

Exhaustive Characterization of TCR – pMHC Binding Energy Estimated by the String Model and Miyazawa-Jernigan Matrix

Hiromichi Tsurui^{1*}, Takuya Takahashi², Yuko Matsuda³, Qingshun Lin¹, Aya Sato-Hayashizaki^{1,4} and Sachiko Hirose¹

¹Department of Pathology, School of Medicine, Juntendo University, 2-1-1 Bunkyo-ku, Hongo, Tokyo 113-8421, Japan

²Department of Bioscience and Bioinformatics, College of life science, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga-ken 525-8577, Japan

³Symbolic Systems, Inc., 1-33-6 Sinyosida-Higashi, Kohoku, YOKOHAMA, 223-0058, Japan

⁴F.G.J Co. Ltd., 3-25-18 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan

Abstract

Accurate calculations of protein–protein binding free energies based on rigorous models, which consider the binding complex structure in atomic detail, are computationally expensive and impracticable to apply to T cell repertoire formation that occurs in the thymus because this process involves the interactions among numerous combinations of T cell receptors (TCRs) and presented peptides. By comparison, an evaluation of binding free-energy using a combination of the string model and Miyazawa-Jernigan matrix is very efficient and was therefore applied to estimate interaction energies between T cell receptor–peptide–MHC (TCR–pMHC) complexes, which appeared to successfully explain the effects of binding capacity of MHC on repertoire–formation and the reason for the presence of elite-controllers of some viral infections. However, this evaluation method is overly simplified and requires more detailed considerations when applied to evaluating TCR–pMHC interactions. In this study, we examined this method exhaustively and revealed the limitations of the method. Following features necessitate cautious attitude when interpreting the calculation results: first, the apparent increase in the number of hot spots in accordance with an increase of educational epitope pool size does not mean an increased TCR specificity of surviving clones; second, strong binders to any TCR converge to some limited sequences that are determined by the physical nature of the Miyazawa-Jernigan matrix.

Keywords: T cell receptors; Miyazawa-Jernigan matrix; Repertoire–formation; Thermodynamic integration; Amino acids; Epitopes

Introduction

The education of T cells in the thymus, primarily involving positive and negative selection, is an essential process for establishing immunological recognition of self and non-self. Among numerous proteins involved in the interactions between thymocytes and epithelial cells, the T cell receptor (TCR) and peptide-MHC (pMHC) play the most important roles. The interactions between them, binding free energies, and to some extent the kinetic features, determine the fate of these cells [1-5]. Several attempts have been made to calculate the binding free energy between TCR–pMHC as precisely as possible despite the extremely high calculation costs [6,7]. A continuum solvent dielectric model was developed, which was far more efficient for calculations than a rigorous method, such as thermodynamic integration, and had substantial accuracy [8,9]. However, this method is not feasible for calculating the interactions among numerous combinations of TCRs and the epitopes presented on thymic epithelia. Miyazawa and Jernigan devised the Miyazawa–Jernigan matrix (M–J matrix) in which the inter–residue potentials were extracted from the crystallography results of 1168 proteins [10]. The principle adopted for this contact potential estimation was that the frequencies of residue–residue contacts observed in a large number of protein crystals would represent the actual intrinsic inter–residue interactions. Li et al. defined the designability of a structure on a hydrophobic-polar (HP)–lattice model as the number of sequences that bear this structure as the ground state (most stable structure) [11]. In other studies, instead of an HP lattice, 20 amino acids (AAs) with the M–J matrix was used, and the results confirmed that designability also applied as an indicator of the folded state in these cases [12-14]. Wang and Wang determined the minimal protein folding alphabets required to form a structured protein using the M–J matrix [15]. The M–J matrix was also used to categorize 20 AAs into several groups [16,17].

String model, which approximates protein–protein interactions as the sum of relevant, ladder-forming AA pair interactions, has been

used to study TCR–pMHC interactions [18-21]. The calculation cost of using a combination of the string model and M–J matrix to estimate TCR–pMHC interactions is quite low, and several attempts have been made to evaluate TCR–pMHC interactions based on this model [22-25]. Kosmrly et al. applied this method to explain the effect of an MHC molecule on T cell repertoire formation, particularly its capacity to bind diversified peptides, and seem to have succeeded to some extent [26,27]. Despite these attempts, the actual configuration of a TCR–pMHC complex may not be so simple as to be simulated by a string model. The M–J matrix has been used primarily to determine the native folded state. However, the density of the interface between TCR–pMHC is not so high as observed with a folded protein [28]. Thus, it is necessary to consider the validity of applying this model to calculations of TCR–pMHC binding energies. Here, we exhaustively characterize this estimation method.

Materials and Methods

TCRs and 9-mer epitopes were generated randomly based on the frequency distribution for mouse endoplasmic reticulum [29]. Epitopes restricted to MHC molecules were generated using Gibbs sampling, taking into consideration both AA frequency and binding scores estimated by position-specific scoring matrix (PSSMs) [30]. PSSMs for Class I were from the Immune Epitope Data Base (IEDB, <http://www.immuneepitope.org/>), and those for Class II were

*Corresponding author: Hiromichi Tsurui, Department of Pathology, School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan, Tel: 81-3-5802-1671; E-mail: tsurui@juntendo.ac.jp

Received October 01, 2013; Accepted December 19, 2013; Published December 27, 2013

Citation: Tsurui H, Takahashi T, Matsuda Y, Lin Q, Sato-Hayashizaki A (2013) Exhaustive Characterization of TCR–pMHC Binding Energy Estimated by the String Model and Miyazawa-Jernigan Matrix. General Med 1: 126. doi: 10.4172/2327-5146.1000126

Copyright: © 2013 Tsurui H, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

calculated using the Multiple Em for Motif Elicitation (MEME) suite (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) based on epitope data registered in IEDB. Mouse cDNA sequences to extract self-epitopes were from the Riken Expression Array Database (<http://read.gsc.riken.jp/fantom2/>) and FANTOM (<http://www.osc.riken.jp/contents/fantom/>). Computers used for calculations included a X8DA3 mother board (SuperMicro, California, USA) and dual Xeon-E5620s (Intel Corp., California, USA). The calculation algorithm used for determining the binding free energy was same as that in a previous study [26]. Programs used for the calculations were coded using Mathematica v.8 (Wolfram Research, Illinois, USA). Molecular images were drawn using Chimera 1.5.3 [31].

Results

String model and M–J matrix

Several studies have applied the string model to approximate protein–protein interactions in immunology. Kosmrlj et al. used the string model and M–J matrix to calculate TCR–pMHC binding energies and explained the effects of the binding capacity of MHC molecule on repertoire formation [26,27]. As shown in Figure 1A, this model assumes that the CDR1 and CDR2 of a TCR interact with MHC helices, and only CDR3 interacts exclusively with a peptide presented on the MHC. For simplicity, Kosmrlj et al. took into account the interaction only between CDR3 and the presented peptide, and we also did so. If the

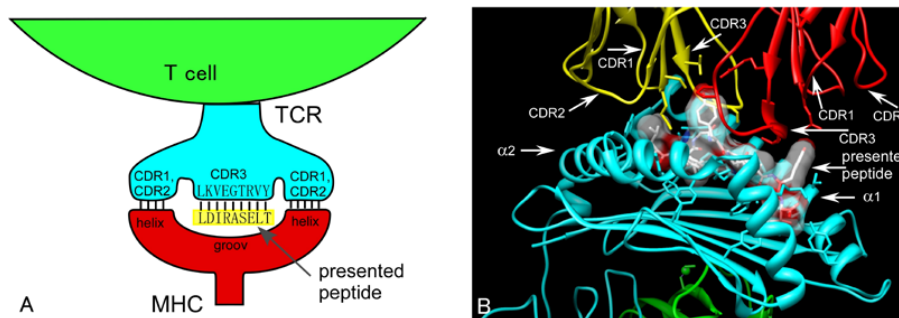


Figure 1: String model for TCR–pMHC complex

In this simplified model, CDR1 and CDR2 of TCRs mainly interact with helices running along both sides of MHC molecules, between which the presented peptide is located. In this case, CDR3 (LKVEGTRVY) interacts with a presented peptide (LDIRASELT) so as to form a ladder with 9 rungs (L–L, K–D, V–I, E–R, G–A, T–S, R–E, V–L, Y–T). For simplicity, only the interaction between CDR3 and the presented peptide is considered (SI 1). (A) AA contact energies were assigned respectively to these rungs (AA pairs) by the M–J matrix, and the binding energy between CDR3 and the presented peptide could be calculated by summing these values. (B) Actual 3-dimensional structure of a TCR–pMHC complex (PDBID: 1A07) is shown. Both CDR3 loops comprise α and β chains running across rather than parallel to the presented peptide. CDR2 mainly interacts with helices, whereas CDR1 interacts with both MHC helix and presented peptide. The TCR–pMHC binding mode also varies widely. This image was generated using Chimera 1.5.3.

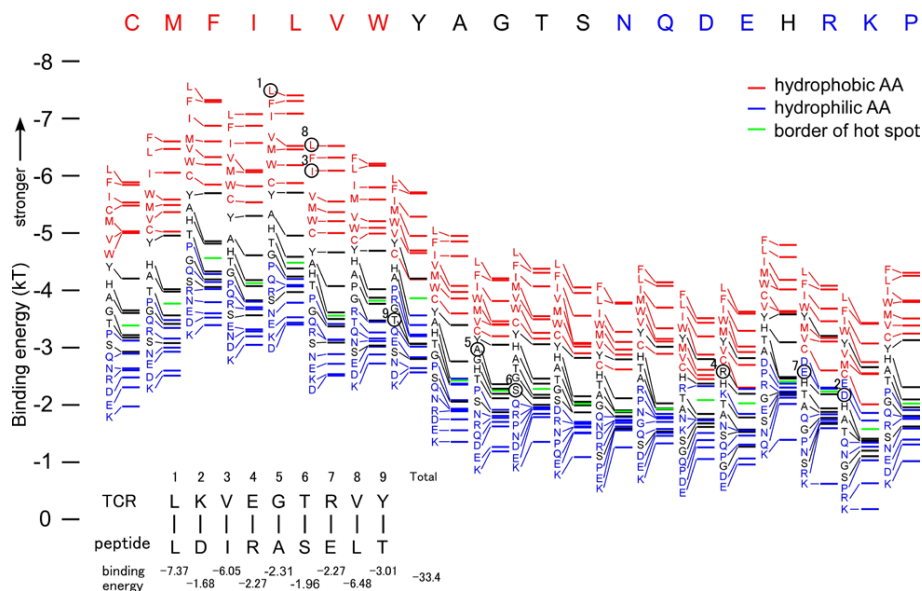


Figure 2: Altered expression of M–J matrix according to energy level

The M–J matrix assigns interaction energies to AA pairs. Columns of an original matrix were transformed so that AAs were arranged according to their binding energy values expressed in units of kT, where k is Boltzmann's constant and T is absolute temperature. Red bar indicates that the AA pair is hydrophobic and blue indicates it is hydrophilic. Green bars were placed between the 10th and 11th AA in the order of energy levels, above which a selected AA behaves as a hot spot. As an example, the binding energy between TCR (LKVEGTRVY) and presented peptide (LDIRASELT) was calculated as follows: interaction between L and L (black-circle) was numbered as 1, –7.37 kT; between K and D was numbered as 2, –1.68 kT; V and I was numbered as 3, –6.05 kT, and so on were read from the M–J matrix and summed to a total of –33.4 kT. Note that the pattern of each column is quite similar; within each column hydrophobic residues (red) are located on the upper side (stronger) and hydrophilic residue (blue) are located on the lower side (weaker). Columns for hydrophobic residue (C, M, F, I, L, V, and W) are also located entirely above (stronger) and those for hydrophilic residue (N, Q, D, E, R, K, P) are located below (weaker).

interaction of CDR1 and CDR2 of any TCR with an MHC molecule can be assumed to be constant and the thresholds for positive and negative selection (E_p and E_N , respectively) are set appropriately, then this approximation seems applicable for considering the principle binding characteristics based on the M-J matrix. However, this model may still seem too simplified. An actual TCR-pMHC complex, as shown in Figure 1B, does not assume such a simple configuration. For example, CDR1 interacts not only with MHC but also with the presented peptide and CDR3 is not configured in parallel with the presented peptide. Moreover, several aberrant binding modes for TCR-pMHC complexes have been reported [32-35]. These results strongly suggest that one should be cautious when applying this method to evaluate TCR-pMHC binding free-energy.

M-J matrix rearrangement according to energy-levels

Figure 2 shows an M-J matrix, in which each column is rearranged such that bars indicating AAs are placed in accordance with the binding free energy, from stronger (upper) to weaker (lower). Red bars indicate hydrophobic AAs and blue bars indicate hydrophilic AAs, as plotted by Kyte and Doolittle [36]. The interaction energy between TCR and the presented peptide is calculated based on the string model as the sum of interacting AA energies. As an example, the interaction energy between

the TCR CDR3 (LKVEGTRV) and a peptide LDKIRASELT, as shown in Figure 2 (lower left), was calculated as $-(7.37 + 1.68 + 6.05 + 2.27 + 2.31 + 1.96 + 2.27 + 6.48 + 3.01)$ kT (k is Boltzmann's constant and T is absolute temperature), where each of these energies corresponds to the binding energy between L-L, K-D, V-I, E-R, G-A, T-S, R-E, V-L, and Y-T, respectively.

Dependence of the strongest binder on the educational epitope pool size

Thymocytes are educated in the thymus through the interactions between TCRs and self-peptides presented on MHC molecules of thymic epithelia. TCRs interact with self-peptides with varying strengths; those that interact with interaction energy weaker than E_p are subject to apoptosis, and those that interact with interaction energy stronger than E_N are deleted (negative selection) [1-5]. Therefore, epitopes that should be recognized as non-self must interact with TCRs with energies stronger than E_N when encountered after thymic education (SI 1). As the size of the pool of educational epitopes increases, the range of interaction energies between individual epitopes and TCRs becomes wider. Figure 3A shows an example of calculated binding profiles, namely histograms of bound peptides, of a TCR with sequence AMLSYCIEK against epitope-pools of 32, 100, and 316 peptides. The

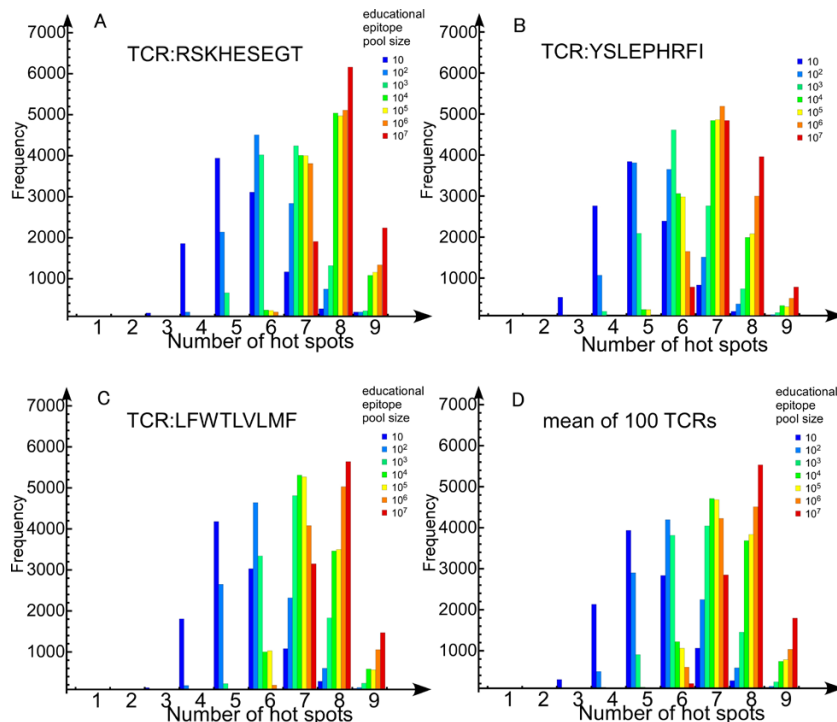


Figure 3: Binding profile of a TCR and the effect of educational epitope pool size

In this model, a TCR binds to any peptide with an energy calculated based on the M-J matrix. A histogram of binding energies (binding profile) against a particular epitope pool can be generated for an individual TCR. One hundred of TCRs (9-mers) and an epitope pool of 10^7 peptides were randomly generated based on the AA frequency of the mouse proteome. (A) For the TCR sequence AMLSYCIEK, which was randomly selected from the 100 TCRs, 32, 100, and 316 peptides were extracted from the epitope pool of 10^7 peptides and the binding energies were calculated. The possible strongest and weakest binders for this TCR, uniquely determined based on the TCR sequence, are LFLFLLLLL and KKKKKKKKEK, respectively. Also shown are the strongest and weakest binders within each pool: EVRCIHVVR and EGREKSPEE, respectively, among 32 peptides; MPKNILLEC and EEEQEEQEQ, respectively, among 100 peptides; and HFRLIVIKR and EEEQEEQEQ, respectively, among 316 peptides. As the pool size increases, the strongest binder becomes stronger and the weakest binder becomes weaker. (B) Dependencies of the strongest and weakest binders based on epitope pool size. A series of epitope pools from 32 to 10^7 peptides were generated. As an example, binding profiles of a TCR (YSLEPHRFI) against this series of epitope pools were calculated and the strongest and weakest binders within each pool were determined. The sequences of bound epitopes are shown. The means and standard deviations of these pools remained nearly constant. (C) Binding profiles are shown for 3 representative TCRs (strong, intermediate, and weak binders selected from the 100 TCRs) against the epitope pool of 10^7 peptides. As the mean of the binding profile becomes stronger, the standard deviation becomes greater. (D) Exemplified relationship among E_p , E_N , and binding profile against educational epitopes (considered as self) and exogenous epitopes for a TCR of a surviving cell is shown. Exogenous epitopes that bind stronger than the discrimination threshold were recognized as non-self. The discrimination threshold should be stronger than E_N . As the size of the pool increases, the strongest binding energy becomes stronger and the weakest becomes weaker. To survive this educational process, the binding energy of the strongest binder among educational epitopes should be weaker than the threshold for negative selection.

profiles appear similar to Gaussian curves with the very near averages and standard deviations. The strongest and weakest binders within the individual pools are EVRCIHVVR (−29.99 kT) and ERGEKSPEE (−18.44 kT), respectively, for a 32 peptide pool; MPKNILLEC (−30.72 kT) and EEEQEEQEQ (−17.6 kT), respectively, for a 100 peptide pool, and HFRLIVIKR (−32.22 kT) and EEEQEEQEQ (−17.6 kT), respectively, for a 312 peptide pool. This TCR binds strongest to LFLFLLLLL (−48.36 kT) and weakest to KKKKKKKKEK (−16.8 kT). As shown in Figure 3B, a wider range of epitope pool size $10^{1.5}$ – 10^7 peptides was investigated. As an example, the TCR with sequence YSLEPHRFI, a medium binder, was adopted. While the size of the epitope pool increased, the mean and standard deviation remained nearly constant (~approximately 29 kT and 3 kT, respectively). In contrast, the strongest binder within the pool became stronger and the weakest binder became weaker. This feature was common to other TCRs. The binding profiles of 10^7 peptides for 3 TCRs, RSKHESEGT (weak binder), YSLEPHRFI (medium binder), and LFWTLVLMF (strong binder), are shown in Figure 3C. Effects of haplotypes on binding profiles for those TCRs as shown in Figure S2, were almost negligible except for on D^d. Scatter graphs of the means and standard deviations against an epitope pool of 10^7 peptides, restricted to several MHC–haplotypes, for 100 TCRs are shown in Figure S3. In this case, epitopes were generated not only randomly, but also with some restriction so as to bind to the indicated MHC molecule more strongly than a given criteria. Binding profiles that had strong mean values had large standard deviations, as indicated by Figure 3C and linear regression analysis shown in Figure S3. These distributions were highly correlated. Exemplified relationships among E_p , E_N , and binding profiles of educational epitopes and epitopes recognized as non-self, are illustrated in Figure 3D. If a clone was to survive after positive and negative selection, the strongest binding energy against an educational

self-epitope should be weaker than E_N , and if an epitope should be recognized as non-self, the binding energy must stronger than E_N . Epitopes that bind more weakly than this threshold will be recognized as self, regardless of their actual origin.

Number of hot spots increases as the educational epitope pool size increases

As shown in Figure 3, an increase in the educational epitope pool size results in a wider range of binding profiles. This means that the discriminating thresholds of the surviving clones become stronger as the size of the educational epitope pool increases. A hot spot is defined as a site where more than 10 (half of the 20 various AAs) mutations weaken the binding energy. This means that an AA in the hot spot site is located above the green bar in Figure 2. As the discrimination threshold becomes stronger, each peptide AA must bind more strongly, on average, causing a decrease in acceptable mutations (which provide stronger binding) and thus increasing the number of hot spots. This can be simulated as shown in Figure 4. In this simulation, 3 representative TCRs were selected (Figure 3). The discrimination threshold was set as the strongest binding energy in the epitope pool; therefore, it behaves as a monotonically increasing function of educational epitope pool size in the range of 10^1 – 10^7 peptides. Gibbs sampling was used to generate 10^4 peptides that bind stronger than the given thresholds, and hot spots for these peptides were calculated. As is clear from all the panels in Figure 4, an increase in the epitope pool size resulted in an increase in hot spots for these 3 TCRs, weak (RSKHESGT, Figure 4A), medium (YSLEPHRFI, Figure 4B), and strong (LFWTLVLMF, Figure 4C) binders. The mean values of 100 TCRs shows similar features (Figure 4D).

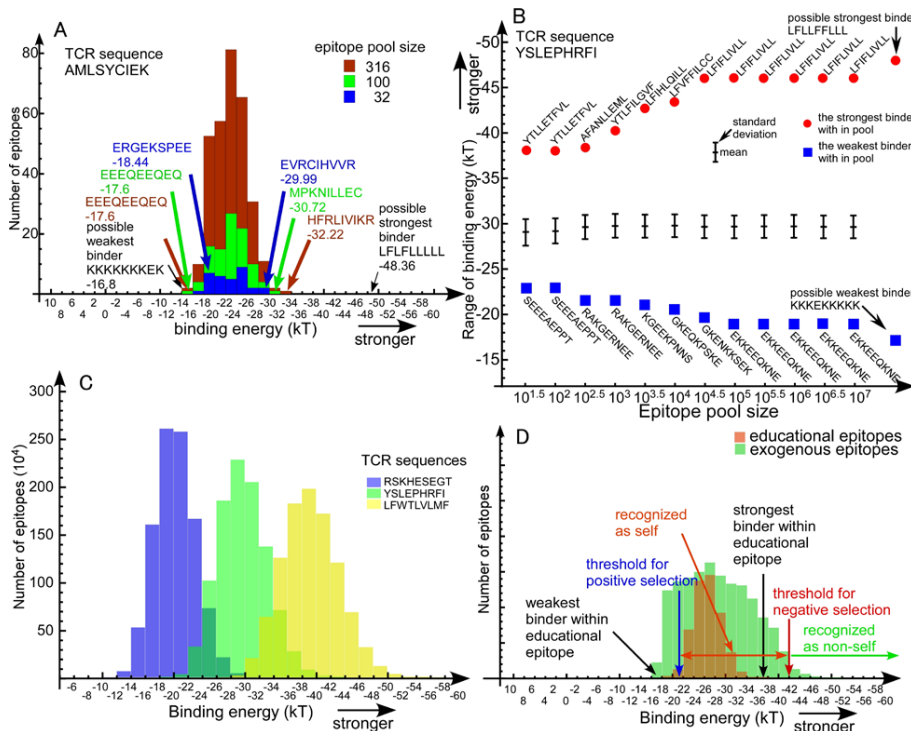


Figure 4: Dependence of the number of hot spots on the size of the educational epitope pool

The strongest binding energies from epitope pools of 10^1 – 10^7 peptides for each TCR obtained in the Figure 3 were adopted as the thresholds. A total of 10^4 peptides with binding energies stronger than these thresholds were generated by Gibbs sampling. For these newly generated epitopes, the numbers of hot spots (as defined in the text) were calculated. Histograms for 3 representative TCRs (A, B, C; same as in Figure 3) and the average of randomly generated 100 TCRs (D) is shown. As the epitope pool size increased, the number of hot spots also increased (i.e., histograms shifted from left to right).

		difference from the mean							
TCR		-5	-10		-15		-20		the strongest	
		the strongest energy in 104			the strongest energy in 107					
RSKHESEGT -18.3	VQYELLSGI	-24.74	IVMLILLGE	-28.46	FLLFYVLL	-34.02			LFLFLLLL	-35.89
	LVMGCYLQS	-23.97	YFLWLLFVH	-31.4	ILFFMLLLV	-33.7				
	KHGIAILLC	-23.88	IFHLLLRK	-28.41	CFYLLLLLL	-33.33				
	VFLRIWPES	-23.32	CLILVLLYI	-31.26	IVLLLLFLL	-34.16				
	ILTLHGLES	-24.14	MFLFWAVF	-30.17	LLFLLSLL	-33.34				
	LPGGLVFGA	-23.66	VIIFVGLVY	-28.84	LFYIWLIFL	-33.44				
	QFCFFGETL	-25.25	RYFWLLFF	-29.86	FVLILLII	-33.45				
	FQLMSHVI	-25.33	PFLFCVFT	-28.97	LLVFLCLFI	-33.41				
	MEFILSFAV	-26.69	LFMLMEIG	-28.57	LWLLLLWFL	-33.9				
RFAVRLSMS	-23.48	IVYLLGLKL	-28.31	LFWFFLWVL	-33.7					
		the strongest energy in 104			the strongest energy in 107					
YSLEPHRFI -26.5	FPPFLVIKK	-32.77	LLLQMGIGF	-38.58	LLLFLPLRF	-41.82	FLLLLWLF	-46.85	LFLFFLLL	-48.02
	FQEFESLVY	-33.05	LLLNL SRMQ	-36.56	VLLMCELF	-42.64	IVLLLFL	-46.58		
	LVISLPVKR	-33.75	LTLYLFSI	-38.69	SFLTILILI	-41.66	LLLLLLLL	-47.64		
	VSSKKFLE	-31.55	VHIYVHWLA	-38.2	IVMFFMMV	-43.3	FLLLLVFL	-46.62		
	GEATYFTML	-34.75	TISSMIYL	-36.93	IILELVML	-42.46	LIIFLIFL	-47.09		
	FRVPIISIR	-35.01	ACLSLENFF	-37.16	CILLTLVLL	-42.47	LLLFLFVL	-46.57		
	YFCRVFTLF	-40.4	MMAKLVVIS	-36.82	VHLMILVL	-41.72	LIIFLLL	-47.04		
	PICYCEGFV	-35.58	QSIISMLLC	-37.31	FSLVFLVLL	-43.77	LLLLLILFL	-47.22		
	VKKIFEPPA	-31.69	AILKFELMI	-39.58	IMFLPLVIL	-42.39	FLILVFL	-46.65		
QPLVELVMA	-34.59	LITLICCVR	-36.97	FMFMLFSL	-42.92	LFLIFLILF	-46.87			
		the strongest energy in 104			the strongest energy in 107					
LFWTLVLMF -35.8	SAYVMLRIE	-43.27	ALLEIFCEA	-46.93	ILLCLNVFY	-52.47	LLCHLLLL	-56.93	LFLLLLFL	-60.21
	GAAAHYPLI	-41.72	FLPMSAVFI	-49.64	FYMILFELH	-50.95	LLFAMVLF	-56.25		
	RGEKIGLW	-41.71	RPRAYLLL	-47.22	FLAMFLVTF	-52.9	LFIIVLV	-56.58		
	NLCFFIEQ	-47.85	QTLTLLTMS	-46.47	FVLSLWYHL	-51.15	VLILMLFLV	-56.75		
	ELAIFFGGG	-43.54	AVLLMYVT	-49.64	LLCFHPVLL	-51.91	VFLLVWMFL	-56.13		
	VTLSVTKGI	-42.4	RWIFALAAM	-47.05	IYVHLLQL	-50.91	LMALLIVFF	-55.81		
	NREYLVNLV	-42.9	VFCFSLSVI	-49.45	LLEAVLFTF	-50.97	FLPVILFLL	-56.24		
	KLPVREMYL	-43.14	LLAMQLYTA	-46.49	ILELLFLT	-53.37	LIFYLLLYV	-55.8		
	GELSFMLM	-45.55	LVLISLWTT	-48.16	LLLKLEFLF	-53.09	FLLHVLILL	-56.81		
	VPEGCRYL	-43.78	GALMQIFFT	-46.83	LGLNCIIF	-51.84	LLLVIYLL	-56.7		

Table 1:

Strongest binders converge to very limited sequences

An increase in the number of hot spots means an increased rigidity for TCR recognition. In this context, rigidity means any perturbation of the sequence weakens binding of TCR–pMHC. Therefore, Figure 4 apparently shows that the increase in the epitope size brings about an increased specificity of TCR-recognition of surviving T cells. By comparison, the M–J matrix, illustrated in Figure 2 shows that the patterns of energy level–distributions are very similar among almost all the columns. This suggests that stronger binders for any TCR converge to some limited common sequence. As can be observed from this matrix, C, F, I, L, V, Y, A, G, T, S, D, E, R, and K bind strongest to L, and M, W, N, Q, H, and P to F. We generated 200 peptides that bind with energies stronger by 5, 10, 15, and 20 kT than the mean binding energy of the epitope pool size of 10⁷ peptide. Ten of these peptides were randomly selected (Table 1). Peptides with binding energy stronger by 5 kT contained a variety of sequences, and this diversity decreased as the binding energy became stronger. Peptides with binding energy stronger by 15 kT contained very similar sequences that were abundant in L and F. This feature was obvious in both YSLEPHRFI and LFWTLVLMF

cases. Especially in the case of YSLEPHRFI with binding energy stronger by 20 kT than the mean, which corresponded to the region that is within 1.5 kT from the strongest binding energy for this TCR, all the sequence consisted of almost L and F. These results derive from the fact that the distributions of energies within the columns of the M–J matrix are quite similar to each other. At the same time, an apparent increase in strictness of recognition (increase in the number of hot spots) involves the convergence of interacting peptides to the common strongest binders irrespective of the TCR sequence.

Discussion

In this study, we considered the interaction between CDR3 and presented peptides only. Kosmrlj et al. also adopted this model [26]. As mentioned in SI 1, this model seems applicable if the interaction of CDR1 and CDR2 with MHC can be presumed as constant and E_p and E_N are set appropriately. The M–J matrix is one of the successful knowledge-based potentials that are widely used as effective free energies to parameterize protein coarse-grained models. This matrix was obtained from databases of known protein structures and has

been used in some of the best known protein structure prediction and folding modeling algorithms [12-14]. In 1985, Miyazawa and Jernigan formalized the theory for contact interactions among 20 natural AAs based on the quasi-chemical approximation, which treats the chain connectivity effects and solvent effects with simple inter-residue parameters [37,38]. The M-J matrix was revised in 1996, in which there were 6 times more residue-residue contact pairs and used 1168 protein structures containing 1661 subunit sequences. The new attractive contact energies were nearly identical to the older ones, except for those with Leu and the rarer AAs Trp and Met, which showed maximum variation, and the estimates of hydrophobicity from the contact energies for non-polar side-chains agreed well with experimental values. In an application of these contact energies, the sequences of 88 structurally distinct proteins in the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) were threaded at all possible positions without gaps into 189 different folds of proteins which differed from each other by at least 35% in their sequence identities. The native structures of 73 of 88 proteins, excluding 15 exceptional proteins such as membrane proteins, were all demonstrated to have the lowest alignment energies [10].

The essence of the M-J matrix is summarized in two simple first degree inequalities:

$$ePP > eHP > eHH$$

$$eHP > (ePP + eHH) / 2,$$

where ePP is the interaction energy between arbitrary polar residues, eHP is the interaction energy between arbitrary polar and hydrophobic residues, and eHH is the interaction energy between arbitrary hydrophobic residues. This study shows that these inequalities nearly explain the statistical results obtained in many previous studies, such as that of Kosmrlj et al. [26].

Moreover, Li et al. analyzed the physical meaning of the M-J matrix, which in principle could have 210 independent elements, by using only 20 parameters q_i , associated with the 20 AAs, and 3 interaction coefficients. Pair-wise inter-residue interactions responsible for folding could be attributed to the hydrophobic force and a force of demixing; the latter obeying Hildebrand's solubility theory (HST). Although HST was derived for simple non-polar molecules, it was previously found that this theory describes the behavior of polymer blends well. The application to proteins is another example of the more general scope of HST [39]. Subsequently, Keskin et al. arrived at the same conclusion and proposed a reduced set of energy parameters comprising 20 one-body and 3 two-body terms (as opposed to the 20×20 tables of inter-residue potentials). This reproduced the conventional 20×20 tables with a correlation coefficient of 0.99 [40].

Because the potential parameters of the M-J matrix are obtained from known protein structures, the parameter values may implicitly contain physico-chemical properties such as: (i) solvation effect (i.e., hydrophobic groups of a soluble protein tend to distribute inner side), (ii) electrostatic stabilizing effect (i.e., formation of hydrogen bonds and salt bridges); and (iii) steric hindrance of side chains (i.e., tendencies for secondary structure formation, such as Ala in an α -helix and Val in a β -sheet). In addition to the physico-chemical factors, biological factors subject to natural selection during evolutionary processes also constrain the amino acid sequences and structures (i.e., a specific sequence is conserved for a specific biological function).

As previously mentioned, the major factor is (i) HST, which is the origin of the 2 inequalities, and the effects described in (ii) and (iii) are weak in the M-J potential. The effect of higher order structure was

independently considered in a subsequent study, where the potential used included a secondary structure potential representing short-range interactions for secondary structures of proteins, and a tertiary structure potential consisting of a long-range, pairwise contact potential and a repulsive packing potential [41]. Because the M-J matrix used by Kosmrlj et al. was the 1996 version, such higher order structure contributions were not considered. Moreover, because the quasi-chemical approximation breaks down at low temperature, we should be careful when applying the M-J matrix and similar models to temperature-dependent phenomena [38].

For all T cells, E_N is set at quite a narrower range compared to E_p , because of the compartmentalization of Ras/MAP signaling [2]. In the calculations for Figure 4, E_N for each cell was set equal to the strongest binding energy within the interacting epitope pool for counting the number of hot spots. This was due to the request that the boundary between self and non-self should be stronger than E_N . To be close to the actual repertoire-formation process, surviving clones should first be selected rigorously based on their binding profiles and a pre-determined E_N , and the number of hot spots should be calculated. Because E_N was stronger than the strongest energy within the pool, the number of hot spots increased in accordance with the epitope pool size, which resulted in a shift in the histogram to the right. Adopting a common E_N or individual mean-based E_N was not essential for these results.

The differences in MHC haplotypes mainly affect two factors: interactions of CDR1 and CDR2 with MHC helices, and the selection of peptides through binding to MHC molecules. The effect of peptide-selection is shown in Figure S3 as scatter graphs of the mean and standard deviations for 100 TCRs for several MHC haplotypes. The effect of peptide-selection is quite small, with the exception that D^b selects weak binders (approximately by 2-3 kT). The reason for this is not certain.

At present, how the epitope pool size affects repertoire-formation, both in terms of positive- and negative-selection, remains controversial [42,43]. Kosmrlj et al. provided a conclusion from the theoretical point of view to some extent. However, as shown in Figure 4, the relationship between the educational epitope pool size and the specificity of surviving clones is not simple.

Although careful consideration is necessary to interpret the calculation results obtained by using the string model and the M-J matrix, the study by Kosmrlj et al. to evaluate the interaction energies between numerous TCRs and pMHCs rather than a single TCR-pMHC pair was a significant achievement in investigating the mechanisms involved in repertoire formation [26]. The methods, in terms of both hardware and software, to calculate protein-protein or protein-peptide interactions, are under rapid development [44]. More precise calculations with practical computational costs will be feasible in the near future.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 22590194 and 25330349 for HT and, partly, Grants-in-Aid for Scientific Research on Innovative Areas (No. 20118003) from the MEXT for TT.

References

1. Starr TK, Jameson SC, Hogquist KA (2003) Positive and negative selection of T cells. *Annu Rev Immunol* 21: 139-176.
2. Daniels MA, Teixeira E, Gill J, Hausmann B, Roubaty D, et al. (2006) Thymic selection threshold defined by compartmentalization of Ras/MAPK signalling. *Nature* 444: 724-729.

3. Klein L, Hinterberger M, Wirnsberger G, Kyewski B (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol* 9: 833-844.
4. Palmer E, Naeher D (2009) Affinity threshold for thymic selection through a T-cell receptor-co-receptor zipper. *Nat Rev Immunol* 9: 207-213.
5. Gascoigne NR, Palmer E (2011) Signaling in thymic selection. *Curr Opin Immunol* 23: 207-212.
6. Michielin O, Karplus M (2002) Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: a simulation analysis. *J Mol Biol* 324: 547-569.
7. Wan S, Coveney PV, Flower DR (2005) Molecular basis of peptide recognition by the TCR: affinity differences calculated using large scale computing. *J Immunol* 175: 1715-1723.
8. Gohlke H, Case DA (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem* 25: 238-250.
9. Zoete V, Michielin O (2007) Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCR-p-MHC complex. *Proteins* 67:1026-1047.
10. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 256: 623-644.
11. Li H, Helling R, Tang C, Wingreen N (1996) Emergence of preferred structures in a simple model of protein folding. *Science* 273: 666-669.
12. Li H, Tang C, Wingreen NS (1998) Are protein folds atypical? *Proc Natl Acad Sci U S A* 95: 4987-4990.
13. Li H, Tang C, Wingreen NS (2002) Designability of protein structures: a lattice-model study using the Miyazawa-Jernigan matrix. *Proteins* 49: 403-412.
14. Wingreen NS, Li H, Tang C (2004) Designability and thermal stability of protein structures. *Polymer* 45: 699-705.
15. Wang J, Wang W (1999) A computational approach to simplifying the protein folding alphabet. *Nat Struct Biol* 6: 1033-1038.
16. Cieplak M, Holter NS, Maritan A, Banavar JR (2001) Amino acid classes and the protein folding problem. *The Journal of Chemical Physics* 114: 1420.
17. Esteve JG, Falceto F (2004) A general clustering approach with application to the Miyazawa-Jernigan potentials for amino acids. *Proteins* 55: 999-1004.
18. Detours V, Mehr R, Perelson AS (1999) A quantitative theory of affinity-driven T cell repertoire selection. *J Theor Biol* 200: 389-403.
19. Detours V, Perelson AS (1999) Explaining high alloreactivity as a quantitative consequence of affinity-driven thymocyte selection. *Proc Natl Acad Sci U S A* 96: 5153-5158.
20. Scherer A, Noest A, de Boer RJ (2004) Activation-threshold tuning in an affinity model for the T-cell repertoire. *Proc Biol Sci* 271: 609-616.
21. Chao DL, Davenport MP, Forrest S, Perelson AS (2005) The effects of thymic selection on the range of T cell cross-reactivity. *Eur J Immunol* 35: 3452-3459.
22. Castiglione F (2011) A Monte Carlo Simulation for the Construction of Cytotoxic T Lymphocytes Repertoire. Application of Monte Carlo Methods in Biology, Medicine and Other Fields of Science, ed Mode C (InTech, Rijeka, Croatia), pp 131-146.
23. Rapin N, Lund O, Bernaschi M, Castiglione F (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One* 5: e9862.
24. Castiglione F, Santoni D, Rapin N (2011) CTLs' repertoire shaping in the thymus: a Monte Carlo simulation. *Autoimmunity* 44: 261-270.
25. Woelke AL, von Eichborn J, Murgueitio MS, Worth CL, Castiglione F, et al. (2011) Development of immune-specific interaction potentials and their application in the multi-agent-system Vacclmm. *PLoS One* 6: e23257.
26. Kosmrlj A, Jha AK, Huseby ES, Kardar M, Chakraborty AK (2008) How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc Natl Acad Sci U S A* 105: 16671-16676.
27. Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, et al. (2010) Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* 465: 350-354.
28. Reiser JB, Darnault C, Grégoire C, Mosser T, Mazza G, et al. (2003) CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol* 4: 241-247.
29. Flícek P, Aken BL, Beal K, Ballester B, Caccamo M, et al. (2008) Ensembl 2008. *Nucleic Acids Res* 36: D707-714.
30. Bishop CM (2006) Pattern recognition and machine learning (springer New York).
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
32. Kato Z, Stern JN, Nakamura HK, Kuwata K, Kondo N, et al. (2008) Positioning of autoimmune TCR-Ob.2F3 and TCR-Ob.3D1 on the MBP85-99/HLA-DR2 complex. *Proc Natl Acad Sci U S A* 105: 15523-15528.
33. Maynard J, Petersson K, Wilson DH, Adams EJ, Blondelle SE, et al. (2005) Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity* 22: 81-92.
34. Wucherpfennig KW, Call MJ, Deng L, Mariuzza R (2009) Structural alterations in peptide-MHC recognition by self-reactive T cell receptors. *Curr Opin Immunol* 21: 590-595.
35. Gras S, Burrows SR, Kjer-Nielsen L, Clements CS, Liu YC, et al. (2009) The shaping of T cell receptor recognition by self-tolerance. *Immunity* 30: 193-203.
36. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105-132.
37. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18: 534-552.
38. Betancourt MR (2009) Another look at the conditions for the extraction of protein knowledge-based potentials. *Proteins* 76: 72-85.
39. Li H, Tang C, Wingreen NS (1997) Nature of driving force for protein folding: a result from analyzing the statistical potential. *Physical review letters* 79: 765-768.
40. Keskin O, Bahar I, Badretdinov AY, Ptitsyn OB, Jernigan RL (1998) Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci* 7: 2578-2586.
41. Miyazawa S, Jernigan RL (1999) An empirical energy potential with a reference state for protein fold and sequence recognition. *Proteins* 36:357-369.
42. Huseby ES, White J, Crawford F, Vass T, Becker D, et al. (2005) How the T cell repertoire becomes peptide and MHC specific. *Cell* 122: 247-260.
43. Wang B, Primeau TM, Myers N, Rohrs HW, Gross ML, et al. (2009) A single peptide-MHC complex positively selects a diverse and specific CD8 T cell repertoire. *Science* 326: 871-874.
44. Tsurui H, Takahashi T (2012) Practical estimation of TCR-pMHC binding free-energy based on dielectric model and coarse-grained model. *Molecular Dynamics - Studies of Synthetic and Biological Macromolecules*, (InTech, Rijeka, Croatia), pp 107-134.

Citation: Tsurui H, Takahashi T, Matsuda Y, Lin Q, Sato-Hayashizaki A (2013) Exhaustive Characterization of TCR–pMHC Binding Energy Estimated by the String Model and Miyazawa-Jernigan Matrix. *General Med* 1: 126. doi: [10.4172/2327-5146.1000126](https://doi.org/10.4172/2327-5146.1000126)