

Supporting information: Influence of correlated antigen presentation on T cell negative selection in the thymus

Soumya Banerjee¹ and S. Jonathan Chapman¹

¹ Mathematical Institute, University of Oxford, Oxford, United Kingdom.

S1 Appendix. Probability of negative selection.

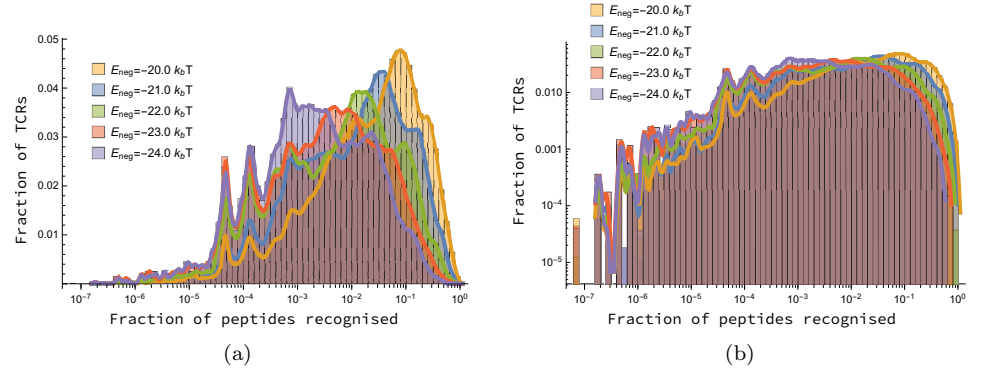
When $p = 2$, the probability of negative selection in a single interaction is

$$\begin{aligned}
 P_1 &\sim \sum_{r=0}^{m/K} \left(1 - \frac{d}{m}\right)^{m/k-r} \left(\frac{d}{m}\right)^r \binom{m/K}{r} \left(\left(1 - \frac{rK}{m}\right)^s + \frac{sKr}{m} \left(1 - \frac{rK}{m}\right)^{s-1} \right) \\
 &\sim \sum_{r=0}^{m/K} \left(1 - \frac{d}{m}\right)^{m/k-r} \left(\frac{d}{m}\right)^r \binom{m/K}{r} \left(1 + \frac{sKr}{m}\right) e^{-rKs/m} \\
 &\sim \left(1 - \frac{d}{m} + \frac{de^{-sK/m}}{m}\right)^{m/K} \left(1 + \frac{s}{1 + (m/d - 1)e^{sK/m}}\right) \\
 &\sim \exp\left(-\frac{d}{K} + \frac{d}{K}e^{-sK/m}\right) \left(1 + \frac{sd}{m}e^{-sK/m}\right),
 \end{aligned}$$

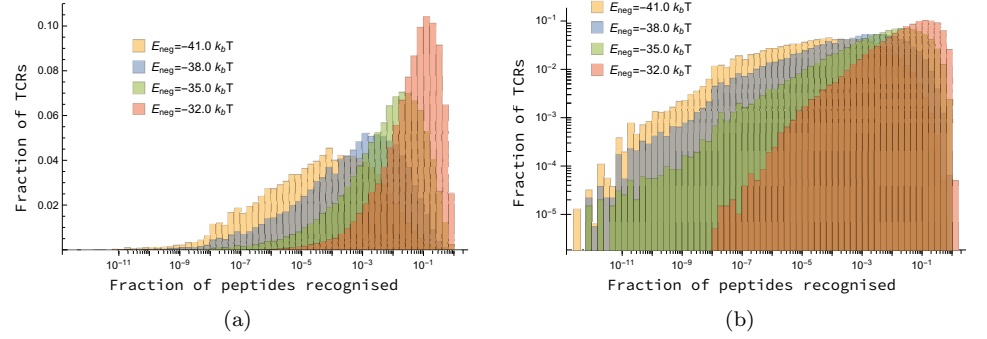
giving Eq (8). For $p = 3$ we have instead

$$\begin{aligned}
 P_1 &\sim \sum_{r=0}^{m/K} \left(1 - \frac{d}{m}\right)^{m/k-r} \left(\frac{d}{m}\right)^r \binom{m/K}{r} \\
 &\quad \times \left(\left(1 - \frac{rK}{m}\right)^s + \frac{sKr}{m} \left(1 - \frac{rK}{m}\right)^{s-1} + \frac{s(s-1)K^2r^2}{2m^2} \left(1 - \frac{rK}{m}\right)^{s-2} \right) \\
 &\sim \sum_{r=0}^{m/K} \left(1 - \frac{d}{m}\right)^{m/k-r} \left(\frac{d}{m}\right)^r \binom{m/K}{r} \left(1 + \frac{sKr}{m} + \frac{s(s-1)K^2r^2}{2m^2}\right) e^{-rKs/m} \\
 &\sim \left(1 - \frac{d}{m} + \frac{de^{-sK/m}}{m}\right)^{m/K} \\
 &\quad \times \left(1 + \frac{s}{1 + (m/d - 1)e^{sK/m}} + \frac{s(s-1)}{2} \frac{((K/d - K/m)e^{sK/m} + 1)}{(1 + (m/d - 1)e^{sK/m})^2}\right) \\
 &\sim \exp\left(-\frac{d}{K} + \frac{d}{K}e^{-sK/m}\right) \left(1 + \frac{sd}{m}e^{-sK/m} + \frac{s^2dK}{2m^2}e^{-sK/m} \left(1 + \frac{d}{K}e^{-sK/m}\right)\right),
 \end{aligned}$$

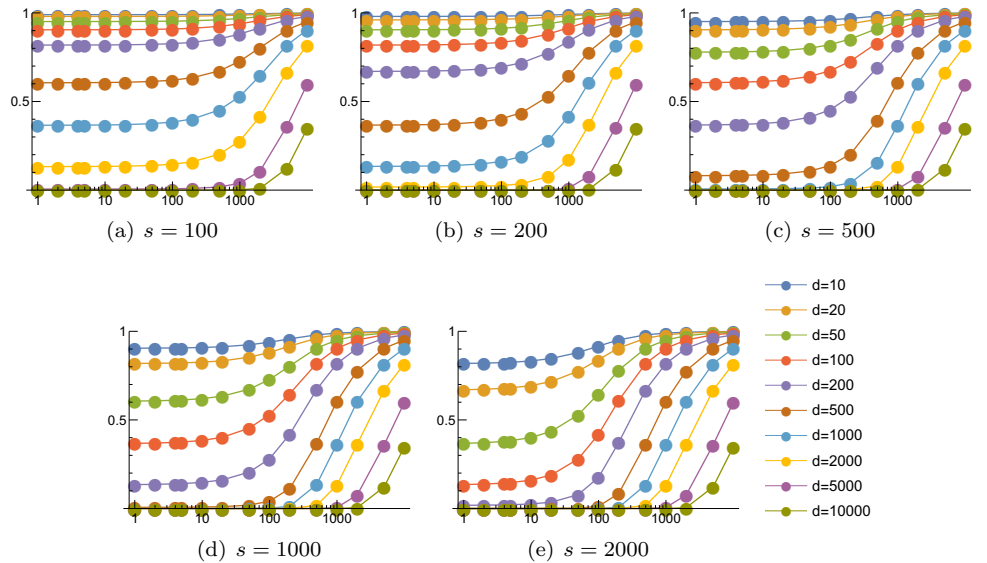
which is Eq (9).



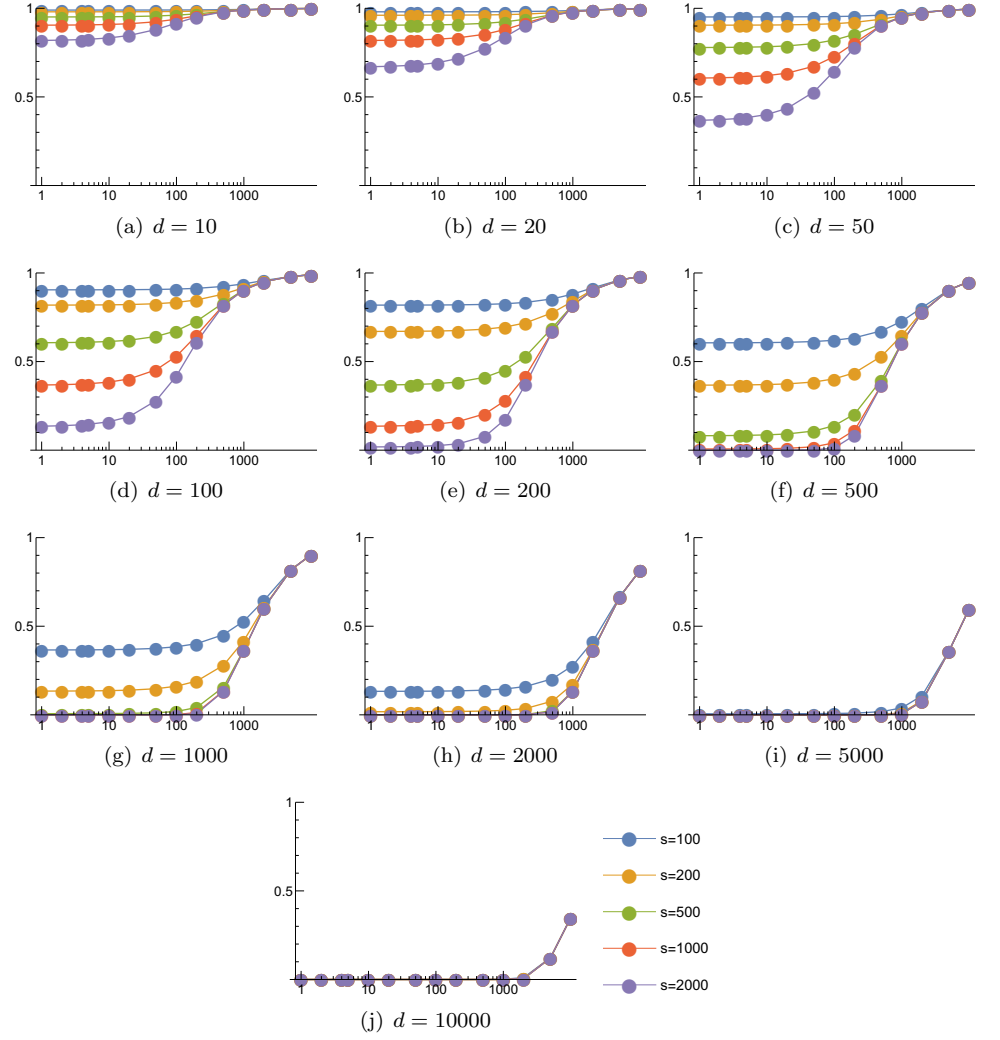
S1 Fig. Distribution of the proportion of peptides recognised by each of 200,000 randomly generated TCR sequences shown on (a) log-linear and (b) log-log scales for a variety of values of E_{neg} , with $N = 5$. The proportion of TCRs not recognising any peptides was approximately 6%, 12%, 19%, 31% and 44% for $E_{\text{neg}} = -20.0k_bT$, $-21.0k_bT$, $-22.0k_bT$, $-23.0k_bT$, and $-24.0k_bT$, respectively.



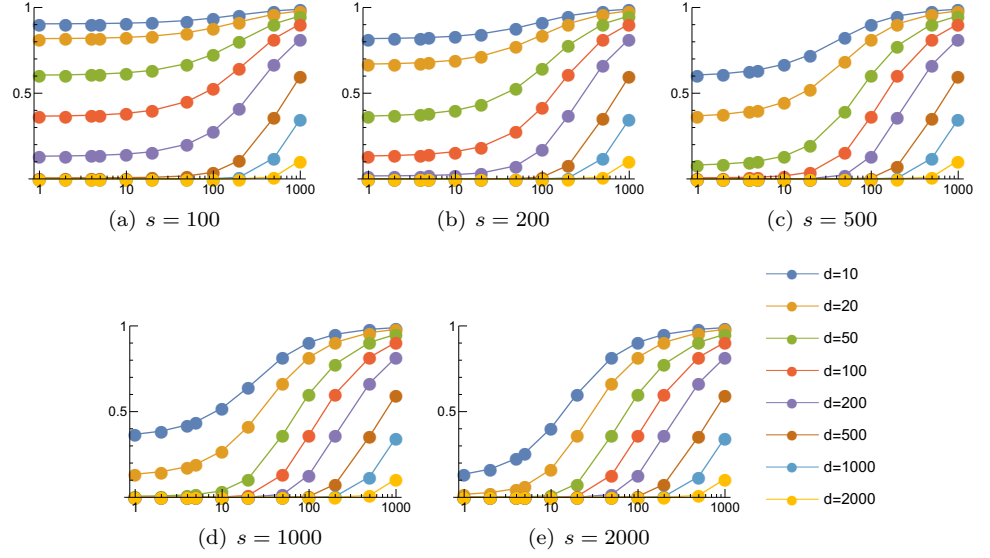
S2 Fig. Distribution of the proportion of peptides recognised by each of 200,000 randomly generated TCR sequences shown on (a) log-linear and (b) log-log scales for a variety of values of E_{neg} , with $N = 9$. The proportion of TCRs not recognising any peptides was approximately 0%, 0.4%, 4.4%, and 21% for $E_{\text{neg}} = -32.0k_bT$, $-35.0k_bT$, $-38.0k_bT$, and $-41.0k_bT$, respectively.



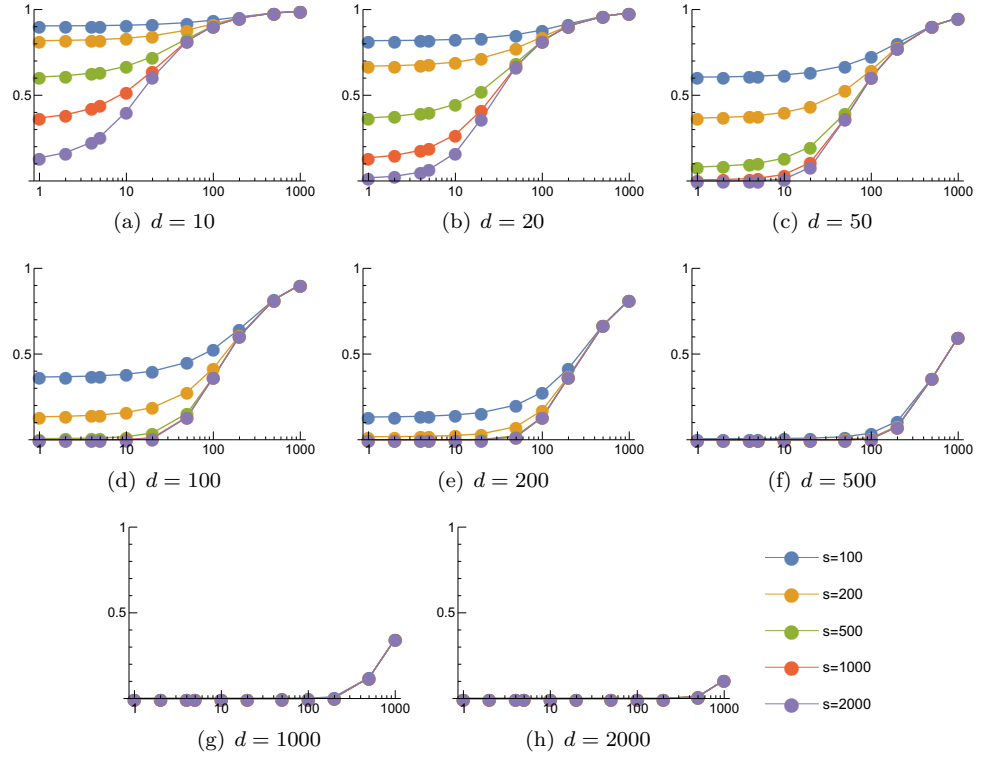
S3 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various TCR degeneracies, d . Other model parameters are $m = 100000$, $p = 1$.



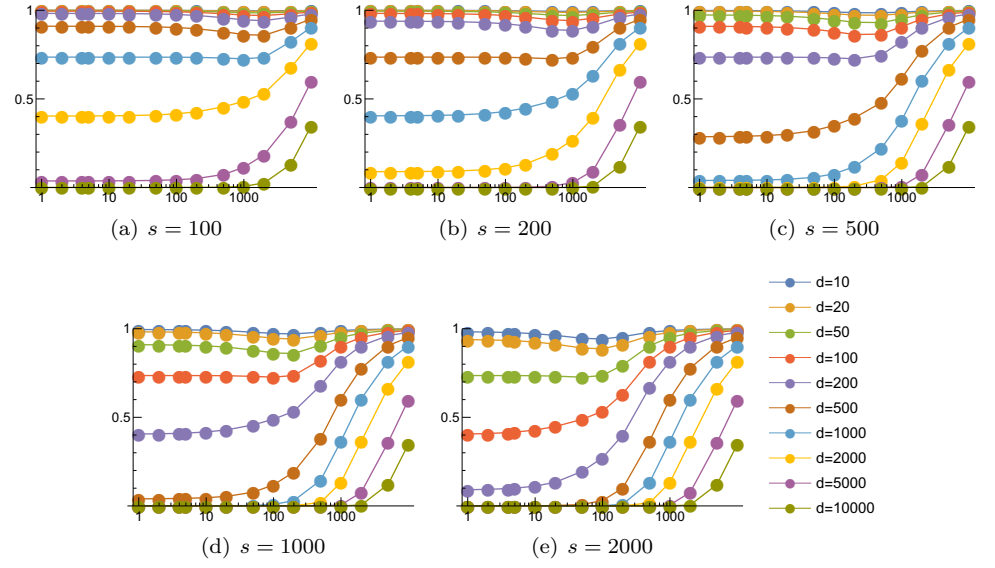
S4 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 100000$, $p = 1$.



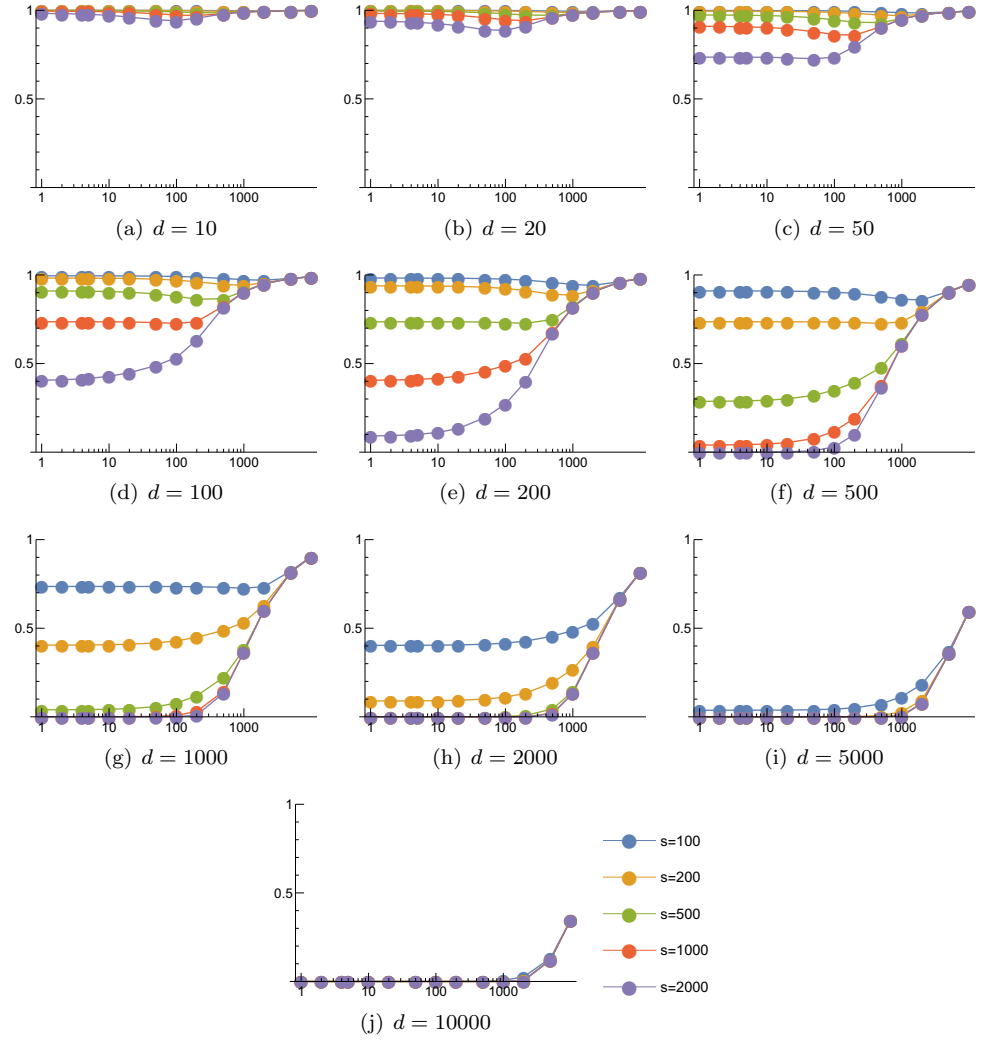
S5 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various TCR degeneracies, d . Other model parameters are $m = 10000$, $p = 1$.



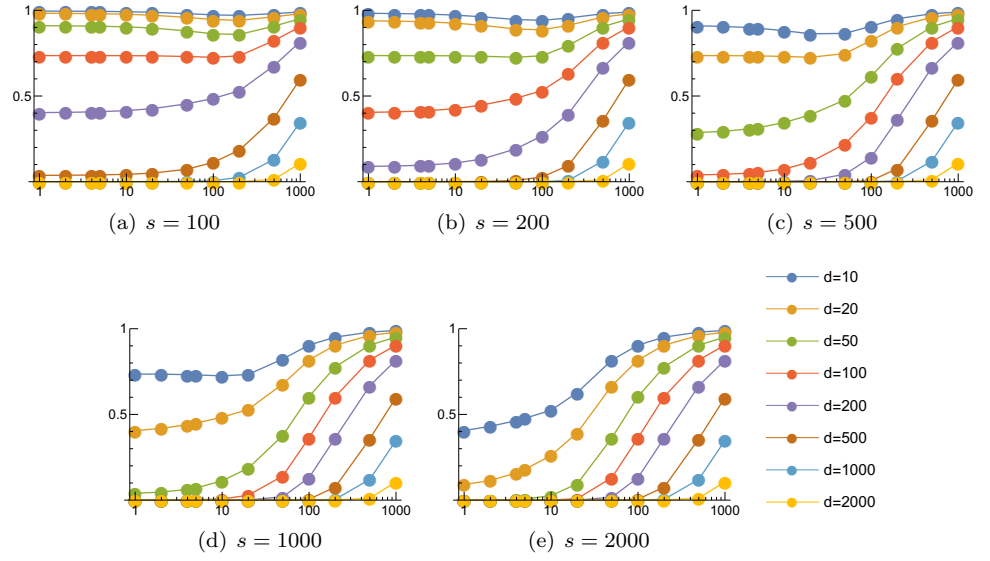
S6 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 10000$, $p = 1$.



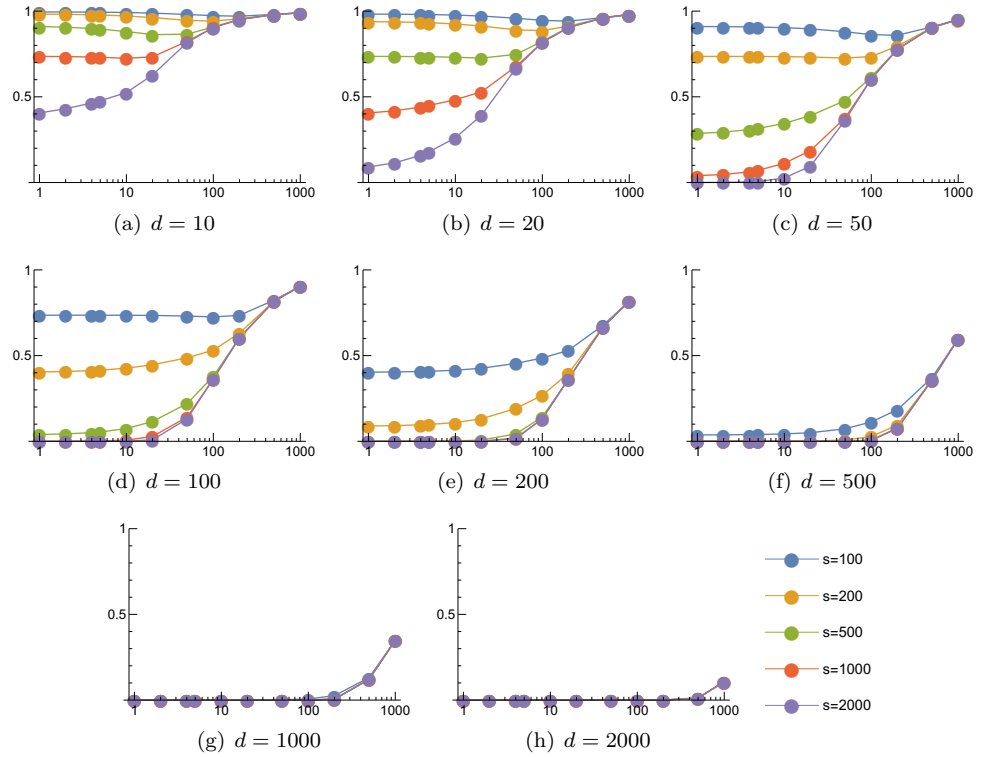
S7 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various TCR degeneracies, d . Other model parameters are $m = 100000$, $p = 2$.



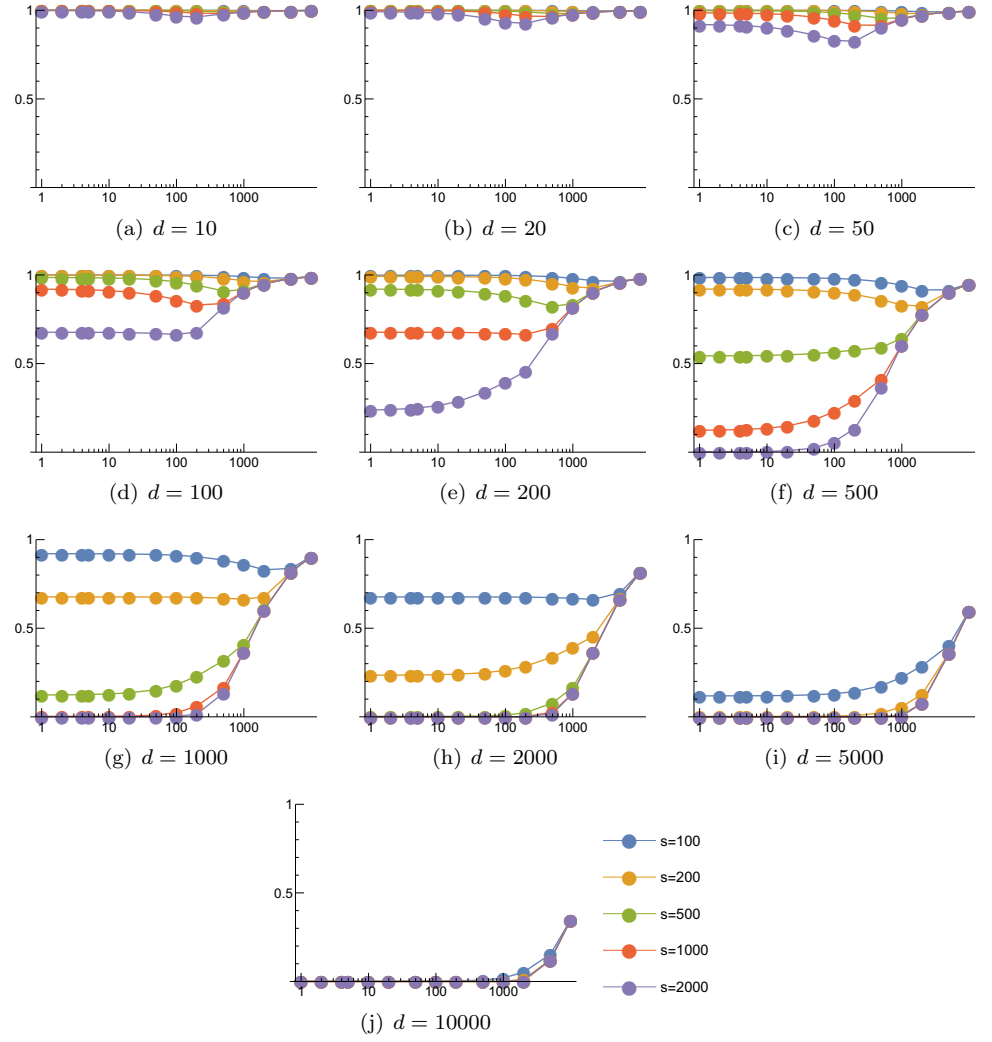
S8 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 100000$, $p = 2$.



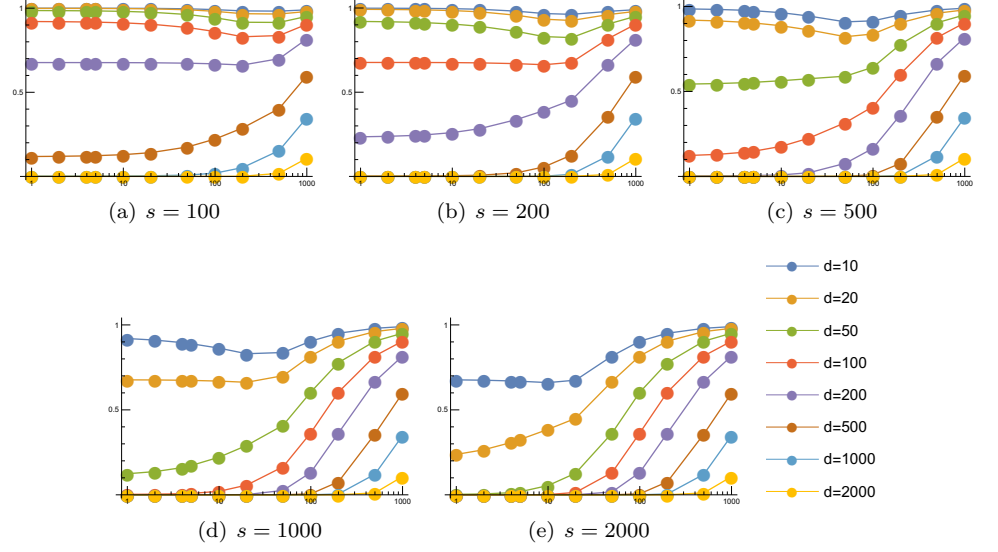
S9 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various TCR degeneracies, d . Other model parameters are $m = 10000$, $p = 2$.



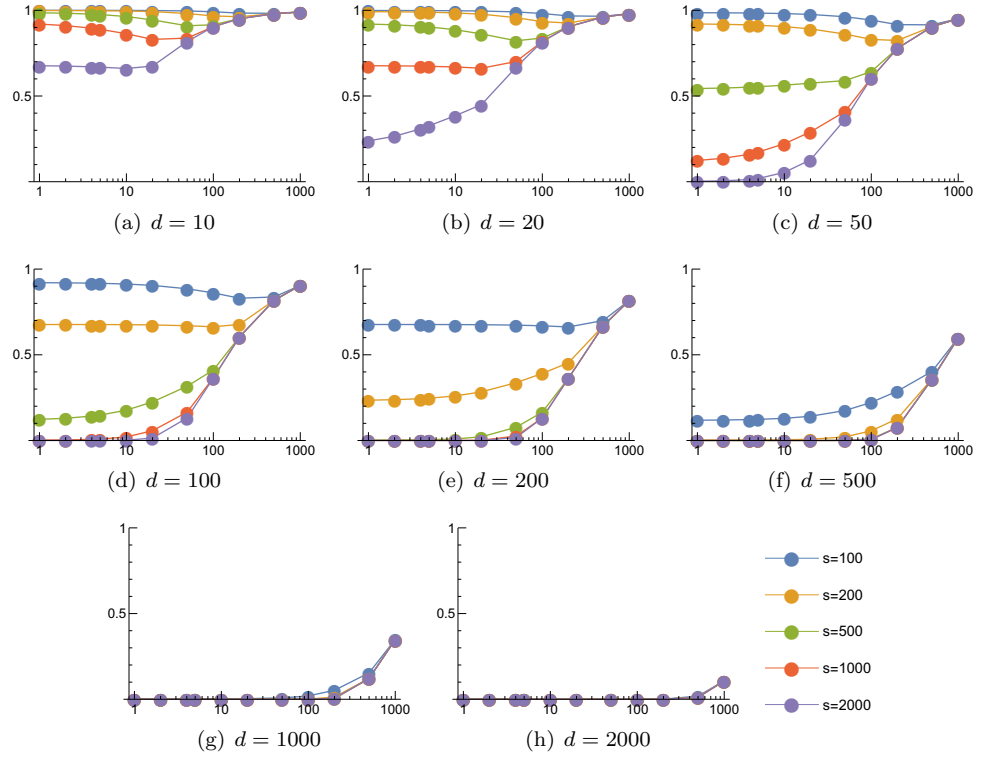
S10 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 10000$, $p = 2$.



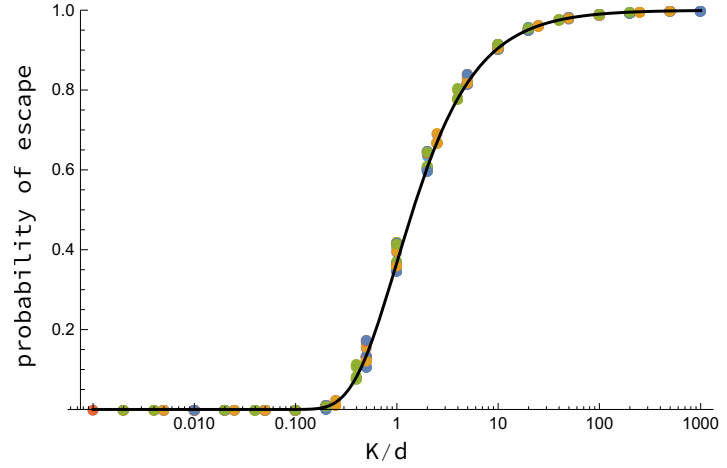
S11 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 100000$, $p = 3$.



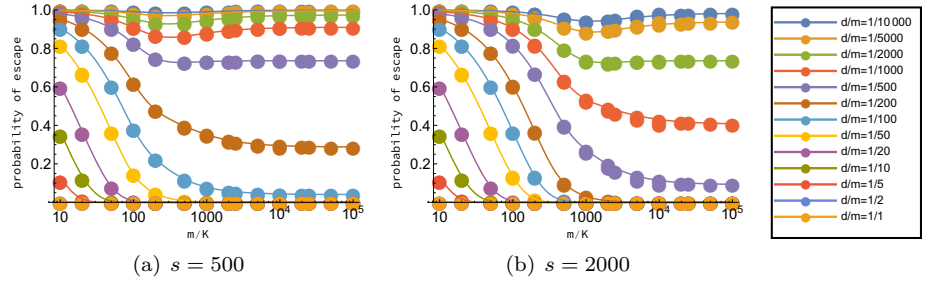
S12 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various TCR degeneracies, d . Other model parameters are $m = 10000$, $p = 3$.



S13 Fig. Mean probability of escape in a single interaction, P_1 , as a function of the number of mTEC classes, K , for various numbers of TCR-pMHC complexes in the immunological synapse, s . Other model parameters are $m = 10000$, $p = 3$.



S14 Fig. The mean probability of escape as a function of K/d , illustrating the collapse of the data when $s > 2m/K$. The curve is given by Eq (6). The points are calculated using all those combinations of $s = 100, 200, 500, 1000, 2000$, $K = 1, 2, 4, 5, 10, 20, 50, 100, 200, 500, 1000$, $m = 10^4, 10^5$, and $d = 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000$, which satisfy the constraint $s > 2m/K$ (and $d \leq m$). The plot shows data for $p = 1$. For $p = 2$ the requirement is more like $s > 3.5m/K$; for $p = 2$ the requirement is $s > 5m/K$.



S15 Fig. The mean probability of escape as a function of m/K . The curves are given by Eq (8). The points are calculated values, using both $m = 10^4$ and 10^5 . For $s = 500$ these points lie on top of one another. For $s = 2000$ the difference between the two sets of points is just visible near $m/K = 10^4$. Here $p = 2$.