Biophysical Journal
Article

Biophysical Society

Integration of kinetic data into affinity-based models for improved T cell specificity prediction

Zahra S. Ghoreyshi,[1,2] Hamid Teimouri,[2,3] Anatoly B. Kolomeisky,[2,3,4,*] and Jason T. George[1,2,*]
[1]Department of Biomedical Engineering, Texas A&M University, College Station, Texas; [2]Center for Theoretical Biological Physics, Rice University, Houston, Texas; [3]Department of Chemistry, Rice University, Houston, Texas; and [4]Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas

ABSTRACT T cell receptor (TCR) and peptide-major histocompatibility complex (pMHC) interactions that result in T cell activation are complex and have been distinguished by their equilibrium affinity and kinetic profiles. While prior affinity-based models can successfully predict meaningful TCR-pMHC interactions in many cases, they occasionally fail at identifying TCR-pMHC interactions with low binding affinity. This study analyzes TCR-pMHC systems for which empirical kinetic and affinity data exist and prior affinity-based predictions have failed. We identify criteria for TCR-pMHC systems with available kinetic information where the introduction of a correction factor improves energy-based model predictions. This kinetic correction factor offers a means to refine existing models with additional data and offers molecular insights to help reconcile previously conflicting reports concerning the influence of TCR-pMHC binding kinetics and affinity on T cell activation.

SIGNIFICANCE This study develops a method to integrate available kinetic data on the TCR-pMHC interactions into biophysical affinity-based models. By analyzing available TCR-pMHC binding affinity and kinetic data, we identify empirical criteria for which the incorporation of kinetic data is expected to benefit predictions based solely on affinity data. The application of our kinetic correction factor benefitted model accuracy in several TCR-pMHC systems for which energy-based modeling previously failed to adequately explain favorable interactions. Our approach offers a refinement to existing biophysical models in cases where kinetic data are available for improved binding predictions, which is of general interest for predicting T cell activity in the setting of infectious disease, autoimmunity, and cancer immunotherapy.

INTRODUCTION

The adaptive immune response relies on T cell activation upon encountering antigenic peptide sequences. T cell specificity stems from the T cell receptor (TCR)'s selective recognition of antigenic peptides presented on the cell surface via major histocompatibility complex (MHC) molecules (1). The human T cell repertoire is comprised of a huge number ($\sim 10^8$) (2) of unique TCRs, which collectively confer broad immunity against a variety of antigenic peptides presented on MHC (pMHC). Central thymic selection and peripheral tolerance mechanisms train T cells to distinguish self from non-self signatures (3,4) and result in a TCR repertoire having variable specificities directed against a particular set of non-self antigens.

Reliable prediction of TCR specificity against antigens of interest remains an active area of research with broad implications for a better, more microscopic understanding of immune responses during infection, autoimmunity, and cancer. This problem is challenging because all available training data are sparse in the immense sequence space of TCR and antigen pairs. Several inferential modeling strategies have been proposed, including those based on TCR-peptide primary amino acid sequences (5–11). As an alternative, recently developed random energy models provided a theory-driven approach to understanding repertoire-level T cell tolerance and antigen recognition (12–15). These affinity-based models have subsequently led to the development of inferential biophysical approaches, which leverage known crystal structures to predict TCR-pMHC interactions based on pairwise amino acid potentials trained using available affinity (K_D) data on TCR-pMHC systems (16–18).

The above biophysical models perform well relative to sequence-based models when evaluated using distinct unseen TCRs during training. Incorporating structural

Submitted June 19, 2024, and accepted for publication November 4, 2024.
*Correspondence: tolya@rice.edu or jason.george@tamu.edu
Editor: Abhishek Singharoy.
https://doi.org/10.1016/j.bpj.2024.11.002

information permits reliable predictions using a small fraction of the data needed to train alternative models. Despite these significant advantages, affinity-based modeling still poorly explains some systems, and the molecular reasons for these observations are not well understood. TCR-pMHC interaction kinetics seem to be important in eliciting T cell responses (19–21), but relative to affinity measurements, there is a very limited amount of kinetic data (e.g., association and dissociation rate constants k_{on} and k_{off}) available for previously studied human TCR-pMHC systems. In some situations, more information is available on the kinetic responses of TCR-pMHC interactions. The observation of decreased performance of the affinity-based model in certain cases led us to investigate whether incorporating relevant kinetic information, where available, could improve the model predictions.

Here, we studied whether the relatively weaker performance in certain cases could be improved by the inclusion of kinetic information, such as effective TCR-pMHC association (k_{on}) and dissociation (k_{off}) rates. To analyze this, we developed a simple method to directly integrate kinetic data, where available, into our previously developed affinity-based models. This approach employs a correction factor derived from kinetic arguments to adjust the predicted affinity values, aiming to more accurately reflect the true binding dynamics observed in physiological conditions. We apply our integrated method to evaluate HLA-A*02-restricted 1G4 TCR binding to mutations of NY-ESO-1 peptide ligands (22). These ligands are currently under investigation as candidate antigens for antitumor vaccines aimed at enhancing immune responses against a wide range of tumors. Challenges in T cell activation, such as suboptimal TCR-pMHC binding kinetics or rapid dissociation, make NY-ESO-1 an ideal model for evaluating kinetic corrections. Although affinity-based models like Rapid Coarse-grained Epitope TCR (RACER), which rely solely on K_D, often fail to accurately predict specificity in this case, incorporating kinetic parameters like k_{on} and k_{off} improves immune response predictions and enhances model accuracy. Our findings demonstrate that the inclusion of the kinetic parameter k_{on} is required to resolve TCR specificity, while considering the thermodynamic K_D alone is insufficient. Moreover, the incorporation of k_{on} significantly improves the prediction accuracy in our affinity-driven model, which previously was ineffective at resolving NY-ESO-1-specific TCRs. This approach offers a method by which complementing affinity-driven modeling with kinetic information can dramatically enhance the quantitative description and our understanding of complex processes in immune response.

MATERIALS AND METHODS

TCR-pMHC binding dynamics

A TCR with concentration $[TCR]$ binds a peptide-bound major histocompatibility species that has concentration $[pMHC]$ to create a TCR-pMHC complex with concentration $[TCR : pMHC]$. This process can be described as a reversible chemical reaction,

$$TCR + pMHC \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} TCR : pMHC, \tag{1}$$

for which the equilibrium dissociation constant of this process (K_D) can be expressed as the ratio of corresponding equilibrium concentrations or the ratio of the association and dissociation rate constants,

$$K_D = \frac{[TCR]_{eq}[pMHC]_{eq}}{[TCR : pMHC]_{eq}} = \frac{k_{off}}{k_{on}}. \tag{2}$$

The equilibrium constant K_D provides a static thermodynamic view of the binding strength, while $1/k_{off}$ can be associated with the duration of time that the single TCR-pMHC complex remains intact before disassembly, which might be relevant for T cell signaling response. Therefore, the experimental values of K_D, k_{off}, and k_{on} that are used as input in our model are directly applied to evaluate the efficacy of different TCR-pMHC interactions, guiding the development of our computational models and the interpretation of their outputs.

Predictive structural affinity model

To accurately estimate the binding strength of a specific TCR-pMHC pair, we utilized our previously developed RACER pairwise energy framework (16,17). RACER calculates binding energies by integrating high-throughput data on previously confirmed TCR-peptide interactions and crystal structures to train a residue-specific energy matrix. The energy matrix and available structural templates are then used to quantify TCR-peptide binding. To accomplish this, RACER evaluates the binding interface between peptide antigen and variable (CDR3 α/β) regions via

$$V_{direct} = \sum_{\substack{i \in \text{TCR} \\ j \in \text{peptide}}} \gamma(a_i, a_j)\Theta^I_{ij}. \tag{3}$$

In Eq. 3, $\gamma(a_i, a_j)$ denotes the pairwise interaction strength between amino acids a_i and a_j at positions i and j within the indexed TCR and peptide, respectively. The entries of γ represent all symmetric pairwise amino acid energies, which are optimized during model training (23). Each interaction is weighted as a sigmoidal function Θ_{ij} of the pairwise distances between residues:

$$\Theta_{i,j} = \frac{1}{4}\left(1+\tanh\left[5\left(r_{i,j} - r_m\right)\right]\right)\left(1+\tanh\left[5\left(r_M - r_{i,j}\right)\right]\right), \tag{4}$$

where $r_m = 6.5$ Å (respectively $r_M = 8.5$ Å) defines the minimum (respectively maximum) distances over which the interaction remains significant.

RACER computes the energy of test TCR-pMHC pairs (E_{test}). These values are compared against energy values for an ensemble of randomized weak binding pairs (E_{weak}) binding TCR-pMHC pairs, and the free energy is defined relative to the mean of weak binders: $\Delta G \simeq E_{test} - \langle E_{weak}\rangle$. Experimentally obtained K_D and calculated approximated free energy changes ΔG are assumed to be related via

$$\ln K_D = \Delta G/k_B T. \tag{5}$$

To compare standardized free energy across distinct TCR-pMHC pairs and structural templates, the Z score is introduced as a normalized free energy parameter, defined as follows:

$$Z = \frac{-(E_{test} - \langle E_{weak}\rangle)}{\sqrt{\langle E^2_{weak}\rangle - \langle E_{weak}\rangle^2}} = \frac{-\Delta G}{\sigma_{weak}}. \tag{6}$$

Integrating kinetic data into the affinity model

To incorporate additional experimental kinetic data into the RACER framework, we modify the imputed K_D values. As one can see from Eq. 5, the parameters K_D and k_{off} are expected to vary proportionally across TCR-pMHC pairs that have comparable values for k_{on}. Previous work has, however, noted that the variations in on-rate k_{on} can significantly influence the overall interaction dynamics in the TCR-pMHC binding case (24). As we will demonstrate, the linear relationship between k_{off} and K_D breaks down in systems where k_{on} shows high variability. To account for this, we introduce a kinetic correction term, η_i, to modify the K_D values for TCR-pMHC pair i in cases where the on rate ($k_{on,i}$) is available according to the following rule:

$$\eta_i = \ln(k_{on,i}) - \ln(k_{on,0}). \quad (7)$$

In the above equation, cases where $k_{on,i}$ exhibits significant deviation away from typical values of $k_{on,0}$ of observed systems, η can be used to create an "effective dissociation equilibrium constant," $\tilde{K}_D$, according to

$$\ln(\tilde{K}_D) = \ln(K_D) - \eta_i. \quad (8)$$

The application of Eq. 7 relies on an empirical analysis of experimentally obtained k_{on} values for relevant cases. Our analysis (see the results) ultimately leads us to select a lower limit for $k_{on} \simeq 10^5$ below which the correction η is subsequently applied.

Data acquisition and analysis

Affinity data for 137 TCR-pMHC systems were sourced from the ATLAS database (25). All available crystal structures for HLA A*02-restricted systems were obtained via the RCSB Protein Data Bank (26). Additional kinetic data were obtained from a recent analysis on the NY-ESO-related peptides' interaction with the 1G4 TCR (24). This augmented set provided a total of 11 additional TCR-pMHC pairs (Tables S1 and S2). In the dataset used for this research, various TCR-pMHC interactions from the ATLAS dataset were analyzed alongside the binding interactions of the 1G4 TCR with its corresponding pMHC variant (24) using surface plasmon resonance (SPR) on a BIAcore 3000 system (GE Healthcare, Chicago, IL, USA). All experiments from the ATLAS dataset and (24) were conducted at 25°C and 37°C, respectively. Biotinylated pMHC molecules were immobilized onto a CM5 sensor chip (GE Healthcare) indirectly via covalently coupled streptavidin at varying levels.

RESULTS

Fig. 1 presents an overview of our approach, which involves integrating kinetic data in the form of a corrected affinity score. Our primary goal is motivated by the question of whether or not the incorporation of additional kinetic information into a structural model trained on affinity data could improve the predictions, particularly in cases where available affinity data are insufficient for explaining relevant TCR-pMHC interactions.

1G4-NY-ESO as a candidate TCR-pMHC system

The NY-ESO peptide and corresponding 1G4 TCR variants served as appropriate candidates as sufficient structural information exists for this case, yet despite this, RACER-m, a multitemplate approach within the RACER framework, was unable to correctly assess 50% of available cases even when they were explicitly included in training (17). An additional independent dataset generated affinity and kinetic data in the form of k_{on}, k_{off}, and K_D for additional 1G4-NY-ESO (TCR-pMHC) mutational variants (24). Plotting the effective off-rate k_{off} as a function of the equilibrium constant K_D for NY-ESO-associated cases demonstrated a statistically significant direct correlation between affinity and kinetic parameters, and expected inverse trends were also found for k_{on} versus K_D, respectively (Fig. 2, *a* and *b*). A significant inverse correlation was also observed between k_{on} and k_{off} (Fig. 2 *c*). However, this behavior was inordinately influenced by a minority of cases with large k_{on}. Removal of the three NY-ESO-mutant TCR-peptide pairs having the highest K_D values was enough to break

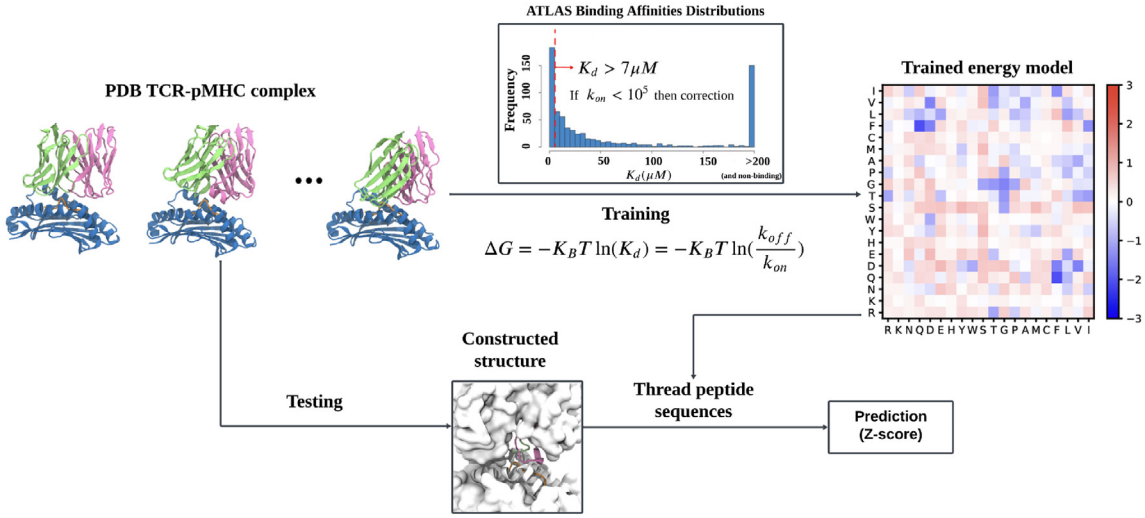

FIGURE 1 Schematic depiction of the training and testing phases within RACER, illustrating the impact of the k_{on} binding parameter on the trained energy model. For each TCR-pMHC pair with $K_D > 7$ μM and $k_{on} < 10^5$, a correction factor was applied.

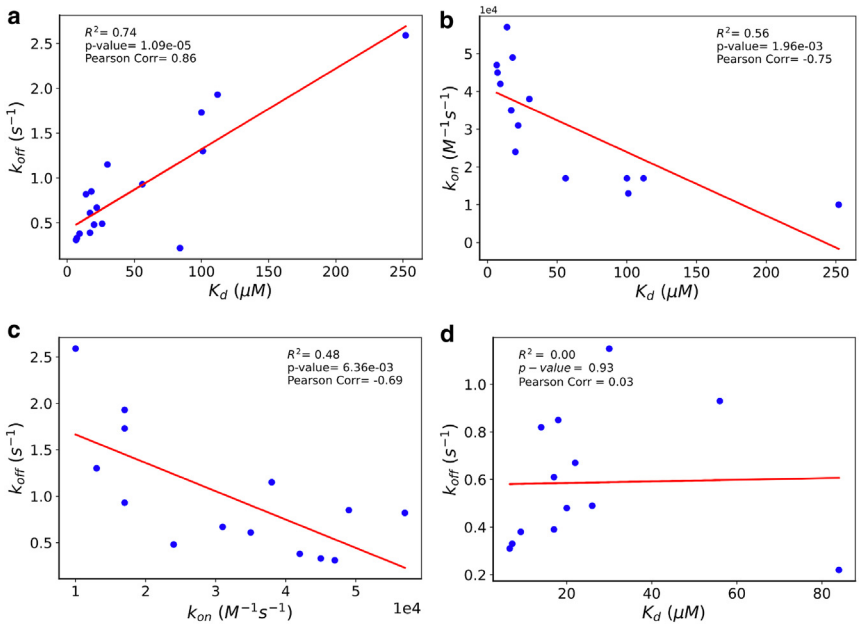

FIGURE 2 Correlation analyses of TCR-pMHC kinetic and affinity parameters for NY-ESO cases. This figure explores the interrelations between kinetic and affinity parameters across all NY-ESO cases. (*a*) demonstrates the correlation between dissociation constant K_D and dissociation rate k_{off}, illustrating their mutual dependencies. (*b*) examines the relationship between K_D and association rate k_{on}, assessing how association dynamics influence binding affinity. (*c*) analyzes the interplay between k_{off} and k_{on}, highlighting the balance of association and dissociation in binding kinetics. (*d*) shows the correlation between K_D and k_{off} specifically for a subset of NY-ESO cases.

the observed correlation between k_{off} and K_D (Fig. 2 *d*; Table S3). Remaining TCR-peptide pairs also exhibited high variability in k_{on}. In these cases, the proportional mapping between binding kinetics and affinity breaks down, thereby necessitating correction to account for kinetics in the affinity-trained model. Because the RACER-m model attempts to resolve meaningful TCR-peptide pairs based on an affinity threshold (defined to be $K_D \approx 7\ \mu M$), we hypothesized that cases with predicted K_D values exceeding this threshold, especially those with the smallest k_{on} values, would benefit significantly from a kinetic normalization of the affinity model's value. In the regime of low k_{on}, small fluctuations can have a large impact on the relationship between K_D and k_{off}. This supports the need for the kinetic correction term, which is further discussed in the results.

Selection of a kinetic correction term recovers identification of favorable interaction pairs

We applied a similar analysis on the subset of the ATLAS dataset for which TCR-pMHC systems had an affinity and kinetic data (Fig. 3). Again, significant correlations have been found in relations between k_{off} versus K_D, and the removal of cases with low k_{on} disrupts the observed correlation between k_{off} and K_D (Fig. 3, *a* and *b*). For these cases, when $k_{on} > 10^5$, the values of K_D and k_{off} are concentrated in a narrow range. Conversely, when $k_{on} < 10^5$, K_{off} and K_D are highly variable (Fig. 3, *c* and *d*). While this behavior appears to affect NY-ESO cases most strongly, additional TCR-pMHC systems from distinct functional clusters also exhibit significant variation in k_{off} and K_D, including TCRs specific for GILGFVFTL and ILAKFLHWL peptides. These findings are consistent with another report analyzing the 1G4-NY-ESO TCR-peptide, wherein the authors utilized a slightly different approach to account for widely variable k_{on} values (24).

Given the above findings, we now introduce a correction factor based on kinetic normalization into our affinity estimates (Eq. 8) for systems having low k_{on} (Eq. 7). The correction term effectively incorporates kinetic information in the regime of $k_{on} < 10^5$, for which the explicit incorporation of observable k_{off} is expected to reflect the true binding dynamics more accurately. To evaluate whether this correction term can successfully distinguish meaningful TCR-pMHC pairs, we incorporate this adjustment into the discrimination step based on standardized Z score (Eq. 6). The predicted Z scores as a function of $\log(K_D)$ are plotted both before (Fig. 4 *a*) and following (Fig. 4 *b*) correction for NY-ESO-specific TCRs. Using the previously established Z score cutoff, our correction demonstrates effectively recognized NY-ESO-specific TCRs, most of which would have been unidentified. This simple correction results in over 90% of cases being correctly identified. Specifically, for the six (two training, four test) NY-ESO cases in the RACER dataset, we correctly predicted three out of four cases after applying the correction factor. Additionally, all 11 NY-ESO test cases from (24) were correctly predicted. In total, the implementation of the correction factor achieved a diagnostic accuracy of 93%. Tables 1 and 2 display a summary of each case's corrections.

Incorporation of kinetic correction term leads to improved model predictions for meaningful TCR-pMHC interactions

The kinetic correction factor can improve the predictions on NY-ESO-specific TCRs. In the next step, we implement

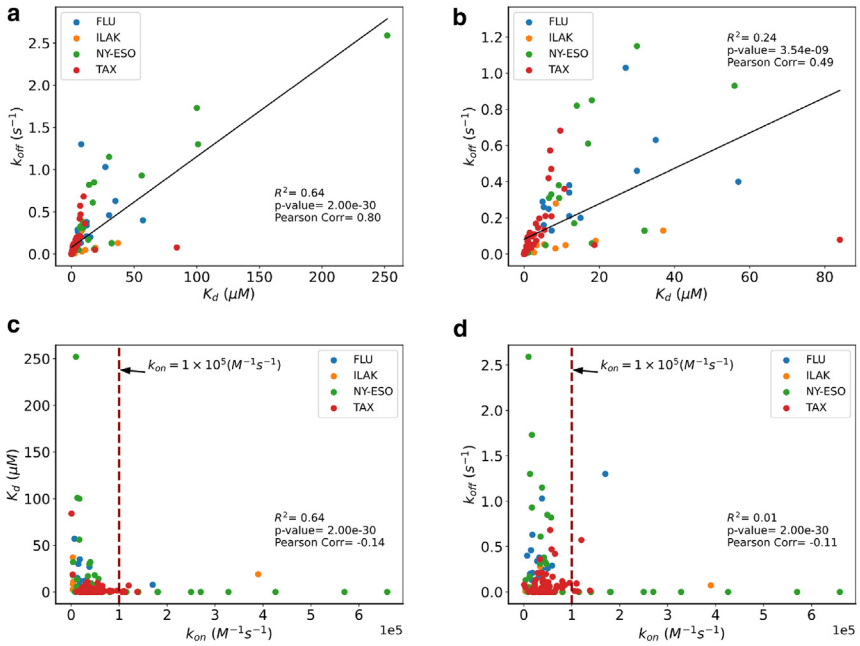

FIGURE 3 Analysis of model fitting to data subsets with larger variability in k_{on}. (*a*) displays the correlation between K_D and k_{off} for the entire ATLAS dataset. (*b*) focuses on a subset of the ATLAS dataset, excluding cases with high variability in k_{on}. These images collectively highlight how different rates of k_{on} influence the model's ability to fit the data, emphasizing the impact of k_{on} on binding parameter correlations. (*c*) illustrates the distribution of k_{on} versus K_D, and (*d*) shows k_{on} versus k_{off} for all cases in the ATLAS dataset.

this correction into the existing RACER approach applied to distinct TCR-pMHC pairs with experimentally confirmed thermodynamic parameters (25,27–30). To assess this, we applied the optimized RACER model to predict TCR-pMHC interactions, restricting our analysis to TCR-pMHC systems in ATLAS with available kinetic data. Our results are reported in Fig. 5 and compare RACER predictions prior to and following kinetic correction. RACER contained 163 unique TCR-peptide pairs, 66 of which had crystal structures and were identified for training, while the remaining 97 pairs were utilized for testing. For the training set with kinetic correction, RACER-m utilized 40 experimentally determined TCR-pMHC complex structures with available kinetic data in training. These structures included both NY-ESO and a total of 15 other functional TCR-peptide systems. Additionally, 40,000 decoy binders were generated by randomizing the peptide sequences, with 1000 decoys created for each strong binder, producing a comprehensive negative dataset that balanced the strong binders as done previously (17). The model's predictive accuracy and generalizability were subsequently evaluated using both the original training set and test TCR-peptide cases, including those from all 16 functional groups in the ATLAS dataset. Cases for which RACER had previously predicted strong binding did not observe significant changes in the predicted binding energy value. This was most evident in the GILGFVFTL (FLU) antigen-specific TCR system. Such cases also tended to correspond to TCR-pMHC pairs where the correction criterion was not satisfied (Fig. 5, *gray values*). This contrasted starkly with the remaining cases, most of which RACER either predicted as a borderline favorable interaction or failed outright. Our kinetic correction successfully improved the original affinity-based predictions in these cases, representing experimentally confirmed favorable TCR-pMHC systems. Lastly, to observe the effects of available kinetic corrections as a function of K_D, we plotted prenormalization and postnormalization binding energies for all cases with kinetic data rank ordered by the original K_D values (Fig. S1). As expected, we find smaller

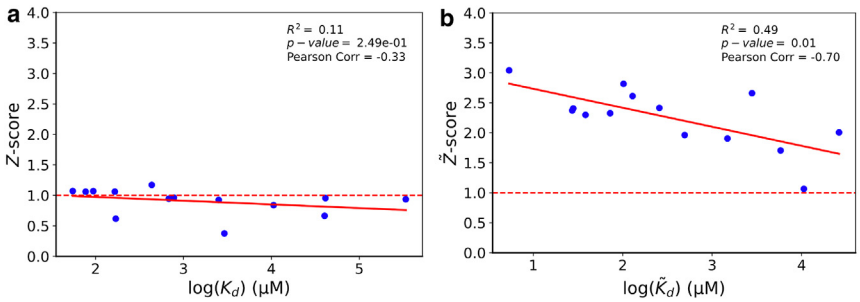

FIGURE 4 Detailed analysis of TCR-pMHC binding parameters for select NY-ESO cases. Comparison of RACER *Z* scores and experimental K_D values for NY-ESO cases with high K_D values, providing insight into the model's effectiveness in predicting TCR-pMHC interactions. (*a*) *Z* scores before applying the correction factor. (*b*) *Z* scores after applying the correction factor, illustrating the improved predictions. This comparison highlights the precision of the RACER model in handling cases with significant variability, thereby illustrating its utility in refining predictions for complex immune responses.

TABLE 1 ***Z*** **scores and** ΔG **values for 1G4 TCR with NY-ESO peptide and its mutations before and after applying the correction factor**

Peptide name	$k_{on} \times 10^3$ (M^{-1} s^{-1})	η_i	Z score	$\tilde{Z}$ score	ΔG (kcal mol^{-1})	$\widetilde{\Delta G}$ (kcal mol^{-1})
ESO-9C	57	0.63	1.1706	2.8172	−2.6642	−4.2259
ESO-9L	17	0.58	0.8390	2.6618	−2.5551	−3.9928
ESO-9V	45	0.39	1.0693	2.2976	−2.4801	−3.4468
ESO-9A	47	0.44	1.0598	2.4064	−2.5189	−3.6096
ESO-3I	35	0.14	0.9459	1.9619	−2.5145	−2.9429
ESO-3M	42	0.36	1.061	2.3252	−2.5959	−3.4878
ESO-3Y	38	0.23	0.9250	1.9042	−2.2861	−2.8562
ESO-4D	10	1.11	0.9358	2.0054	−2.4172	−5.1687
ESO-6V	49	0.48	0.9615	2.4148	−2.4324	−3.6222
ESO-6T	13	0.85	0.9548	1.7045	−2.2861	−4.3931
ESO-7H	17	0.58	0.5530	1.0649	−1.3070	−2.7447

The crystal structures used for RACER analysis are based on PDB: 2BNQ, 2BNR, and 2P5E.

prenormalization versus postnormalization case changes as K_D values decrease.

DISCUSSION

Reliable predictions of TCR specificity are still quite challenging to estimate, partly because a complete description of TCR-pMHC interactions must account for complex interaction dynamics, downstream signaling, and orchestration of cytokine release. Binding affinity, the most widely available thermodynamic characterization of TCR-pMHC interactions, has been successfully applied to train structural models to achieve TCR-pMHC specificity predictions with reasonable accuracy. Despite this, including additional physical parameters, such as binding kinetics, p-MHC abundance on the cell surface, and T cell costimulatory signals, seems necessary to characterize T cell behavior fully. Our goal here was to investigate with more detail the TCR-pMHC situations for which predictive models trained strictly on binding affinity data performed poorly. Given their inability to be accurately accounted for using available affinity data, these cases, including NY-ESO and several other TCR-pMHC functional systems, provided an opportunity to incorporate and test meaningful kinetic data into model predictions. Across all publicly available TCR-pMHC systems, a dramatic increase in the variability of k_{off} emerged for sufficiently low k_{on} values. This observation motivated us to introduce kinetic threshold criteria for refining our earlier modeling approach. The notable increase observed in RACER's predictive accuracy likely result from the fact that the original K_D omitted important kinetic information relevant to specificity, which was then recovered with an appropriate correction. While successful on the limited set of currently available systems, subsequent thermodynamic characterization of additional TCR-pMHC cases will be needed to evaluate whether this empirical cutoff scheme can be applied to immunologically distinct systems. One advantage of our threshold-based approach is that it offers a simple criterion for determining when kinetic corrections are needed. For instance, in several TAX (LLFGYPVYV)- and FLU (GILGFVFTL)-specific cases in which our threshold criterion was not met, utilizing the kinetic corrections yielded minimal deviations from the original Z score predictions. This observation was held for cases with predicted scores both close to and far from the critical cutoff for activation. We assessed the binding affinity of both wild-type and variant 1G4 TCRs to NY-ESO and NY-ESO-variant peptides. Our dataset reveals that pMHC recognition does not always align well with either K_D or k_{off} for cases having lower k_{on}. We augmented our analysis to include additional TCR-pMHC systems with considerable variation in k_{on} to more comprehensively explore k_{on}'s impact on pMHC recognition. Our results align with prior studies that found that consistent k_{on} values were most important in successfully correlating TCR-pMHC recognition with either K_D or k_{off} (31). In many cases, however, TCR-pMHC recognition has been found to successfully correlate with either K_D (1,32,33) or k_{off}(34–38). The corrections in our model are based on SPR-derived kinetic data, as SPR is highly sensitive and precise, especially for low-affinity interactions. Other empirical techniques like bio-layer interferometry and microscale thermophoresis may require

TABLE 2 ***Z*** **scores and** ΔG **values for 1G4 TCR and its mutations before and after applying the correction factor**

TCR name	$k_{on} \times 10^3$ (M^{-1} s^{-1})	η_i	Z score	$\tilde{Z}$ score	ΔG (kcal mol^{-1})	$\widetilde{\Delta G}$ (kcal mol^{-1})
1G4-mu1	17.8	0.16	0.3758	2.3714	−0.9691	−6.6050
1G4-mu2	34	0.12	0.6167	2.6135	−0.9633	−5.2277
1G4-mu3	4	2.03	0.5893	3.9464	−2.0482	−7.8929
1G4-mu4	11	1.01	1.0706	3.0421	−2.9691	−6.0842

The crystal structures used for RACER analysis are based on PDB: 2BNQ, 2BNR, and 2P5E.

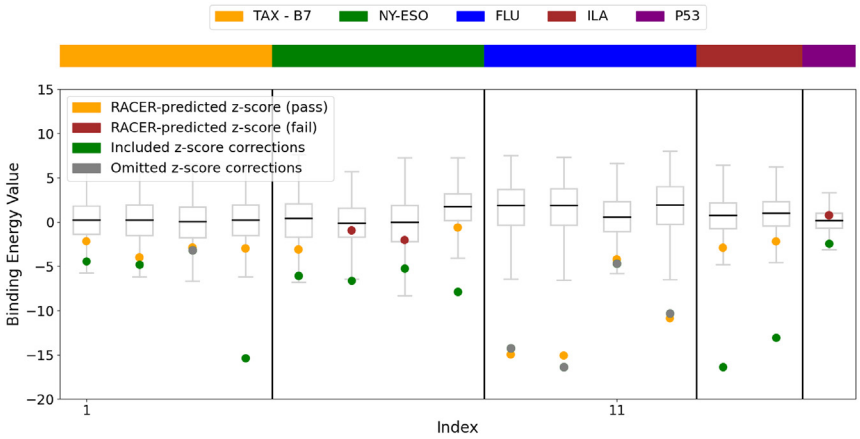

FIGURE 5 Performance on ATLAS dataset before and after applying the correction factor. The boxplots illustrate the distribution of binding energy values, highlighting the variability. The correction factor improves prediction accuracy, particularly in the higher-affinity interactions (represented by the *orange dots*), which is consistent with the lower K_D values demonstrating less improvement.

specific adjustments due to their lower sensitivity for weak interactions (bio-layer interferometry) or the potential variability introduced by fluorescence labeling (microscale thermophoresis). Since these methods focus more on binding affinities than precise kinetics, additional calibration between measurements should be verified for cases that contain affinity and kinetic data from separate experimental setups. Empirical measurements originating from diverse methodologies, where identified, will require calibration and validation. Our results demonstrate that one can address the practical question of when and how to effectively incorporate variable thermodynamic features into inferential models of TCR-pMHC specificity by identifying and incorporating suitable correction factors. Moreover, such adjustments, when properly introduced, can successfully encode independent information (off rate) that is otherwise distinct from the original model biophysical feature (affinity). We remark that while our correction factor is incorporated using thermodynamic theory, the precise criteria for inclusion were made empirically based on all available observations. As a result, our prediction refinements are limited by the availability and accuracy of kinetic data. Our theoretical method works reasonably well for low values of the on rates, and one can provide a more microscopic explanation of this observation. The equilibrium constant K_D provides a measure of binding affinity if these events can be measured for a long time. For large on rates, after some time, the K_D can be well evaluated. However, when the on rates are small due to limited time for immune response processes, one cannot rely on K_D, and the kinetic information provides a better description of the effective binding affinity. One intriguing future direction concerns extending this analysis so that future on rates would no longer be required for applying the kinetic correction to new systems. Such an approach could be implemented by utilizing the mean correction factors over a collection of empirical k_{on} values within a particular functional group, provided the within-cluster variance is adequately controlled. Such an approach requires additional kinetic data for particular systems of interest for the necessary evaluation but could replace altogether the need for additional kinetic data for a new TCR-peptide pair of interest. While successfully applied to multiple TCR-pMHC systems, this approach will require further validation on distinct functional systems. These follow-up analyses will benefit from additional TCR-pMHC kinetic information, particularly in cases having low on rates. The above approach demonstrates promise for broadening the class of encodable information useful to an existing, affinity-based model. In this case, kinetic data were included in a straightforward way using foundational thermodynamic principles. Future research should focus on expanding this approach to include additional features that are commonly measured, such as pMHC abundance and cytokine release, in a biophysically meaningful manner. Such model improvements will broaden the utility of predictive models and, in doing so, will enhance therapeutic strategies to predict TCR-pMHC interactions with applications to cancer immunology, infectious diseases, and autoimmunity.

ACKNOWLEDGMENTS

J.T.G. was supported by the Cancer Prevention Research Institute of Texas (RR210080) and the National Institute of General Medical Sciences of the NIH (R35GM155458). J.T.G. is a CPRIT Scholar in Cancer Research. A.B.K. was supported by the Welch Foundation (C-1559) and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745).

AUTHOR CONTRIBUTIONS

A.B.K. and J.T.G. designed and supervised the research. Z.S.G. and H.T. performed the research. Z.S.G., H.T., A.B.K., and J.T.G. analyzed the data and wrote the paper. All authors approve of the final manuscript.

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2024.11.002.

REFERENCES

1. Boulter, J. M., N. Schmitz, …, A. M. Gallimore. 2007. Potent T cell agonism mediated by a very rapid TCR/pMHC interaction. *Eur. J. Immunol.* 37:798–806.
2. Sun, X., T. Nguyen, …, N. P. Weng. 2022. Longitudinal analysis reveals age-related changes in the T cell receptor repertoire of human T cell subsets. *J. Clin. Invest.* 132
3. Klein, L., B. Kyewski, …, K. A. Hogquist. 2014. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* 14:377–391.
4. Davis, M. M. 2015. Not-so-negative selection. *Immunity.* 43:833–835.
5. Kwee, B. P., M. Messemaker, …, T. N. Schumacher. 2023. STAPLER: efficient learning of TCR-peptide specificity prediction from full-length TCR-peptide data. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.25.538237.
6. Meynard-Piganeau, B., C. Feinauer, …, T. Mora. 2024. TULIP: A transformer-based unsupervised language model for interacting peptides and T cell receptors that generalizes to unseen epitopes. *Proc. Natl. Acad. Sci. USA.* 121:e2316401121.
7. Jurtz, V. I., L. E. Jessen, …, M. Nielsen. 2018. NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks. Preprint at bioRxiv. https://doi.org/10.1101/433706.
8. Glanville, J., H. Huang, …, M. M. Davis. 2017. Identifying specificity groups in the T cell receptor repertoire. *Nature.* 547:94–98.
9. Dash, P., A. J. Fiore-Gartland, …, P. G. Thomas. 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* 547:89–93.
10. Weber, A., J. Born, and M. Rodriguez Martínez. 2021. TITAN: T-cell receptor specificity prediction with bimodal attention networks. *Bioinformatics.* 37:i237–i244.
11. Asgari, E., A. C. McHardy, and M. R. K. Mofrad. 2019. Probabilistic variable-length segmentation of protein sequences for discriminative motif discovery (DiMotif) and sequence embedding (ProtVecX). *Sci. Rep.* 9:3577.
12. George, J. T., D. A. Kessler, and H. Levine. 2017. Effects of thymic selection on T cell recognition of foreign and tumor antigenic peptides. *Proc. Natl. Acad. Sci. USA.* 114:E7875–E7881.
13. Ng Chau, K., J. T. George, …, H. Levine. 2022. Contact map dependence of a T-cell receptor binding repertoire. *Phys. Rev. E.* 106:014406.
14. Košmrlj, A., A. K. Jha, …, A. K. Chakraborty. 2008. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. USA.* 105:16671–16676.
15. Detours, V., and A. S. Perelson. 1999. Explaining high alloreactivity as a quantitative consequence of affinity-driven thymocyte selection. *Proc. Natl. Acad. Sci. USA.* 96:5153–5158.
16. Lin, X., J. T. George, …, H. Levine. 2021. Rapid assessment of T-cell receptor specificity of the immune repertoire. *Nat. Comput. Sci.* 1:362–373.
17. Wang, A., X. Lin, …, J. T. George. 2024. RACER-m leverages structural features for sparse T cell specificity prediction. *Sci. Adv.* 10:eadl0161.
18. Ghoreyshi, Z. S., and J. T. George. 2023. Quantitative approaches for decoding the specificity of the human T cell repertoire. *Front. Immunol.* 14:1228873.
19. François, P., and G. Altan-Bonnet. 2016. The case for absolute ligand discrimination: Modeling information processing and decision by immune T cells. *J. Stat. Phys.* 162:1130–1152.
20. Teimouri, H., and A. B. Kolomeisky. 2020. Relaxation times of ligand-receptor complex formation control T cell activation. *Biophys. J.* 119:182–189.
21. Gálvez, J., J. J. Gálvez, and P. García-Peñarrubia. 2019. Is TCR/pMHC affinity a good estimate of the T-cell response? An answer based on predictions from 12 phenotypic models. *Front. Immunol.* 10:421889.
22. Chen, Y.-T., M. J. Scanlan, …, L. J. Old. 1997. A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc. Natl. Acad. Sci. USA.* 94:1914–1918.
23. Davtyan, A., N. P. Schafer, …, G. A. Papoian. 2012. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B.* 116:8494–8503.
24. Aleksic, M., O. Dushek, …, P. A. van der Merwe. 2010. Dependence of T cell antigen recognition on T cell receptor-peptide MHC confinement time. *Immunity.* 32:163–174.
25. Borrman, T., J. Cimons, …, Z. Weng. 2017. ATLAS: a database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes. *Proteins.* 85:908–916.
26. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.
27. Davis-Harrison, R. L., K. M. Armstrong, and B. M. Baker. 2005. Two different T cell receptors use different thermodynamic strategies to recognize the same peptide/MHC ligand. *J. Mol. Biol.* 346:533–550.
28. Davis-Harrison, R. L., F. K. Insaidoo, and B. M. Baker. 2007. T cell receptor binding transition states and recognition of peptide/MHC. *Biochemistry.* 46:1840–1850.
29. Ishizuka, J., G. B. E. Stewart-Jones, …, E. Y. Jones. 2008. The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vβ domain. *Immunity.* 28:171–182.
30. Wu, D., D. T. Gallagher, …, R. A. Mariuzza. 2020. Structural basis for oligoclonal T cell recognition of a shared p53 cancer neoantigen. *Nat. Commun.* 11:2908.
31. Chervin, A. S., J. D. Stone, …, D. M. Kranz. 2009. The impact of TCR-binding properties and antigen presentation format on T cell responsiveness. *J. Immunol.* 183:1166–1178.
32. Andersen, P. S., C. Geisler, …, K. Karjalainen. 2001. Role of the T cell receptor ligand affinity in T cell activation by bacterial superantigens. *J. Biol. Chem.* 276:33452–33457.
33. Holler, P. D., and D. M. Kranz. 2003. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. *Immunity.* 18:255–264.
34. Carreño, L. J., S. M. Bueno, …, A. M. Kalergis. 2007. The half-life of the T-cell receptor/peptide–major histocompatibility complex interaction can modulate T-cell activation in response to bacterial challenge. *Immunology.* 121:227–237.
35. Kalergis, A. M., N. Boucheron, …, S. G. Nathenson. 2001. Efficient T cell activation requires an optimal dwell-time of interaction between the TCR and the pMHC complex. *Nat. Immunol.* 2:229–234.
36. Kersh, G. J., E. N. Kersh, …, P. M. Allen. 1998. High-and low-potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. *Immunity.* 9:817–826.
37. Krogsgaard, M., N. Prado, …, M. M. Davis. 2003. Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation. *Mol. Cell.* 12:1367–1378.
38. Qi, S., M. Krogsgaard, …, A. K. Chakraborty. 2006. Molecular flexibility can influence the stimulatory ability of receptor–ligand interactions at cell–cell junctions. *Proc. Natl. Acad. Sci. USA.* 103:4416–4421.