu-Yalcin Z, Lanka M, Frentzen A, Logandha Ramamoorthy Premlal A, Sidney J, Vaughan K, et al. Predicting T cell recognition of MHC class I restricted neoepitopes. Oncoimmunology. 2018;7(11):e1492508.
[11] Dijkstra KK, Cattaneo CM, Weeber F, Chalabi M, van de Haar J, Fanchi LF, et al. Generation of Tumor-Reactive T Cells by Co-culture of Peripheral Blood Lymphocytes and Tumor Organoids. Cell 2018;174(6):1586–98 e12.
[12] Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat Med 2016;22(4):433–8.
[13] de Greef PC, Oakes T, Gerritsen B, Ismail M, Heather JM, Hermsen R, et al. The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes. Elife 2020;9.
[14] Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R. Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. Front Immunol 2013;4:485.
[15] Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. Genome Res 2011;21(5):790–7.
[16] Kunert A, Obenaus M, Lamers CHJ, Blankenstein T, Debets R. T-cell Receptors for Clinical Therapy. In Vitro Assessment of Toxicity Risk. Clin Cancer Res 2017;23(20):6012–20.
[17] Mensali N, Myhre MR, Dillard P, Pollmann S, Gaudernack G, Kvalheim G, et al. Preclinical assessment of transiently TCR redirected T cells for solid tumour immunotherapy. Cancer Immunol Immunother 2019;68(8):1235–43.
[18] Kersh GJ, Kersh EN, Fremont DH, Allen PM. High- and low-potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. Immunity 1998;9(6):817–26.
[19] van der Merwe PA, Davis SJ. Molecular interactions mediating T cell antigen recognition. Annu Rev Immunol 2003;21:659–84.
[20] Rudolph MG, Wilson IA. The specificity of TCR/pMHC interaction. Curr Opin Immunol 2002;14(1):52–65.
[21] Zhu C, Jiang N, Huang J, Zarnitsyna VI, Evavold BD. Insights from in situ analysis of TCR-pMHC recognition: response of an interaction network. Immunol Rev 2013;251(1):49–64.
[22] Liu Y, Blanchfield L, Ma VP, Andargachew R, Galior K, Liu Z, et al. DNA-based nanoparticle tension sensors reveal that T-cell receptors transmit defined pN forces to their antigens for enhanced fidelity. Proc Natl Acad Sci U S A 2016;113(20):5610–5.
[23] Ma R, Kellner AV, Ma VP, Su H, Deal BR, Brockman JM, et al. DNA probes that store mechanical information reveal transient piconewton forces applied by T cells. Proc Natl Acad Sci U S A 2019;116(34):16949–54.
[24] Liu B, Chen W, Evavold BD, Zhu C. Accumulation of dynamic catch bonds between TCR and agonist peptide-MHC triggers T cell signaling. Cell 2014;157(2):357–68.
[25] Liu B, Chen W, Natarajan K, Li Z, Margulies DH, Zhu C. The cellular environment regulates in situ kinetics of T-cell receptor interaction with peptide major histocompatibility complex. Eur J Immunol 2015;45(7):2099–110.
[26] Kolawole EM, Andargachew R, Liu B, Jacobs JR, Evavold BD. 2D Kinetic Analysis of TCR and CD8 Coreceptor for LCMV GP33 Epitopes. Front Immunol 2018;9:2348.
[27] Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, et al. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. Cell 2018;174(3):672–87 e27.
[28] Wu P, Zhang T, Liu B, Fei P, Cui L, Qin R, et al. Mechano-regulation of Peptide-MHC Class I Conformations Determines TCR Antigen Recognition. Mol Cell 2019;73(5):1015–27 e7.
[29] Das DK, Feng Y, Mallis RJ, Li X, Keskin DB, Hussey RE, et al. Force-dependent transition in the T-cell receptor beta-subunit allosterically regulates peptide discrimination and pMHC bond lifetime. Proc Natl Acad Sci U S A 2015;112(5):1517–22.
[30] Robert P, Aleksic M, Dushek O, Cerundolo V, Bongrand P, van der Merwe PA. Kinetics and mechanics of two-dimensional interactions between T cell receptors and different activating ligands. Biophys J 2012;102(2):248–57.
[31] Limozin L, Bridge M, Bongrand P, Dushek O, van der Merwe PA, Robert P. TCR-pMHC kinetics under force in a cell-free system show no intrinsic catch bond, but a minimal encounter duration before binding. Proc Natl Acad Sci U S A 2019;116(34):16943–8.
[32] Borbulevych OY, Santhanagopolan SM, Hossain M, Baker BM. TCRs used in cancer gene therapy cross-react with MART-1/Melan-A tumor antigens via distinct mechanisms. J Immunol 2011;187(5):2453–63.

[33] Ivanciuc O, Midoro-Horiuti T, Schein CH, Xie L, Hillman GR, Goldblum RM, et al. The property distance index PD predicts peptides that cross-react with IgE antibodies. Mol Immunol 2009;46(5):873–83.
[34] Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput 2011;7(2):525–37.
[35] Sondergaard CR, Olsson MH, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput 2011;7(7):2284–95.
[36] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701–18.
[37] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79 (2):926–35.
[38] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B 1998;102(18):3586–616.
[39] Berendsen HJC, Postma JPM, Gunsteren WFv, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81 (8):3684–90.
[40] Evans DJ, Holian BL. The Nose-Hoover thermostat. J Chem Phys 1985;83 (8):4069–74.
[41] Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 1981;52(12):7182–90.
[42] Di Pierro M, Elber R, Leimkuhler B. A Stochastic Algorithm for the Isobaric-Isothermal Ensemble with Ewald Summations for All Long Range Forces. J Chem Theory Comput 2015;11(12):5624–37.
[43] Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Annalen der Physik. 1921;369(3):253-87.
[44] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 1997;18(12):1463–72.
[45] Rollins ZA, Faller R, George SC. T Cell Receptor Non-Equilibrium Kinetics. bioRxiv. 2021:2021.10.27.466112.
[46] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585(7825):357–62.
[47] McKinney W. Data structures for statistical computing in python. Proceedings of the 9th Python in Science Conference. 2010;445.
[48] Waskom MB, Olga; O'Kane, Drew; Hobson, Paul; Lukauskas, Saulius; Gemperline, David C; Augspurger, Tom; Halchenko, Yaroslav; Cole, John B; Warmenhoven, Jordi; de Ruiter, Julian; Pye, Cameron; Hoyer, Stephan; Vanderplas, Jake; Villalba, Santi; Kunter, Gero; Quintero, Eric; Bachant, Pete; Martin, Marcel; Qalieh, Adel. mwaskom/seaborn: v0.8.1. 0.8.1 ed. Meyrin, Switzerland: Zenodo; 2017.
[49] Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng 2007;9 (3):90–5.
[50] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261–72.
[51] Beckstein OD, Jan; Somogyi, Andy. GromacsWrapper: v0.3.3 (release-0.3.3). 0.3.3 ed. Meyrin, Switzerland: Zenodo; 2015.
[52] Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al., editors. API design for machine learning software: experiences from the scikit-learn project. European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases; 2013 2013-09-23; Prague, Czech Republichttps://hal.inria.fr/hal-00856511/document.
[53] https://hal.inria.fr/hal-00856511/file/paper.pdf.
[54] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12 (null):2825–30..
[55] Raschka S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. The Journal of Open Source Software 2018;3(24):638.
[56] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005;67(2):301–20.
[57] Rivoltini L, Squarcina P, Loftus DJ, Castelli C, Tarsini P, Mazzocchi A, et al. A superagonist variant of peptide MART1/Melan A27–35 elicits anti-melanoma CD8+ T cells with enhanced functional characteristics: implication for more effective immunotherapy. Cancer Res 1999;59(2):301–6.
[58] Hellman LM, Foley KC, Singh NK, Alonso JA, Riley TP, Devlin JR, et al. Improving T Cell Receptor On-Target Specificity via Structure-Guided Design. Mol Ther 2019;27(2):300–13.
[59] Solbach F, Bernardi A, Bansal S, Budamagunta MS, Krep L, Leonhard K, et al. Determining structure and action mechanism of LBF14 by molecular simulation. J Biomol Struct Dyn 2021;1–12.
[60] Hwang W, Mallis RJ, Lang MJ, Reinherz EL. The alphabetaTCR mechanosensor exploits dynamic ectodomain allostery to optimize its ligand recognition site. Proc Natl Acad Sci U S A 2020;117(35):21336–45.
[61] Welch DA, Woehl TJ, Park C, Faller R, Evans JE, Browning ND. Understanding the Role of Solvation Forces on the Preferential Attachment of Nanoparticles in Liquid. ACS Nano 2016;10(1):181–7.
[62] Huang Y, Harris BS, Minami SA, Jung S, Shah PS, Nandi S, et al. SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction. Biophys J 2022;121(1):79–90.
[63] Xiong Y, Karuppanan K, Bernardi A, Li Q, Kommineni V, Dandekar AM, et al. Effects of N-Glycosylation on the Structure, Function, and Stability of a Plant-Made Fc-Fusion Anthrax Decoy Protein. Frontiers. Plant Sci 2019;10(768).
[64] Martínez L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLoS ONE 2015;10(3):e0119264.
[65] Trunk GV. A problem of dimensionality: a simple example. IEEE Trans Pattern Anal Mach Intell 1979;1(3):306–7.
[66] McLachlan GJ. Discriminant analysis and statistical pattern recognition 2004; xv:526 pp..
[67] Zollanvari A, James AP, Sameni R. A Theoretical Analysis of the Peaking Phenomenon in Classification. J Classif 2020;37(2):421–34.