Reliable *in silico* ranking of engineered therapeutic TCR binding affinities using MMPBSA and MMGBSA calculations.

Rory M. Crean [1,2,#a,*], Christopher R. Pudney [1,3], David K. Cole[4,5], Marc W. van der Kamp[6,*]

[1]Department of Biology and Biochemistry, [2]Doctoral Training Centre in Sustainable Chemical Technologies, [3]Centre for Therapeutic Innovation, University of Bath, Bath, UK.
[4]Immunocore Ltd., Milton Park, Abingdon, UK.
[5]Division of Infection & Immunity, Cardiff University, Cardiff, UK.
[6]School of Biochemistry, University of Bristol, Biomedical Sciences building, Bristol, UK.
[#a]Current Address: Science for Life Laboratory, Department of Chemistry − Biomedicinska Centrum, Uppsala University, Uppsala, Sweden.

* Corresponding authors

E-mail: rory.crean@kemi.uu.se (RMC).

E-mail: marc.vanderkamp@bristol.ac.uk (MWvdK).

Abstract

Accurate and efficient in silico ranking of protein-protein binding affinities is useful for protein design with applications in biological therapeutics. One popular approach to rank binding affinities is to apply the molecular mechanics Poisson Boltzmann/generalized Born surface area (MMPB/GBSA) method to molecular dynamics trajectories. This provides a compromise between rapid but approximate scoring functions of single structures and more sophisticated methods such as free energy perturbation. Optimal MMPB/GBSA parameters tend to be system specific. Here, we identify protocols that enable reliable evaluation of the effect of mutations in a T-cell receptor (TCR) in complex with its natural target, the peptide-human leukocyte antigen (pHLA). The development of affinity-enhanced engineered TCRs towards a specific pHLA is of great interest in the field of immunotherapy. Our study highlights the importance of using a higher than default internal dielectric constant, especially in the case of charge changing mutations. Including explicit solvation and/or entropy corrections may deteriorate the ranking of single point variants due to the errors associated with these additions. For multi-point variants,

however, these corrections were important for accurate ranking. We also demonstrate how potential outliers could be identified in advance by analyzing changes in the hydrogen bonding networks at the binding interface. Finally, using bootstrapping we show that as few as 5-10 replicas of short (4 ns) MD simulations may be sufficient for reproducible and accurate ranking of candidate TCR variants. Our work demonstrates that reliably ranking TCR variant binding affinities can be achieved at moderate computational cost. The protocols developed here can be applied towards *in silico* screening during the optimization of therapeutic TCRs, potentially reducing both the cost and time taken for biologic development.

Author summary

T-cell receptors (TCRs) recognise an extracellular presentation platform that displays peptides that can be derived from infectious agents, or mutated/dysregulated cellular proteins. Soluble TCRs can be engineered to bind with enhanced affinity and specificity towards a given protein-peptide complex displayed only on diseased cells (*i.e.,* cancerous cells), providing the basis for potent biologic drugs. Here, we performed computer simulations that describe how atoms and molecules interact with one another with atomistic detail in order to predict how the binding affinity (*i.e.*, how sticky the drug is to its target) changes upon mutation of the TCR. By systematically testing different settings and methodologies for our binding affinity calculations, we identified optimized protocols to more accurately predict how a given mutation (or set of mutations) will impact the binding affinity. We also identified differences in the optimal protocol depending on the number of mutations that separate the TCRs. Our resulting calculated affinities correlated very well with experimental results, showcasing the promise of incorporating our computational approach into the design of affinity-enhanced TCRs for therapeutic use. The protocols developed can lead to faster and less expensive development of effective TCR-based treatments for various cancers and other diseases.

Introduction

Computational methods to predict the binding affinities of protein-protein interactions (PPIs) that are sufficiently accurate, reliable and high throughput have clear potential for application towards the rational design of biologic drugs. Many approaches (all with many variations available) including free energy perturbation (FEP), umbrella sampling, molecular docking and machine learning have all been applied to predict or rank order PPI binding affinities.[1–4] Here, we focused on the molecular mechanics Poisson Boltzmann/generalized Born surface area (MMPB/GBSA) approach,[5] which combines conformational sampling using molecular dynamics (MD) simulations with empirical calculations on these snapshots to estimate the binding free energy. This approach can be thought of as sitting somewhere in between the more accurate but more computationally expensive FEP method, and less accurate but computationally cheaper methods like docking.[6] This approach should only be relied on for relative binding affinities (*i.e.,* $\Delta\Delta G$ not ΔG) to rank order a set of similar potential drug candidates.[6] An advantage of the MMPB/GBSA approach is that it can be decomposed to obtain per-residue contributions to the binding energy, which we and others have used to identify key residues and interactions which drive protein-protein binding.[7–9] The information obtained from this decomposition analysis can be used to inform (semi-)rational design efforts towards enhanced affinity and/or selectivity drug candidates.[8,10]

MMPB/GBSA has been used and evaluated extensively for many applications, and it is clear that tuning of the parameters and protocols applied can give significant improvements in accuracy, with such tuning typically being system specific (see *e.g.* refs. [6,11–16]). With this in mind, we aimed to identify an MMPB/GBSA protocol that provides reliable and accurate relative binding free energies for a PPI of great interest in the field of immuno-oncology,[17] the T-cell receptor (TCR) peptide-human leukocyte antigen (pHLA) complex (TCR-pHLA, **Fig 1**). The TCR-pHLA interaction is a vital component of the adaptive immune system, with the TCR ultimately responsible for selectively binding specific peptide sequences presented on the surface of cells by the HLA. For HLA Class 1 proteins (the focus of this

study), the peptides in pHLA complexes are sourced from proteins digested inside the cell: each cell presents peptide fragments of its cellular proteins on the extracellular surface. In the natural immune system, TCRs can specifically identify antigenic peptide sequences presented by unhealthy (*e.g.* cancerous) cells, enabling the immune system to subsequently eradicate these cells.[18]

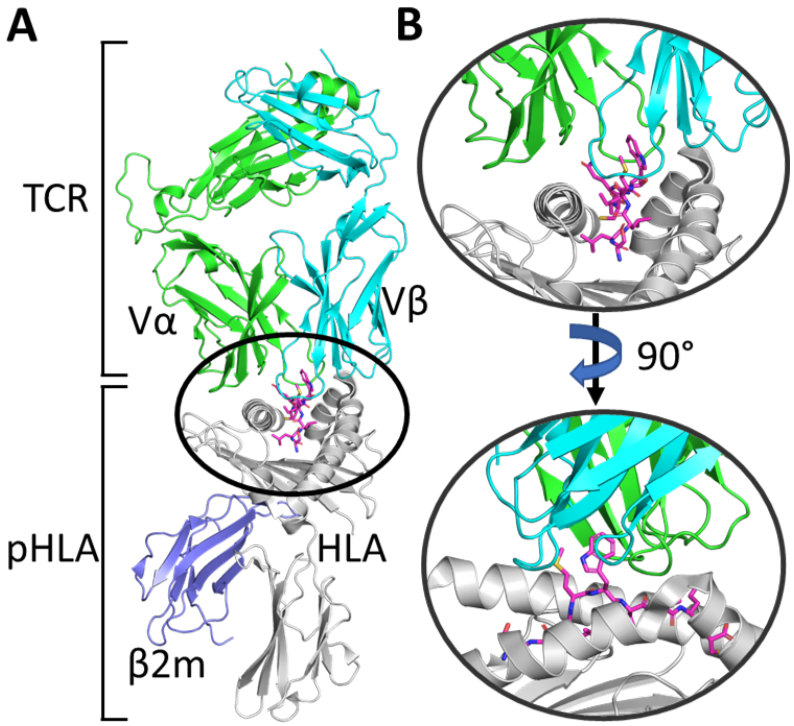


Fig 1. A. Overview of the TCR-pHLA complex. The T-cell receptor (TCR) is comprised of two (α and β) domains, which engage the peptide-human leukocyte antigen (pHLA) complex. **B.** Zoom in on the TCR-pHLA binding site from two different angles, demonstrating that the binding interface is composed of six CDR loops on the TCR, which engage both the peptide and two α-helices on the pHLA complex.

Taking advantage of these natural mechanisms of the immune system, affinity-enhanced, soluble, engineered TCRs are a class of therapeutic molecules which are designed to target a specific antigenic pHLA complex presented only by unhealthy (*e.g.* cancerous) cells, whilst simultaneously not binding the considerably large number of other pHLA complexes presented by "healthy" cells (in order to avoid

off-target toxicity).[19] This provides two deeply intertwined engineering challenges which must be addressed in order to produce a therapeutic TCR.[20] That is, TCRs must have both strong affinity (natural TCRs have affinities in the ~μM range,[21] whilst therapeutic soluble TCRs are in the ~pM range) and high specificity (to avoid the large number of off-targets). Whilst most TCR affinity engineering studies reported in the literature have obtained affinity enhancement through experimental approaches (primarily those that utilize *in vitro* selection),[22–27] docking[28,29] and structure based-rational design[30] have also been successfully applied to engineer TCRs. Here, we envisage MMPB/GBSA as a technique that could be used to filter promising candidate mutations generated through a more high-throughput technique such as docking prior to experimental screening.

To date there has been no systematic study on how best to predict TCR-pHLA binding affinities using MMPB/GBSA, and herein, we aim to resolve this. To do this, we have evaluated a variety of MMPB/GBSA calculation protocols using two different TCR-pHLA test sets, one with 18 TCR variants with between 3-14 mutations (spread over most CDR loops) and one with 29 variants, of which 25 have just one mutation. The use of these two disparate test sets should allow us to identify a single protocol (if one exists) that works for both TCR-pHLA complexes and thus may be generally applicable for this biologically and therapeutically important protein-protein interaction.

Results and discussion

In this study, we assessed the capability of MMPB/GBSA calculations to reproduce TCR-pHLA binding affinity relationships on two different test sets. The first (1G4) test set, is composed of 18 TCR variants all containing between 3-14 mutations from the WT, with mutations spread between 1-5 CDR loops.[26,31] In contrast, the second (A6) test set is composed of 29 TCR variants, with 25 of these being single point variants and the remaining 4 baring between 2-4 mutations.[29,32] The aim is to identify a protocol that is not only reliable and reproducible, but that works well for both test sets studied here. With this in mind, we build on recommendations of others[15,33–36] in our use of many replicas of short MD simulations to obtain snapshots for MMPB/GBSA calculations. This "ensemble" based approach has been shown to outperform single or few replica simulations of much longer length, both in terms of reliability and predictability.[33,35] Specifically, we performed 25 independent MD simulations of 4 ns long were performed, and used frames collected every 10 ps from the last 3 ns of each as input for MMPB/GBSA calculations (meaning a total of 7500 frames were used per TCR-pHLA complex). The prediction accuracy was assed using both Pearson's r (r_p) and Spearman's rank (r_s), as r_p measures the linear correlation between experiment and simulation whilst r_s asses the ability to rank order binding affinities (arguably of more importance in a design context).

Modulation of the internal dielectric constant drastically improves predictability.

For both the 1G4 and A6 TCRs test sets investigated herein, we assessed the ability of both MMPBSA and MMGBSA to predict relative binding affinities (**Fig 2**). Further, given previously reported successes at improving the quality of prediction for other systems,[6,37–40] we assessed the benefit of modifying the internal dielectric constant (ϵ_{int}) for the MMPBSA calculations.

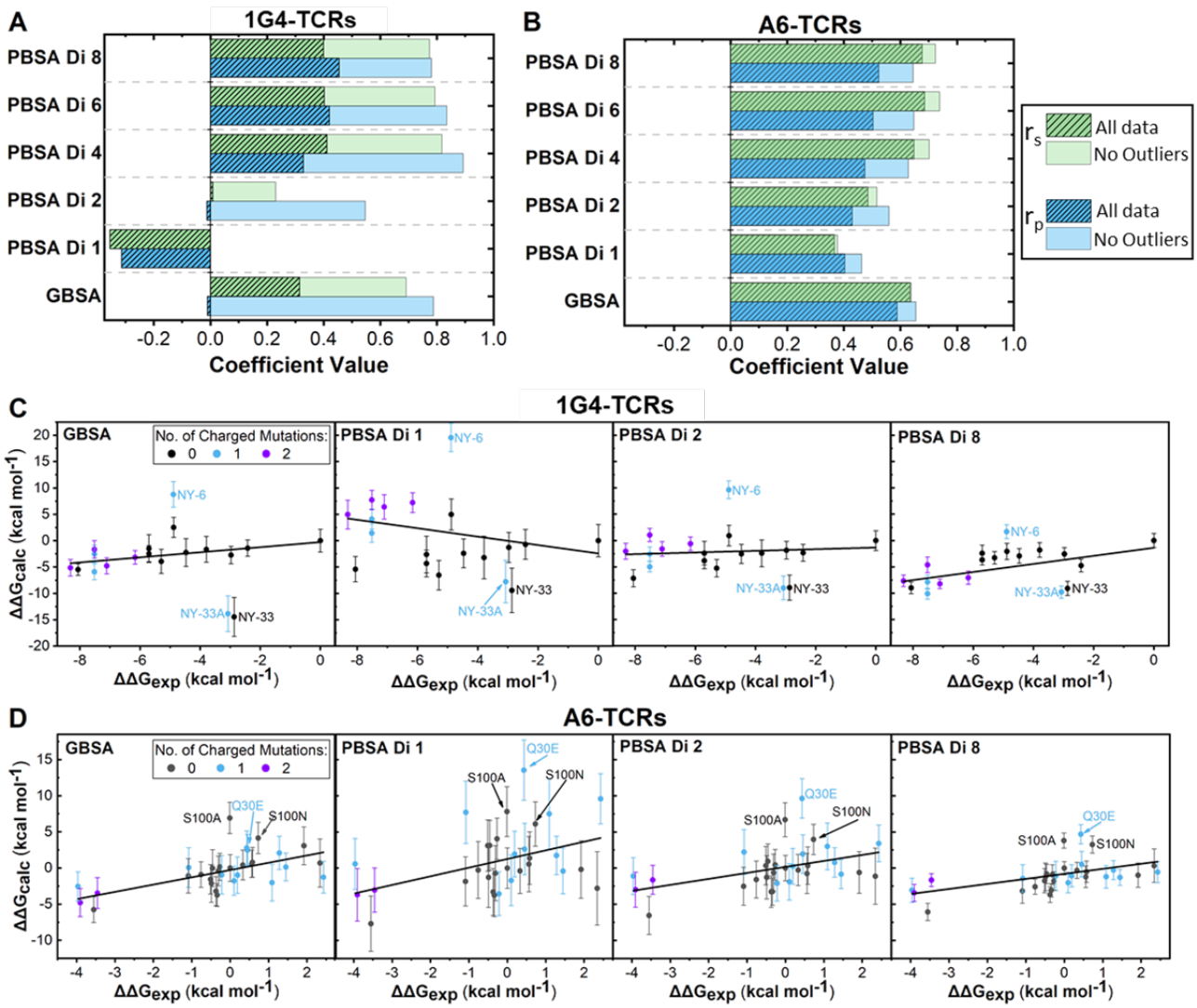


Fig 2. Modulation of the interior dielectric constant improves MMPBSA predictability. A+B. Determined Spearman's rank (r_s) and Pearson's r (r_p) values for MMPB/GBSA calculations for the 1G4 (**A**) and A6 (**B**) test sets. Results are plotted with and without the three identified outliers described in the text for both data sets. **C+D.** Exemplar scatter plots with lines of best fit for the 1G4 (**C**) and A6 (**D**) test sets using either MMGBSA or MMPBSA (at different internal dielectric constants) methodology. For **C**+**D**, outliers are labelled. Scatter plots in panels **C**+**D** are also colored according to the number of charged mutations made between the variant and the WT. Complete scatter graphs for all results are provided in **S1** and **S2 Figs**.

First, in the 1G4 test set, and to a lesser extent the A6 test set, increasing the ϵ_{int} used in the MMPBSA calculations progressively improves the prediction quality, with the effect largely flattening

out for internal dielectric constants in the range of 4-8 (**Fig 2A+B**). We also note that the standard deviations (SD) obtained from the 25 replicas for individual $\Delta\Delta G$ measurements reduce as ϵ_{int} increases, with this effect again flattening out for ϵ_{int} values between 4-8 (**Fig 2C+D** and **S1** and **S2 Figs**). For example, the average SD reduces from 2.8 to 1.3 kcal mol^{-1} and 3.5 to 1.7 kcal mol^{-1} when ϵ_{int} is increased from 1 to 4 for the 1G4 and A6 TCR systems, respectively. This data suggests that fewer replicas per variant may be required in order to obtain converged results when a higher ϵ_{int} value is used. Interestingly, for both test sets, the GB solvent model significantly outperforms the PB solvent model (at an ϵ_{int} of 1). This is perhaps surprising given that the GB solvent model is designed to reproduce the PB solvent model with an ϵ_{int} of 1.[41]

It is challenging to provide a concrete answer as to the reason why increasing the ϵ_{int} can improve the quality of prediction, and why the 1G4 test set is more sensitive to this effect than the A6 test set. A recent MMPBSA study focused on predicting the correct binding pose for PPIs observed a weak relationship between the polar buried area (PBA) and the optimal ϵ_{int} to use.[37] Systems with increasing PBA were recommended higher ϵ_{int} values, and based on the PBA of the WT TCR-pHLA complexes studied here (1310 Å^2 and 1250 Å^2 for WT-1G4 and WT-A6 respectively, determined using the COCOMAPS webserver[42]) a ϵ_{int} of approximately 2-4 would be recommended. Further, several MMPBSA alanine scanning studies have found the use of ϵ_{int} values greater than 1 to greatly improve the quality of prediction for the exchange of charged residues.[16,43–46] To assess the possibility that the outliers observed in the MMPBSA calculations with an ϵ_{int} of 1 were induced by changes in the charge of the TCR, we colored variants in **Fig 2C+D** according to the total number of charged mutations made from the WT. The benefit that increasing ϵ_{int} has on charged variants is clear for both data sets, but particularly striking for the 1G4 test set, as several affinity-enhanced variants (with $\Delta\Delta G$ values < −6 kcal mol^{-1}) are progressively reordered from some of the lowest affinity variants to some of the highest affinity variants.

For the 1G4 TCRs, three apparent outliers can be identified even at higher ϵ_{int} values or when using the GBSA approach (**Fig 2C**), and their negative impact is clear when comparing the prediction quality with and without the outliers included (**Fig 2A**). Analysis of the CDR loop sequences of these TCR variants (**Fig 3A**) shows 5 mutations are made in their CDR3α loop which are not present in any of the other variants studied here (see **S4 Table** for all sequences used). These differences in the CDR3α loop could therefore explain why these variants are outliers in the above data set. That is, these mutations may have notably altered the conformational dynamics/sampling of the loop (and/or neighboring regions), and this would likely not be accounted for by the short MD simulations (which start using the same backbone crystal structure as described in the **Methods**) performed here. Alternatively, there could be a significant change in the rigidity of the CDR3α-loop, such that the contribution from changes in solute entropy upon binding cannot be ignored for accurate ranking of these variants.

For the A6 TCRs, 3 single point variants (Q30E and S100A/N on the TCR α-chain) were consistently underestimated (**Fig 3D**). Our 3D-RISM calculations on WT A6 (used to solvate the protein due to the lack of water molecules available in the X-ray structure, see **Methods**) predicted strong affinity bridging water molecule sites at both of the above mutation sites in the WT A6-TCR (**Fig 3B+C**), with the 3D-RISM distribution function ($g(\boldsymbol{r})$) for water oxygen atoms calculated to be > 10 (note that the $g(\boldsymbol{r})$ of bulk water is by definition 1). Further, both mutated side chains are predicted to make water-bridged hydrogen bonds with pHLA residues (**Fig 3B+C**). Specifically, HLA residues K66 and T163 for Q30 and peptide residue P6 for S100). Taken together, our data suggests that the outlier mutations may be poorly described due to not explicitly describing key solvent meditated water bridged hydrogen bonds through the use of an implicit solvent model (PB or GB) in our calculations.

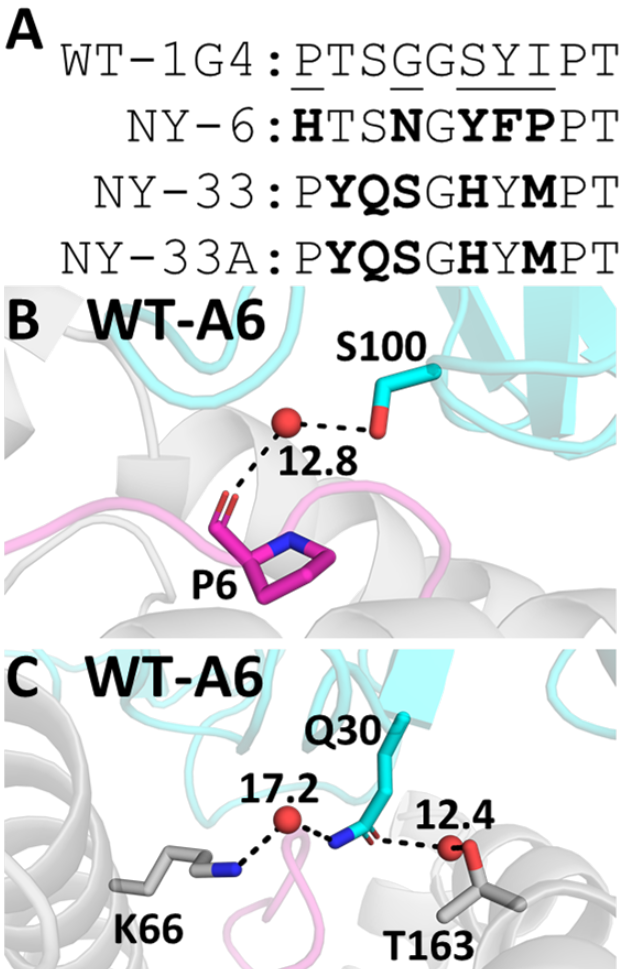


Fig 3. A potential rational for outliers identified in our MMPB/GBSA Calculations. **A.** Sequences of the CDR3α loop of the three 1G4 outliers, with positions mutated shown in bold. All 1G4 variant sequences are provided in **S4 Table**. **B+C.** WT A6 TCR-pHLA structure with the two outlier mutation sites S100 (**B**) and Q30 (**C**) labelled. Predicted water sites (using 3D-RISM[47,48] and Placevent[49], see **Methods**) which form bridged water hydrogen bonds to pHLA residues are shown (here, all donor-acceptor heavy atom distances are within 3 Å). The calculated water density distribution function $g(\boldsymbol{r})$ is shown for the water molecules, demonstrating that they are all predicted to have a very high occupancy.

Given the above observations, in the following sections we aimed to try to correct the outliers observed in both data sets and improve our overall prediction accuracy. We did this by (1) including explicit water molecules into our MMPB/GBSA calculations, and (2) introducing a correction for the

change in solute entropy. Further, we note that our primary aim is to identify an approach that is ideally suitable for all TCR-pHLA complexes. It is therefore important to assess whether the inclusion of explicit water molecules and entropic corrections could have a deleterious effect on the overall quality of prediction (*i.e.,* through the introduction of an additional source of error and/or noise).

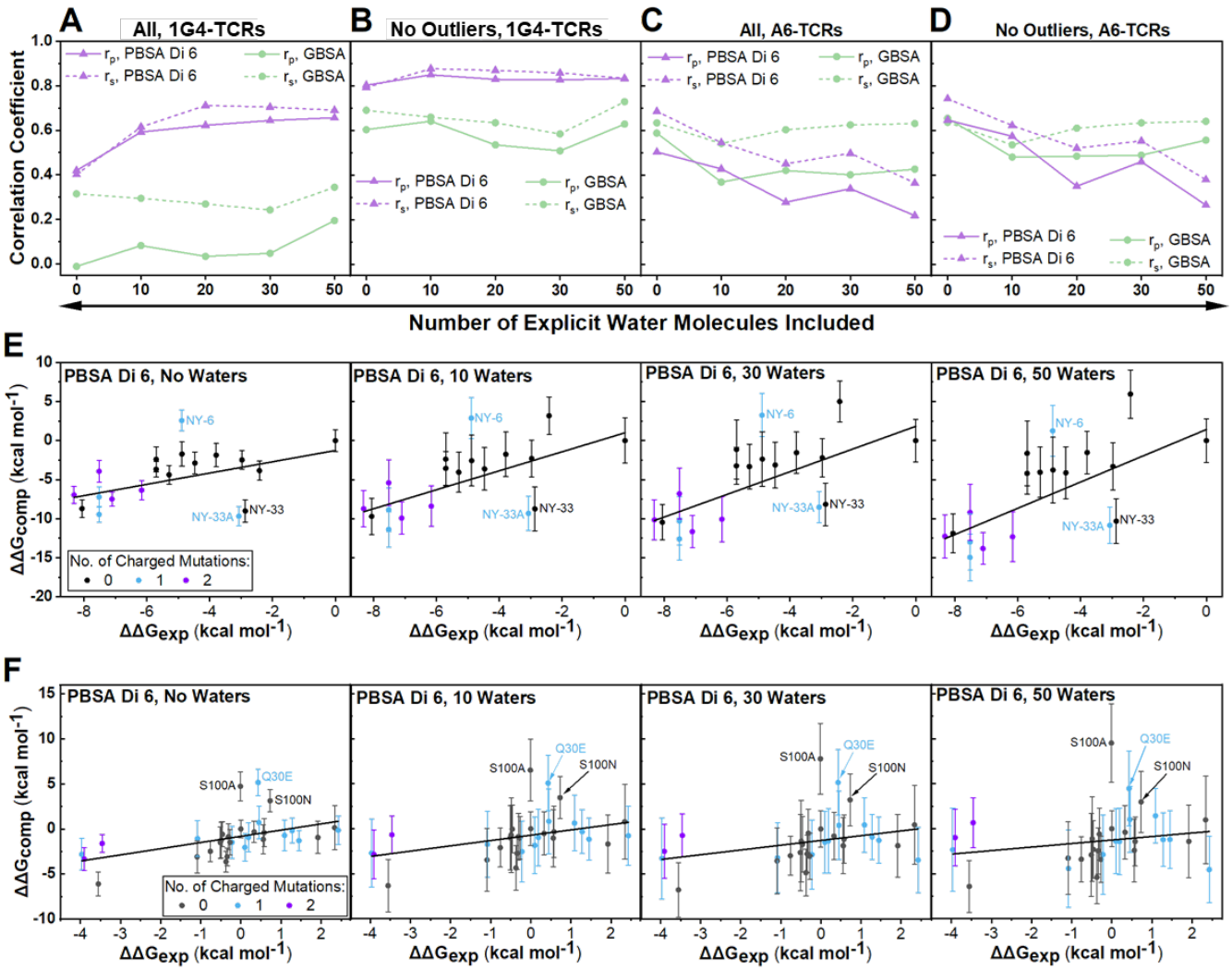
Effect of Inclusion of Explicit Water Molecules.

The inclusion of explicit water molecules has shown mixed success in the context of MMPB/GBSA calculations.[11,14,50–52] In the case of a small molecule ligand binding a protein receptor it is common practice to include the "X" closest water molecules to the ligand and retain these water molecules for the receptor calculation (as well as the complex calculation). In contrast, for a PPI, there are many possible ways to define which water molecules should be kept in the calculation and further, whether these waters are retained on the receptor or the ligand or some combination of both. Here, we took the X (where X is 10, 20, 30 or 50) closest waters to any oxygen or nitrogen atom on a selection of binding site residues located on the pHLA, and included these waters as part of the pHLA (*i.e.,* receptor) calculation, as well as the complex calculation (see **Methods** for further details). We choose to keep all waters on the pHLA over a combination of the TCR and pHLA to ensure that all retained waters were close to a protein atom in both the bound and unbound MMPB/GBSA calculations. Given the results obtained for different solvent models and dielectric constants as described in **Fig 2**, we assessed the benefit of including explicit water molecules using both the MMGBSA and MMPBSA methods, setting ϵ_{int} to 6 for the MMPBSA calculations (**Fig 4**).

Focusing first on the 1G4 set of TCRs, a beneficial effect is observed when explicit water molecules are included in the MMPBSA calculation with ϵ_{int} set to 6 (**Fig 4A+B**). Improvements in the prediction quality are observed both including and excluding the previously described outliers. This additional benefit appears to be largely due to correctly ranking the highest affinity TCR variants (those with $\Delta\Delta G_{\text{exp}} < -6$ kcal mol^{-1}). Further, most of the beneficial effect of including explicit water molecules is

observed after only 10 waters are included, with the improved prediction quality remaining fairly stable with increasing numbers of waters included. This observation of a lack of sensitivity to differing numbers of explicit waters is reassuring to note (as it is impossible to know the optimal number of waters to include *a priori*).

Fig 4. Impact of explicit water molecules on binding affinity predictions. A-D. Determined Spearman's rank (r_s) and



Pearson's R (r_p) values for MMPB/GBSA calculations on the 1G4 (**A**+**B**) and A6 (**C**+**D**) test sets for different numbers of explicit water molecules included in the calculation. **E+F**. Exemplar scatter plots for the 1G4 (**E**) and A6 (**F**) test sets showing the impact of the inclusion of an increasing number of explicit water molecules when using the MMPBSA method with ϵ_{int} set to 6. Scatter points are colored according to the number of charged mutations made between the variant and the WT. Complete scatter graphs for all results are provided in **S3-S6 Figs**.

In contrast to the MMPBSA calculations, the inclusion of explicit water does not significantly improve the correlations for the MMGBSA approach. It should be noted, however, that the inclusion of explicit solvent increases the standard deviation obtained for the individual affinity estimates (**Fig 4E+F** and **S3** and **S4 Figs**). For both the MMPBSA and MMGBSA simulations of the 1G4 test set, this increased deviation is partially compensated for by sampling a larger range of affinities. For instance, the MMPBSA $\Delta\Delta G_{calc}$ values vary by up to 12.2 kcal mol^{-1} for calculations with no water as compared to up to 20.9 kcal mol^{-1} for calculations with 50 waters included (with the three outliers described above removed, the variations for no water or 50 waters are 9.5 and 20.9 kcal mol^{-1}, respectively). In contrast, the range of $\Delta\Delta G_{calc}$ values obtained for the A6 TCRs does not change significantly with increasing numbers of waters (**Fig 4F** and **S5** and **S6 Figs**), indicating that the impact of the increased standard deviations observed may be particularly detrimental to the prediction accuracy for the A6 test set (as this therefore implies increased noise in the dataset).

As the A6 TCR data set consists primarily of single point mutations, whist the 1G4 set is composed entirely of multi-point variants, it is important to consider how significant the contribution of explicit water molecules is in describing the differences in affinity between variants (*i.e.,* $\Delta\Delta G$ not ΔG). That is, mutations that do not (significantly) interrupt the solvation environment between the TCR and pHLA may not require explicit solvation in order to correctly rank their relative affinities, and instead the increased noise associated with the calculation may just worsen the prediction quality. One would expect single point mutations to significantly disrupt the water network less often than the multi-point mutants present in the 1G4 test set, which is consistent with our observations shown in **Fig 4**. Further, the A6 TCR model was solvated based on 3D-RISM calculations, as no crystallographic waters were resolved (due to the low resolution of the structure). This may therefore also provide a source of error, if any key binding site water molecules were incorrectly placed.

To try to identify how the binding site solvation environment may have changed for TCR-pHLA complexes with different TCR variants, we calculated the total average number of bridged water hydrogen bonds (Hbonds) as well as solute-solute Hbonds formed between the TCR and pHLA during our MD simulations (**S7 Fig**). Whilst in the A6 data set we do observe a notable decrease in the average number of bridged water Hbonds for the Q30E variant (one of the outliers described above) as compared to the WT, other variants show largely similar values, consistent with a largely unchanged binding site water network. We also note that the 1G4 outliers NY-33 and NY-33A have the largest number of solute-solute Hbonds (**S7 Fig**), approximately 3 more Hbonds than most of the rest of the 1G4 test set. Our binding affinity calculations overestimate these outliers' affinities (**Figs 2** and **4**, where no solute entropy correction term has yet been considered). This could suggest that enthalpy-entropy compensation is important for correctly ranking these outliers.[53] That is, with additional Hbonds between the TCR and pHLA, one may expect a more favorable binding enthalpy term, which could be offset to a large degree by a less favorable binding entropy term.

Impact of Solute Entropy Corrections.

We evaluated two different methods to determine the change in solute entropy ($T\Delta S$). The first is a modified version of the normal mode analysis (NMA) approach. In this approach, snapshots from MD are subjected to energy minimization and vibrational frequency calculations to obtain an estimate of the configurational entropy for each state. This approach is often not used in MMPB/GBSA applications due to its sizable computational cost and the large standard deviations obtained, which can often worsen the prediction quality.[6,54] However, Kongsted and Ryde introduced a modified approach whereby NMA is performed on a truncated region around the binding site, with a "buffer" region of amino acids and water molecules fixed in place to stabilize the conformation of the structure (**Fig 5**).[55] This approach, referred to as truncated-NMA (Trunc-NMA) has been demonstrated to significantly reduce

the computing time associated with the calculations, as well as reducing the magnitude of the error values obtained.[55–57] Given that we did not expect entropy corrections to improve predictions for the A6 test set with (primarily) single point mutations, alongside the substantial computational cost of Trunc-NMA, we applied this approach on only the 1G4 set of TCRs.

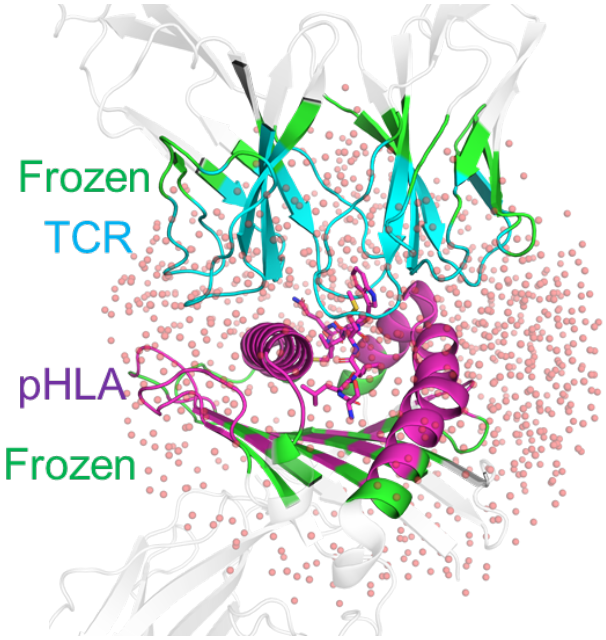


Fig 5. Illustration of the truncated-normal mode analysis (Trunc-NMA) method used to calculate a solute entropy correction for the 1G4 test set. Residues included in the Trunc-NMA calculations are colored in blue (TCR) or magenta (pHLA) if they are flexible in the NMA calculations or green if they are frozen (and therefore make up part of the buffer region). Residues colored in white are not included in the calculation (see **Methods**). The 1000 water molecules retained in the calculation are shown as transparent spheres.

The second method evaluated in this manuscript is known as the interaction entropy[58] (Int-Entropy) approach, which determines the solute contributions to $-T\Delta S$ from the fluctuations in the change in the gas phase interaction energy (*i.e.,* larger average fluctuations result in a larger value of $-T\Delta S$, see **Methods** for further details). This approach has the advantage of not requiring additional simulations (as fluctuations of the gas phase interaction energy can be taken directly from the original

MMPB/GBSA calculations) and has shown great promise as a correction for protein-ligand binding free energy calculations.[40,58–60]

For both test sets, there is a clear reduction in the quality of prediction when the Int-Entropy corrections are applied (**Fig 6**). Analysis of individual scatter plots with and without this approach included (**Fig 6B+D**) illustrate that the Int-Entropy approach has a negative effect on the prediction accuracy. We note that the error bars plotted for calculations with the Int-Entropy approach do not include an estimate of the uncertainty of the Int-Entropy correction itself (as all frames are combined for a single estimate). Nevertheless, it is clear that the noise and/or error induced from the Int-Entropy approach has had an unfavorable impact on the prediction accuracy.

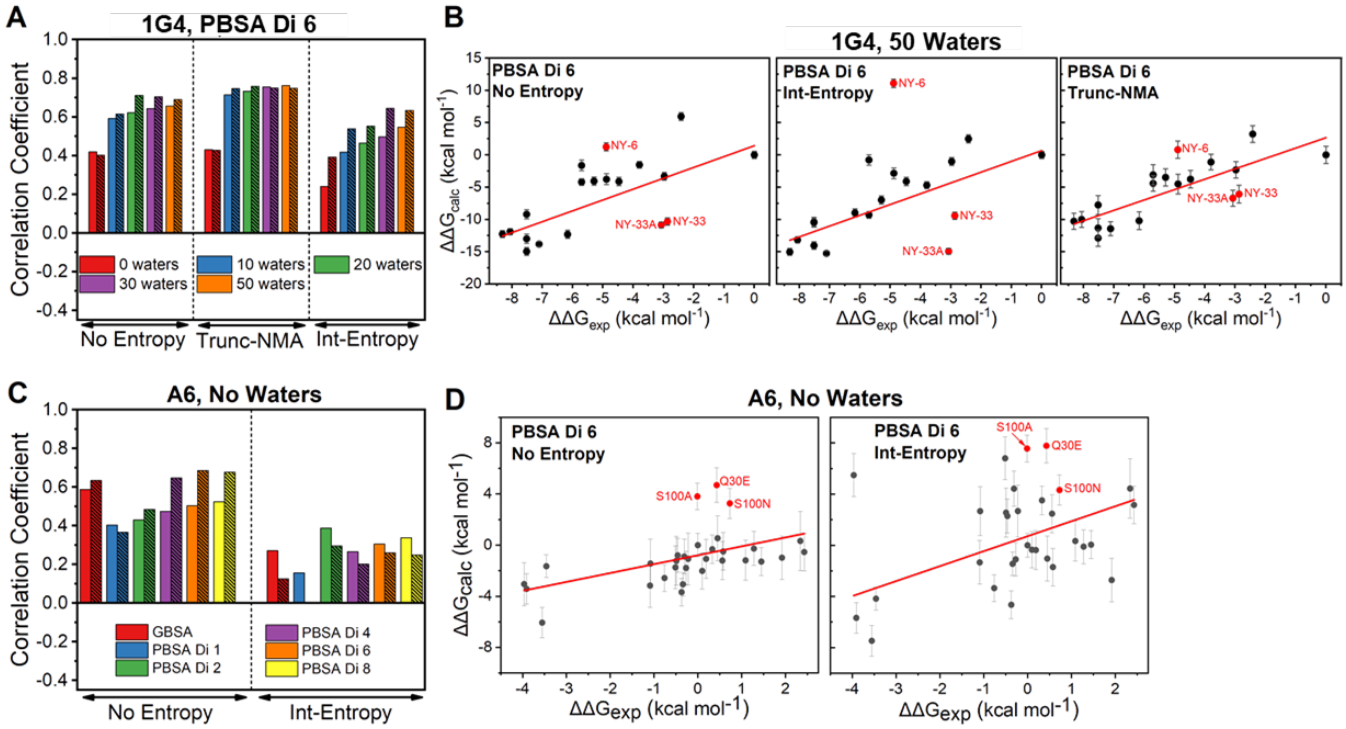


Fig 6: Impact of Solute Entropy Corrections on our MMPB/GBSA Calculations. A. Spearman's rank (r_s, unhashed bars) and Pearson's R (r_P, hashed bars) values determined for MMPB/GBSA calculations on the 1G4 test set with ϵ_{int} set to 6. Results are presented using a variable number of waters without any entropy corrections included as well as with the Trunc-NMA and Int-Entropy approaches. **B**. Exemplar scatter plots for the 1G4 test set with the PBSA approach (with ϵ_{int} set to 6) including 50 explicit water molecules. Panels compare no entropy corrections (left), with Int-Entropy corrections (middle)

and with Trunc-NMA corrections (right). **C**. Impact of the inclusion of the Int-Entropy correction to the A6 data set, with the r_s and r_p values colored as in **A**. All results are without any explicit water molecules included. **D**. Exemplar scatter plots for the A6 test set with the PBSA approach (with ϵ_{int} set to 6) and no explicit water molecules. Panels compare no entropy corrections (left), to with Int-Entropy corrections (right). More complete results, including comparing the effect of removing outliers, are provided in **S9 Fig**.

Whilst the Int-Entropy has been successfully applied to several small-molecule MMPB/GBSA studies, its application to PPIs has proven more challenging.[61] This is largely a consequence of the large binding interfaces (TCR-pHLA buried surface areas tend to be ~2000-2500 Å^2) which give rise to a correspondingly large amount of variance in the obtained per-frame interaction energies. Thus, without exhaustive sampling, this approach can lead to non-converged and abnormally high entropy corrections.[61] One solution to this problem is to perform MMPB/GBSA calculations using an ϵ_{int} value larger than the default of 1, which notably reduces the variance of the interaction energies obtained, ultimately leading to converged entropy estimates within reasonable simulation times.[16] We indeed observed this behavior with our Int-Entropy corrections for the different MMPBSA methods used in this study, in which only ϵ_{int} values between 2 and 8 showed gaussian-like distributions of the gas phase interaction energy (**S8 Fig**). Regardless, the error/noise associated with the calculation was observed to worsen the prediction accuracy for both test sets. We note that when the Int-Entropy method was first introduced by Duan *et al.,* interaction energies were computed using 100,000 snapshots from a single 2 ns long simulation.[58] In contrast, here we extracted significantly fewer snapshots (7500 frames taken from 25, 3 ns long replicas) and our snapshots were significantly less correlated with one another (frames were taken every 0.02 ps by Duan et al.[58] instead of every 10 ps here). Whilst some more recent attempts have successfully applied the Int-Entropy approach using notably fewer simulation frames than those used in the original study,[40,60,61] collecting a much larger number of frames to

assist with convergence would be significantly more resource intensive, both in terms of the additional MMPB/GBSA calculations needed and the additional storage requirements for the simulations.

We performed Trunc-NMA calculations on only the 1G4 test set (**Fig 6A+B** and **S9 Fig**) and obtained no notable change in the prediction accuracy when applying the method to MMPB/GBSA calculations without explicit water molecules. However, the combination of the explicit waters and Trunc-NMA corrections gave rise to a better prediction quality both when including and excluding the aforementioned three outliers. We further note that the improved prediction accuracy associated with Trunc-NMA corrections is not sensitive to the number of explicit water molecules included in the MMPBSA calculation (**Fig 6A**; similar as observed without applying entropy corrections, **Fig 4**).

Whilst we did not evaluate the non-truncated form of NMA, previous studies have clearly shown the beneficial effects of using a truncated system on both the errors obtained and computational efficiency.[55,56] Given the size of a standard TCR-pHLA complex (~800-900 residues), the Trunc-NMA approach used here would be significantly more efficient than standard NMA. For the 1G4 data set composed of many multi-point mutations, the combination of Trunc-NMA and explicit water molecules was beneficial, and we observed the prediction quality to be highly insensitive to the number of explicit water molecules included in the MMPBSA calculation.

How many replicas are required for reproducible MMPB/GBSA calculations?

The results presented so far have shown a clear benefit of the use of a ϵ_{int} value ≥ 4 for MMPBSA calculations, both in terms of improving the prediction quality and in reducing the magnitude of the errors obtained. Further, beneficial effects were also observed for the 1G4 test set when both explicit water molecules and entropy corrections were applied. However, these methods are likely to increase the noise associated with the predictions. It is therefore important to assess how many replicas may be required for reproducible results with the different approaches performed in this study. We used

'bootstrapping' to do this: a statistical method that involves "resampling with replacement", meaning that from a set of N observables (in our case N is the 25 replicas performed for each complex) a large number of bootstrap "resamples" are constructed by randomly removing or duplicating the individual observations. These bootstrap resamples are then used to recalculate the correlation coefficients many times in order to obtain confidence intervals in the calculated correlation coefficients. For both test sets, we generated 1,000,000 bootstrap resamples of $\Delta\Delta G_{calc}$ for several different MMPB/GBSA protocols used here. We then evaluated the impact of using a reduced number of replicas on both Spearman's rank (**Fig 7**) and Pearson's R (**S10 Fig**) and their corresponding confidence intervals. We note that both types of correlation coefficient showed very similar behavior, so only Spearman's rank (**Fig 7**) is discussed below.

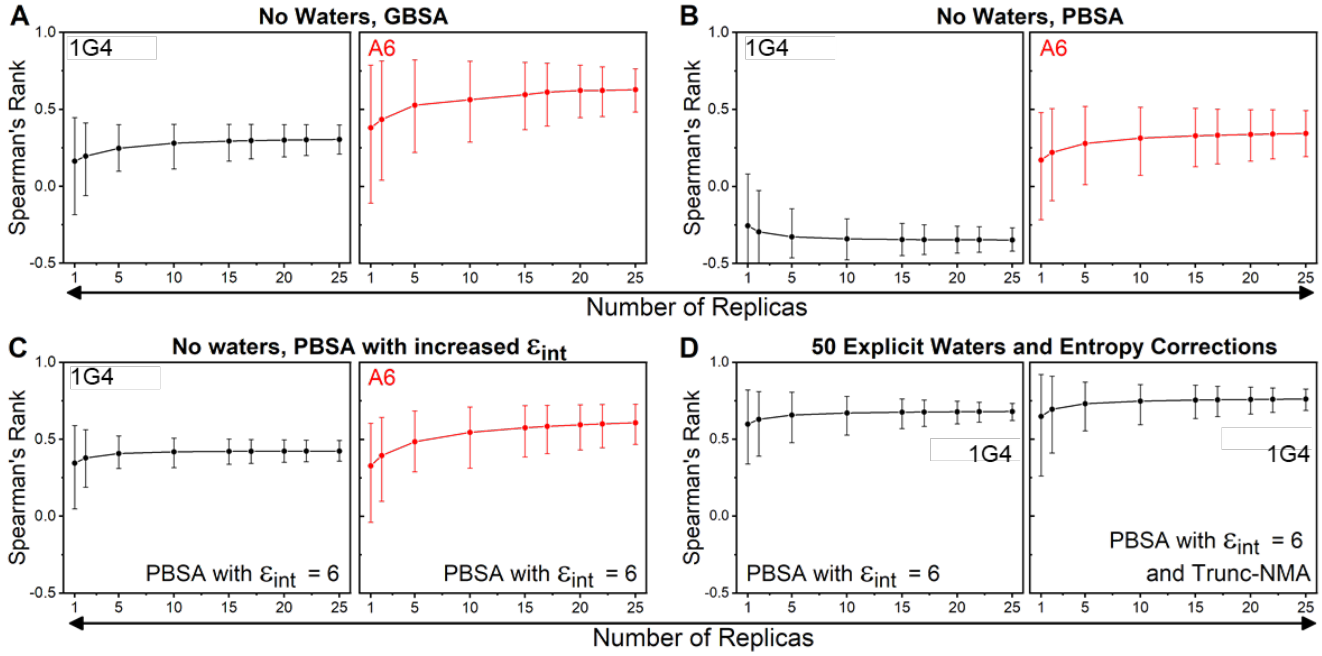


Fig 7: Bootstrapping to assess the impact of using differing numbers of replicas to obtain Spearman's Rank for some of the protocols evaluated in this study. Panels **A+B** focus on the GBSA and PBSA approaches with no explicit waters included. Panel **C** focusses on the PBSA method with ϵ_{int} set to 6. Panel **D** focusses on the PBSA method (ϵ_{int} set to 6) with 50 explicit waters molecules included with and without the Trunc-NMA correction applied. Measurements with the 1G4 and A6 test sets are colored black and red respectively. In each panel, the average of the 1 million bootstrap resamples are used to calculate Spearman's Rank when using a differing number of replicas, with the error bars depicting the 95% confidence

intervals. The complete data is used in all cases (*i.e.,* the outliers discussed above are included). Equivalent results with the Pearson's R metric are provided in **S10 Fig.**

Focusing first on the 1G4 test set (**Fig 7A-D**), there appears to be little benefit for performing more than 15 replicas for the MMGBSA approach, whilst for the MMPBSA simulations with ϵ_{int} set to 6, one could argue that as few as 5 replicas may be sufficient, considering the additional computing cost if more replicas are used. This is also true when explicit water molecules are included and/or Trunc-NMA entropy corrections applied: 5-10 replicas are sufficient to converge the prediction estimates. Comparison of the A6 and 1G4 test sets shows that the A6 test set is generally noisier for each comparable method (**Fig 7A-C**). This is likely in part due to the reduced experimental affinity range in the data set as well as the comparably lower quality of the WT crystal structure (resolutions of 1.9 Å vs 2.6 Å for 1G4[62] and A6[63], respectively). For the A6 test set, a larger number of replicas may therefore be optimal as compared to the 1G4 TCR, in terms of the balance between accuracy and computing cost. Regardless, for both test sets a maximum of 15 replicas would appear to be sufficient when using the optimal parameters previously described.

Conclusions

This study evaluated MMPB/GBSA protocols for relative binding free energy prediction on two different sets of TCR-pHLA variants. The first test set (1G4) consists of 18 multi-point variants which contained between 3-14 mutations across up to 5 CDR loops, and spanned a large range of experimental affinities (~8 kcal mol^{-1}).[26,31] In contrast, the second (A6) set of TCR variants consisted largely of single point mutations (25 of 29, with the four remaining variants having between 2-4 mutations), all located on the CDR3β loop.[29,32] The A6 variants in this set include those that both increase and decrease binding affinity relative to the WT ($\Delta\Delta G_{exp}$ values between +2.3 and −4.0 kcal mol^{-1}). The

contrast between test sets was intentional, as we initially aimed to see if an accurate MMPB/GBSA protocol could be obtained that would be applicable for any set of TCR-pHLA variants. Although there is no single protocol that is highly suitable for both these sets, there are general lessons to be learned and specific recommendations for the application of MMPB/GBSA to TCR-pHLA complexes that can be made based on our results.

First, for MMPBSA calculations of TCR-pHLA complexes, and indeed protein-protein complexes in general, an increased value (between 4-8) of ϵ_{int} is strongly recommended. This should not only improve the prediction quality but also reduce the uncertainties obtained, meaning fewer simulations need be performed per complex (arguably ~3-fold less; see **Fig 7**). Further, the use of a such an increased dielectric constant is likely to be particularly important for describing charge swapping mutations (**Fig 2**). Second, when deciding whether to include explicit water molecules or solute entropy corrections, we observed a clear divergence in the optimal protocol between our two test sets. For our data set composed of many multi-point mutants, inclusion of explicit water molecules (10 to 50) improved the prediction accuracy. In contrast, for our test set (A6) composed of single point (or very few) mutations, the additional errors/noise from adding explicit solvent resulted in a reduced prediction accuracy. Third, Int-Entropy corrections had a negative effect on the prediction quality for both test sets (likely in part due to the relatively small number of snapshots used). However, using the Trunc-NMA approach improved the prediction accuracy for the test set where variants had significantly altered H-bonding across the interface, essentially resolving what were significant outliers. Notably, analysis of the total average number of hydrogen bonding contacts for different variants enabled identification of the main outliers in the data set, indicating that this could be used to "triage" potentially poorly described mutations for further investigation using Trunc-NMA calculations.

Overall, we thus recommend the following for TCR-pHLA relative binding affinity prediction with MMPBSA: (1) Use an internal dielectric constant of ~6; (2) for TCR variants with multiple mutations,

include sufficient explicit water molecules to cover the interface. (3) Only when variants lead to a significant change in the TCR-pHLA Hydrogen bonding network, a truncated NMA based entropy correction should be applied. When computational efficiency is important, MMGBSA could be considered for TCR variants with few mutations.

Finally, our bootstrapping analysis demonstrated that when using MMPBSA with an increased internal dielectric constant (4-8, as recommended above) as few as 5 replicas (20 ns MD in total) can be sufficient to obtain reproducible results. The inclusion of explicit water molecules or Trunc-NMA entropy corrections did not notably increase the number of replicas one would need to perform. Thus, in a practical context, one could envisage a scenario in which all candidate mutants are first subjected to an initial screen of 5 replicas, with promising variants taken forward for a total of 10-15 replicas for increased accuracy.

Computational methods that allow for the accurate ranking of TCR-pHLA binding affinities and those of PPIs more generally have obvious utility in computational drug discovery. Whilst we intended to find a general approach, our results demonstrated the need for two somewhat different approaches for the accurate and reliable ranking of TCR-pHLA binding affinities, one for ranking TCR variants with multiple mutations (>4) from the starting point, and one with few mutations. We believe the MMPB/GBSA approach outlined here has promise as a medium throughput screening tool to select and rank candidate mutations for experimental testing.

Methods

Structure preparation.

X-ray crystal structures of the TCR-pHLA complexes of wild-type (WT) 1G4 and WT A6 were taken from PDBs 2BNR[62] and 1AO7[63] respectively, with the missing residues in PDB 1AO7 (located in the constant domain, away from the binding site) added in using PDB 4FTV,[32] which has an identical (but resolved) constant domain to 1AO7. All simulations of point variants were performed using the WT structure, with mutations inserted using PyMOL[64] (rotamers were selected based on recommendations from PyMOL v2.1, avoiding clashes as much as possible). Optimal His tautomerisation states and Asn and Gln side chain orientations were determined using MolProbity[65], with all residues simulated in their standard protonation states at pH 7 (consistent with PROPKA 3.0[66] predictions). His tautomerisation states were kept consistent between the WT and any variant structure simulated (see **S1 Table** for tautomerisation states used). All structures were solvated in an octahedral water box such that no protein atom was within 10 Å of the box boundary, with the minimum number of either Na^+ or Cl^- ions added as required to ensure total system neutrality. The 1G4 TCRs were solvated retaining the WT crystal structure waters, with any crystal water molecule that clashed with a newly inserted side-chain removed. For the A6 TCRs, the resolution of the WT structure (2.6 Å) is too low to identify (many) waters surrounding the binding site, so 3D-RISM[47,48] was used to calculate the radial distribution function (g(r)) of water surrounding the protein and the "Placevent" algorithm[49] was used to solvate the protein based on the obtained g(r) (see **Supplementary Methods**), prior to solvation in a octahedral box.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed using GPU accelerated Amber16, [67] with the ff14SB[68] force field and TIP3P water model were used to describe the protein and water molecules respectively. For each structure, a protocol of minimization, heating and equilibration (see **Supplementary Methods**) prior to production MD simulations in the NPT ensemble (298 K and 1 atm). For each structure, 25 replicas of 4 ns long were performed, with the

last 3 ns taken forward for MMPB/GBSA calculations. Simulations were performed with a 2 fs time step (with the SHAKE algorithm applied to all bonds containing hydrogen. The default 8 Å direct space non-bonded cut-off was applied with long range electrostatics evaluated using the particle mesh Ewald algorithm. Temperature and pressure regulation were performed using Langevin temperature control (collision frequency of 1 ps^{-1}) and a Berendsen barostat (pressure relaxation time of 1 ps). Trajectory analysis was performed using CPPTRAJ.[69] Hydrogen bonds (both solute-solute and water bridged) were considered formed if the donor-acceptor distances were less than 3 Å and the donor-hydrogen-acceptor angles were between 180±45°.

MMPB/GBSA theory and methodology.

The molecular mechanics generalized Born/Poisson–Boltzmann surface area (MMPB/GBSA) is an end state binding free energy calculation method which calculates the binding free energy (ΔG_{bind}) through the following equation:

$$\Delta G_{bind} = \Delta E_{int} + \Delta G_{pol} + \Delta G_{npol} - T\Delta S \quad \textbf{(1)}$$

Where ΔE_{int} is the difference in the interaction energy, ΔG_{pol} and ΔG_{npol} are the polar and non-polar contributions to the solvation free energy respectively, and ΔS is the change in solute entropy. ΔE_{int} can be obtained directly from the force field energy terms:

$$\Delta E_{int} = \Delta E_{internal} + \Delta E_{ele} + \Delta E_{vdW} \quad \textbf{(2)}$$

Where $\Delta E_{internal}$ is the difference in the internal energy terms (i.e., bonding, angle, dihedral and improper torsions) and ΔE_{ele} and ΔE_{vdW} are the electrostatic and Van der Waals (vdW) contributions. Note that in the single trajectory approach, which is used here, the contributions from $\Delta E_{internal}$ cancel out. ΔG_{pol} is obtained by solving either the Poisson Boltzmann (PB) or Generalized Born (GB) equations respectively. The non-polar contributions to the solvation free energy can be estimated from the solvent accessible surface area (SASA):

$$\Delta G_{npol} = \gamma \times SASA + b \quad \textbf{(3)}$$

Where γ is surface tension (set to 0.00542 kcal mol^{-1} Å^{-2}) and b is an offset (set to 0.92 kcal mol^{-1}).

Finally, $T\Delta S$ is an optional correction that accounts for the change in solute entropy. In this study we tested two different methods in order to calculate this, which are discussed in the section "Solute Entropy Corrections".

For MMPB/GBSA calculations, frames were taken every 10 ps from the last 3 ns of each production MD simulation replica, meaning a total of 300 x 25 (number of replicas performed) frames were used for MMPB/GBSA calculations. Calculations were performed using the MPI version of MMPBSA.py[5], with the GB-Neck2[41] (*i.e.,* igb=8) solvent model for GBSA calculations, and the default PB solvent model for MMPBSA calculations. MMPB/GBSA calculations were performed with an implicit salt concertation of 150 mM (to match experimental assay conditions).

Solute Entropy Corrections.

As discussed in the introduction, the MMPB/GBSA approach does not account for the rigidification of the solutes upon binding. We applied two different methods to predict a "correction" for this effect to the calculated binding free energies, using both the interaction entropy (Int-Entropy[58]) and the truncated normal mode analysis[55] (Trunc-NMA) methods.

The Int-Entropy approach developed by Duan et al.[58] uses the fluctuation of the gas phase contributions to ΔG_{bind} (referred to as the interaction energy, ΔE_{int}) to provide an estimate of $T\Delta S$. **Equation 2** shows how to calculate ΔE_{int}. The per frame fluctuation of ΔE_{int} can then be determined by:

$$\Delta E_{fluc} = \Delta E_{int} - \langle \Delta E_{int} \rangle \quad \textbf{(4)}$$

Where $\langle \Delta E_{int} \rangle$ is the ensemble average of ΔE_{int}. Finally, $T\Delta S$ can be determined by:

$$-T\Delta S = k_B T \ln\langle e^{\beta \Delta E_{fluc}} \rangle \quad \textbf{(5)}$$

Where β is $1/k_B T$. For each different MMPBSA or MMGBSA calculation, we took the ΔE_{int} values obtained from all 7500 frames per complex and used this to calculate $-T\Delta S$.

Normal mode analysis (NMA) uses vibrational frequency calculations of energy minimized structures of each state to determine the change in solute rigidity upon ligand binding and can therefore be used to determine $T\Delta S$ in **Equation 1**. To reduce the computational cost and noise associated with this approach, we used a modified version of this approach referred to as truncated-NMA (Trunc-NMA).[55] In Trunc-NMA, only a subset of atoms located near the binding site are used for the entropy calculation. Residues located close to the binding site are treated as flexible (*i.e.,* allowed to move and therefore contribute to a vibrational frequency calculation), whilst residues further away from the binding site are included in a "buffer zone" and held fixed throughout the minimization and vibrational frequency calculations. Further, water molecules that surround the binding site are also often included as part of the buffer region. In the Trunc-NMA approach used here (see **Supplementary Methods** for further details), we retained all receptor (pHLA) residues within ~16 Å of any ligand (TCR) residue and vice versa, using the WT crystal structure to determine distances. Any breakages introduced into the sequence were acetylated or amidated, using the coordinates from the first deleted residue. Those residues within the range 12-16 Å were kept frozen for both the optimization and vibrational frequency calculations. A shell of 1000 water molecules were also retained (and kept frozen throughout) around the binding site. For the frequency calculations of the free ligand or receptor, 500 water molecules were included for each structure. Energy minimization was performed using sander (Ambertools18[67]) with GB implicit solvent and performed until the RMSD was less than 10^{-6} kcal mol^{-1} Å^{-1}. Frequency calculations were performed *in vacuo* using a modified version of the Nmode program (from Amber14), to allow use of the "ibelly" command, which allows for the freezing of atoms during the energy

minimization and vibrational frequency calculations. Frozen atoms therefore have no (direct) impact on the entropy estimates obtained. The Trunc-NMA approach was only applied to the 1G4 set of TCRs and was performed on frames taken every 100 ps from the last 3 ns of each of the 25 replicas (750 frames per complex).

Assessment of the Quality of Prediction.

Experimentally determined $\Delta\Delta G$'s (obtained from prior studies[26,29,31,32], see **S2** and **S3 Tables** for affinities of all TCR-pHLA complexes studied) were compared to the computationally derived $\Delta\Delta G$'s and assessed using both the Pearson's r (r_p) value and Spearman's rank (r_s). These metrics were chosen as r_p determines how linearly correlated the two data sets are, whilst r_s assesses how monotonic the two data sets are (*i.e.,* how well do the computational results correctly rank order the experimental results). Error values associated with individual $\Delta\Delta G_{bind}$ calculations are the standard deviation obtained from the 25 replicas performed per complex. Bootstrapping with random replacement was performed using the R software package. In all instances, 1 million bootstrap resamples were constructed from the original 25 replicas performed per complex. Each resample was then used to calculate Spearman's rank and Pearson correlation coefficient *r*, with the average values and 95% confidence intervals determined for different numbers of replicas.

Author contributions

RMC and MWvdK designed the study. RMC performed the simulations and analysis. RMC produced the first draft manuscript which was edited through contributions from all authors. All authors discussed and interpreted data. All authors have given approval to the final version of the manuscript.

Acknowledgments

The authors would like to thank Dr. Dimas Suarez (Univ. of Oviedo, Spain) for assistance in modifying the Nmode program to enable the running of Trunc-NMA calculations.

Abbreviations

Complementarity-determining region (CDR), interaction entropy (Int-Entropy), internal dielectric constant (ϵ_{int}), molecular dynamics (MD), Molecular Mechanics Poisson-Boltzmann/generalized Born surface area (MMPB/GBSA), peptide-human leukocyte antigen (pHLA), polar buried area (PBA), protein-protein interaction (PPI), radial distribution function (g(r)), T-cell receptor (TCR), truncated normal mode analysis (Trunc-NMA).

Funding

RMC's PhD studentship was funded by a Engineering and Physical Sciences Research Council (EPSRC) Training Grant (EP/L016354/1). MWvdK is a BBSRC David Phillips Fellow (BB/M026280/1). This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath, as well as the computational facilities of the Advanced Computing Research Centre of the University of Bristol. Further, this project used computing time on ARCHER, granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/L000253/1).

Data availability

All starting structures for simulation (Amber topology files and coordinates), example input files for MD simulations and MMPB/GBSA calculations, example analysis scripts and a copy of the modified NMode code used are available at DOI: 10.5281/zenodo.4805388.

All further relevant data are within the manuscript and its Supporting Information file.

References

1. Clark AJ, Gindin T, Zhang B, Wang L, Abel R, Murret CS, et al. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. J Mol Biol. 429(7):930–47. DOI: 10.1016/j.jmb.2016.11.021

2. Keskin O, Tuncbag N, Gursoy A. Predicting Protein–Protein Interactions from the Molecular to the Proteome Level. Chem Rev. 116(8):4884–909. DOI: 10.1021/acs.chemrev.5b00683

3. Siebenmorgen T, Zacharias M. Evaluation of Predicted Protein–Protein Complexes by Binding Free Energy Simulations. J Chem Theory Comput. 15(3):2071–86. DOI: 10.1021/acs.jctc.8b01022

4. Ramadoss V, Dehez F, Chipot C. AlaScan: A Graphical User Interface for Alanine Scanning Free-Energy Calculations. J Chem Inf Model. 56(6):1122–6. DOI: 10.1021/acs.jcim.6b00162

5. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 8(9):3314–21. DOI: 10.1021/ct300418h

6. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 10(5):449–61. DOI: 10.1517/17460441.2015.1032936

7. Holland CJ, Crean RM, Pentier JM, de Wet B, Lloyd A, Srikannathasan V, et al. Specificity of bispecific T cell receptors and antibodies targeting peptide-HLA. J Clin Invest. 130(5):2673–88. DOI: 10.1172/JCI130562

8. Zoete V, Irving MB, Michielin O. MM-GBSA binding free energy decomposition and T cell receptor engineering. J Mol Recognit. 23(2):142–52. DOI: 10.1002/jmr.1005

9. Crean RM, MacLachlan BJ, Madura F, Whalley T, Rizkallah PJ, Holland CJ, et al. Molecular Rules Underpinning Enhanced Affinity Binding of Human T Cell Receptors Engineered for

Immunotherapy. Mol Ther - Oncolytics. 18(September):443–56. DOI: 10.1016/j.omto.2020.07.008

10. Zoete V, Irving M, Ferber M, Cuendet MA, Michielin O. Structure-based, rational design of T cell receptors. Front Immunol. 4(SEP):1–19. DOI: 10.3389/fimmu.2013.00268

11. Maffucci I, Contini A. Improved Computation of Protein–Protein Relative Binding Energies with the Nwat-MMGBSA Method. J Chem Inf Model. 56(9):1692–704. DOI: 10.1021/acs.jcim.6b00196

12. Wang C, Greene D, Xiao L, Qi R, Luo R. Recent Developments and Applications of the MMPBSA Method. Front Mol Biosci. 4(January):1–18. DOI: 10.3389/fmolb.2017.00087

13. Sun H, Li Y, Tian S, Xu L, Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. Phys Chem Chem Phys. 16(31):16719–29. DOI: 10.1039/C4CP01388C

14. Zhu Y-L, Beroza P, Artis DR. Including Explicit Water Molecules as Part of the Protein Structure in MM/PBSA Calculations. J Chem Inf Model. 54(2):462–9. DOI: 10.1021/ci4001794

15. Godschalk F, Genheden S, Söderhjelm P, Ryde U. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. Phys Chem Chem Phys. 15(20):7731. DOI: 10.1039/c3cp00116d

16. Liu X, Peng L, Zhang JZH. Accurate and Efficient Calculation of Protein–Protein Binding Free Energy-Interaction Entropy with Residue Type-Specific Dielectric Constants. J Chem Inf Model. 59(1):272–81. DOI: 10.1021/acs.jcim.8b00248

17. Goebeler M-E, Bargou RC. T cell-engaging therapies — BiTEs and beyond. Nat Rev Clin Oncol. 17(7):418–34. DOI: 10.1038/s41571-020-0347-5

18. Hewitt EW. The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology. 110(2):163–9. DOI: 10.1046/j.1365-2567.2003.01738.x

19. Oates J, Jakobsen BK. ImmTACs. Oncoimmunology. 2(2):e22891. DOI: 10.4161/onci.22891

20. Singh NK, Riley TP, Baker SCB, Borrman T, Weng Z, Baker BM. Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes. J Immunol. 199(7):2203–13. DOI: 10.4049/jimmunol.1700744

21. Aleksic M, Liddy N, Molloy PE, Pumphrey N, Vuidepot A, Chang K-M, et al. Different affinity windows for virus and cancer-specific T-cell receptors: Implications for therapeutic strategies. Eur J Immunol. 42(12):3174–9. DOI: 10.1002/eji.201242606

22. Richman SA, Healan SJ, Weber KS, Donermeyer DL, Dossett ML, Greenberg PD, et al. Development of a novel strategy for engineering high-affinity proteins by yeast display. Protein Eng Des Sel. 19(6):255–64. DOI: 10.1093/protein/gzl008

23. Harris DT, Wang N, Riley TP, Anderson SD, Singh NK, Procko E, et al. Deep Mutational Scans as a Guide to Engineering High Affinity T Cell Receptor Interactions with Peptide-bound Major Histocompatibility Complex. J Biol Chem. 291(47):24566–78. DOI: 10.1074/jbc.M116.748681

24. Chervin AS, Aggen DH, Raseman JM, Kranz DM. Engineering higher affinity T cell receptors using a T cell display system. J Immunol Methods. 339(2):175–84. DOI: 10.1016/j.jim.2008.09.016

25. Sharma P, Kranz DM. Subtle changes at the variable domain interface of the T-cell receptor can strongly increase affinity. J Biol Chem. 293(5):1820–34. DOI: 10.1074/jbc.M117.814152

26. Li Y, Moysey R, Molloy PE, Vuidepot AL, Mahon T, Baston E, et al. Directed evolution of human T-cell receptors with picomolar affinities by phage display. Nat Biotechnol. 23(3):349–54. DOI: 10.1038/nbt1070

27. Madura F, Rizkallah PJ, Miles KM, Holland CJ, Bulek AM, Fuller A, et al. T-cell Receptor Specificity Maintained by Altered Thermodynamics. J Biol Chem. 288(26):18766–75. DOI: 10.1074/jbc.M113.464560

28. Pierce BG, Hellman LM, Hossain M, Singh NK, Vander Kooi CW, Weng Z, et al. Computational design of the affinity and specificity of a therapeutic T cell receptor. Dunbrack RL, editor. PLoS Comput Biol. 10(2):e1003478. DOI: 10.1371/journal.pcbi.1003478

29. Haidar JN, Pierce B, Yu Y, Tong W, Li M, Weng Z. Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC. Proteins Struct Funct Bioinforma. 74(4):948–60. DOI: 10.1002/prot.22203

30. 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. 27(2):300–13. DOI: 10.1016/j.ymthe.2018.12.010

31. Dunn SM, Rizkallah PJ, Baston E, Mahon T, Cameron B, Moysey R, et al. Directed evolution of human T cell receptor CDR2 residues by phage display dramatically enhances affinity for cognate peptide-MHC without increasing apparent cross-reactivity. Protein Sci. 15(4):710–21. DOI: 10.1110/ps.051936406

32. Cole DK, Sami M, Scott DR, Rizkallah PJ, Borbulevych OY, Todorov PT, et al. Increased peptide contacts govern high affinity binding of a modified TCR whilst maintaining a native pMHC docking mode. Front Immunol. 2013/06/28. 4(JUN):1–8. DOI: 10.3389/fimmu.2013.00168

33. Wan S, Knapp B, Wright DW, Deane CM, Coveney P V. Rapid, Precise, and Reproducible Prediction of Peptide-MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. J Chem Theory Comput. 11(7):3346–56. DOI: 10.1021/acs.jctc.5b00179

34. Wright DW, Hall BA, Kenway OA, Jha S, Coveney P V. Computing clinically relevant binding free energies of HIV-1 protease inhibitors. J Chem Theory Comput. 10(3):1228–41. DOI: 10.1021/ct4007037

35. Genheden S, Ryde U. How to obtain statistically converged MM/GBSA results. J Comput Chem. 32(May):NA-NA. DOI: 10.1002/jcc.21366

36. Wan S, Bhati AP, Zasada SJ, Wall I, Green D, Bamborough P, et al. Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: A Computational Study. J Chem Theory Comput. 13(2):784–95. DOI: 10.1021/acs.jctc.6b00794

37. Chen F, Liu H, Sun H, Pan P, Li Y, Li D, et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. Phys Chem Chem Phys. 18(32):22129–39. DOI: 10.1039/C6CP03670H

38. Sun H, Li Y, Shen M, Tian S, Xu L, Pan P, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. Phys Chem Chem Phys. 16(40):22035–45. DOI: 10.1039/C4CP03179B

39. Wang C, Nguyen PH, Pham K, Huynh D, Le TBN, Wang H, et al. Calculating protein–ligand binding affinities with MMPBSA: Method and error analysis. J Comput Chem. :2436–46. DOI: 10.1002/jcc.24467

40. Liu X, Peng L, Zhou Y, Zhang Y, Zhang JZH. Computational Alanine Scanning with Interaction Entropy for Protein–Ligand Binding Free Energies. J Chem Theory Comput. 14(3):1772–80. DOI: 10.1021/acs.jctc.7b01295

41. Nguyen H, Roe DR, Simmerling C. Improved generalized born solvent model parameters for

protein simulations. J Chem Theory Comput. 9(4):2020–34. DOI: 10.1021/ct3010485

42. Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. Bioinformatics. 27(20):2915–6. DOI: 10.1093/bioinformatics/btr484

43. Ramos RM, Moreira IS. Computational Alanine Scanning Mutagenesis—An Improved Methodological Approach for Protein–DNA Complexes. J Chem Theory Comput. 9(9):4243–56. DOI: 10.1021/ct400387r

44. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis—An improved methodological approach. J Comput Chem. 28(3):644–54. DOI: 10.1002/jcc.20566

45. Martins SA, Perez MAS, Moreira IS, Sousa SF, Ramos MJ, Fernandes PA. Computational alanine scanning mutagenesis: MM-PBSA vs TI. J Chem Theory Comput. 9(3):1311–9. DOI: 10.1021/ct4000372

46. Simões ICM, Costa IPD, Coimbra JTS, Ramos MJ, Fernandes PA. New Parameters for Higher Accuracy in the Computation of Binding Free Energy Differences upon Alanine Scanning Mutagenesis on Protein-Protein Interfaces. J Chem Inf Model. 57(1):60–72. DOI: 10.1021/acs.jcim.6b00378

47. Beglov D, Roux B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. J Phys Chem B. 101(39):7821–6. DOI: 10.1021/jp971083h

48. Kovalenko A, Hirata F. Potential of mean force between two molecular ions in a polar molecular solvent: A study by the three-dimensional reference interaction site model. J Phys Chem B. 103(37):7942–57. DOI: 10.1021/jp991300+

49. Sindhikara DJ, Yoshida N, Hirata F. Placevent: An algorithm for prediction of explicit solvent atom distribution-Application to HIV-1 protease and F-ATP synthase. J Comput Chem.

33(18):1536–43. DOI: 10.1002/jcc.22984

50. Wright DW, Wan S, Meyer C, van Vlijmen H, Tresadern G, Coveney P V. Application of ESMACS binding free energy protocols to diverse datasets: Bromodomain-containing protein 4. Sci Rep. 9(1):6017. DOI: 10.1038/s41598-019-41758-1

51. Mikulskis P, Genheden S, Ryde U. Effect of explicit water molecules on ligand-binding affinities calculated with the MM/GBSA approach. J Mol Model. 20(6). DOI: 10.1007/s00894-014-2273-x

52. Maffucci I, Hu X, Fumagalli V, Contini A. An efficient implementation of the Nwat-MMGBSA method to rescore docking results in medium-throughput virtual screenings. Front Chem. 6(MAR):1–14. DOI: 10.3389/fchem.2018.00043

53. Peccati F, Jiménez-Osés G. Enthalpy–Entropy Compensation in Biomolecular Recognition: A Computational Perspective. ACS Omega. 6(17):11122–30. DOI: 10.1021/acsomega.1c00485

54. Sun H, Duan L, Chen F, Liu H, Wang Z, Pan P, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. Phys Chem Chem Phys. 20(21):14450–60. DOI: 10.1039/C7CP07623A

55. Kongsted J, Ryde U. An improved method to predict the entropy term with the MM/PBSA approach. J Comput Aided Mol Des. 23(2):63–71. DOI: 10.1007/s10822-008-9238-z

56. Genheden S, Kuhn O, Mikulskis P, Hoffmann D, Ryde U. The normal-mode entropy in the MM/GBSA method: Effect of system truncation, buffer region, and dielectric constant. J Chem Inf Model. 52(8):2079–88. DOI: 10.1021/ci3001919

57. Suárez D, Díaz N. Ligand Strain and Entropic Effects on the Binding of Macrocyclic and Linear Inhibitors: Molecular Modeling of Penicillopepsin Complexes. J Chem Inf Model. 57(8):2045–55. DOI: 10.1021/acs.jcim.7b00355

58. Duan L, Liu X, Zhang JZH. Interaction entropy: A new paradigm for highly efficient and reliable

computation of protein-ligand binding free energy. J Am Chem Soc. 138(17):5722–8. DOI: 10.1021/jacs.6b02682

59. Yan Y, Yang M, Ji CG, Zhang JZH. Interaction Entropy for Computational Alanine Scanning. J Chem Inf Model. 57(5):1112–22. DOI: 10.1021/acs.jcim.6b00734

60. Chen J, Wang X, Zhang JZHH, Zhu T. Effect of Substituents in Different Positions of Aminothiazole Hinge-Binding Scaffolds on Inhibitor-CDK2 Association Probed by Interaction Entropy Method. ACS Omega. 3(12):18052–64. DOI: 10.1021/acsomega.8b02354

61. Sun Z, Yan YN, Yang M, Zhang JZH. Interaction entropy for protein-protein binding. J Chem Phys. 146(12):124124. DOI: 10.1063/1.4978893

62. Chen J-L, Stewart-Jones G, Bossi G, Lissin NM, Wooldridge L, Choi EML, et al. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. J Exp Med. 201(8):1243–55. DOI: 10.1084/jem.20042323

63. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature. 384(6605):134–41. DOI: 10.1038/384134a0

64. Schrödinger. PyMol Molecular Graphics System. Schrödinger, LLC;

65. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr. 66(1):12–21. DOI: 10.1107/S0907444909042073

66. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput. 7(7):2284–95. DOI: 10.1021/ct200133y

67. D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,

A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Mo-nard, H DMY and PAK, Case DA, Cerutti DS, T.E. Cheatham I, Darden TA, Duke RE, et al. Amber 2016. San Francisco: University of California, San Francisco.

68. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 11(8):3696–713. DOI: 10.1021/acs.jctc.5b00255

69. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput. 9(7):3084–95. DOI: 10.1021/ct400341p

Supporting information

Reliable *in silico* ranking of engineered therapeutic TCR binding affinities using MMPBSA and MMGBSA calculations.

Rory M. Crean [1,2,#a,*], Christopher R. Pudney [1,3], David K. Cole[4,5], Marc W. van der Kamp[6,*]

[1]Department of Biology and Biochemistry, [2]Doctoral Training Centre in Sustainable Chemical Technologies, [3]Centre for Therapeutic Innovation, University of Bath, Bath, UK.
[4]Immunocore Ltd., Milton Park, Abingdon, UK.
[5]Division of Infection & Immunity, Cardiff University, Cardiff, UK.
[6]School of Biochemistry, University of Bristol, Biomedical Sciences building, Bristol, UK.
[#a]Current Address: Science for Life Laboratory, Department of Chemistry − Biomedicinska Centrum, Uppsala University, Uppsala, Sweden.

* Corresponding authors

E-mail: rory.crean@kemi.uu.se (RMC).

E-mail: marc.vanderkamp@bristol.ac.uk (MWvdK).

Table of contents

Supplementary methods 1

WT A6-TCR Solvation. 1

Structure equilibration procedure. 1

Truncated Normal Mode Analysis. 2

Inclusion of Explicit Water Molecules for MMPB/GBSA Calculations. 2

Supplementary figures 3

Supplementary tables 13

Supplementary references 17

Supplementary methods

WT A6-TCR Solvation. The 3D-Reference Interaction Site Model[1,2] (3D-RISM) was used to predict the density distribution function ($g(\boldsymbol{r})$) for water oxygen atoms across the entire TCR-pHLA binding site of the WT A6 TCR-pHLA structure. For 3D-RISM calculations (performed with AmberTools18), the Kovalenko-Hirata (KH) closure method[3,4] was used, with all other settings kept as default. Following this, Placevent[5] was used to solvate the entire TCR-pHLA complex, by solvating the entire complex with waters molecules up to 5 Å away from any protein atom. Multiple cut-off $g(\boldsymbol{r})$ values (point at which additional waters with smaller $g(\boldsymbol{r})$ values are no longer added) were tested as the default value of 1.5 Å resulted in stability issues (due to "vacuum bubbles") during NPT simulations (due to too much space between solvent atoms in the initial structure). We found that appropriate solvation was achieved with a $g(\boldsymbol{r})$ cut-off of 1.1 Å and following this we solvated the protein and these water molecules in a water box of size 7 Å (away from both any protein and 3D-RISM/Placevent water molecule). This distance of 7 Å was chosen as it gave a slightly bigger box size (in all dimensions) to the box size that would be generated when solvating just the protein in a 10 Å octahedral water box. (This value is likely to be somewhat system dependent.)

Structure equilibration procedure. The following procedure was used to prepare all systems simulated for production MD simulations in the NPT ensemble at 298 K and 1 atm. All dynamics steps applied the SHAKE algorithm to constrain all bonds containing a hydrogen atom. Replica simulations were initiated from the second heating step of the following protocol (with each replica assigned different random velocity vectors at this stage). Simulations performed in the NVT ensemble used Langevin temperature control (with a collision frequency of 1 ps^{-1}) and used a simulation timestep of 1 fs. Simulations performed in the NPT ensemble used Langevin temperature control (collision frequency of 1 ps^{-1}) and a Berendsen barostat (1 ps pressure relaxation time).

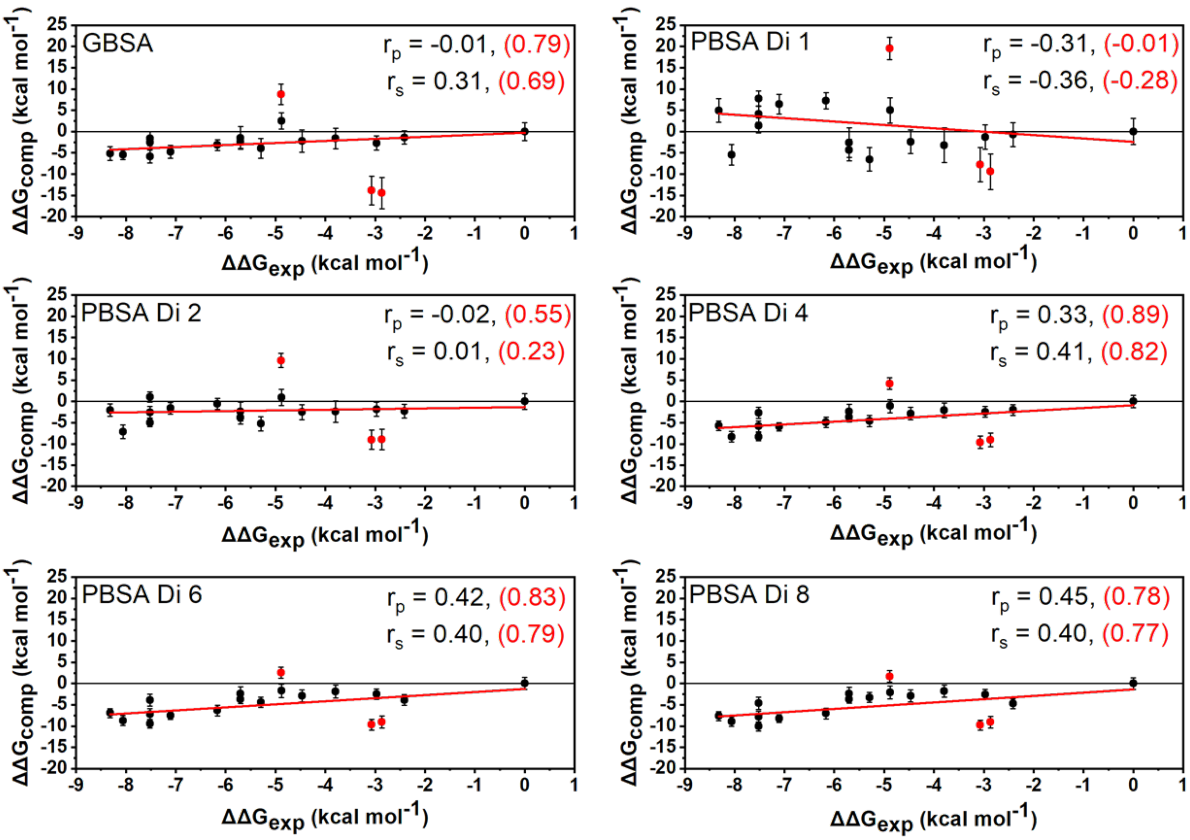
The equilibration protocol is as follows: First, hydrogens atoms and solvent molecules were energy minimized (using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization). To prevent the movement of non-hydrogen and non-solvent atoms during the minimization, 10 kcal mol^{-1} $Å^{-1}$ positional restraints were used to keep all heavy atoms fixed. Then the solvent was heated rapidly from 50 K to 298 K (NVT ensemble, 1 fs timestep) over the course of 200 ps, with the previously described restraints still maintained. The positional restraints were then replaced with 5 kcal mol^{-1} $Å^{-1}$ positional restraints on only the Cα carbon atoms of each residue and subjected to another round of energy minimization (500 steps of steepest descent followed by 500 steps of conjugate gradient). Retaining these

positional restraints, the system was heated from 25 K to 298 K over the course of 50 ps (NVT ensemble, 1 fs time step). Simulations were then performed in the NPT ensemble (1 atm, 298 K, 2 fs time step) by first gradually reducing the 5 kcal mol^{-1} $Å^{-1}$ Cα carbon restraints over the course of 50 ps. This was done by reducing the restraint weight by 1 kcal mol^{-1} $Å^{-1}$ every 10 ps. A final 1 ns long NVT MD simulation with no restraints placed on the system was then performed, with the final structure produced after this run, used as the starting point for production MD simulations.

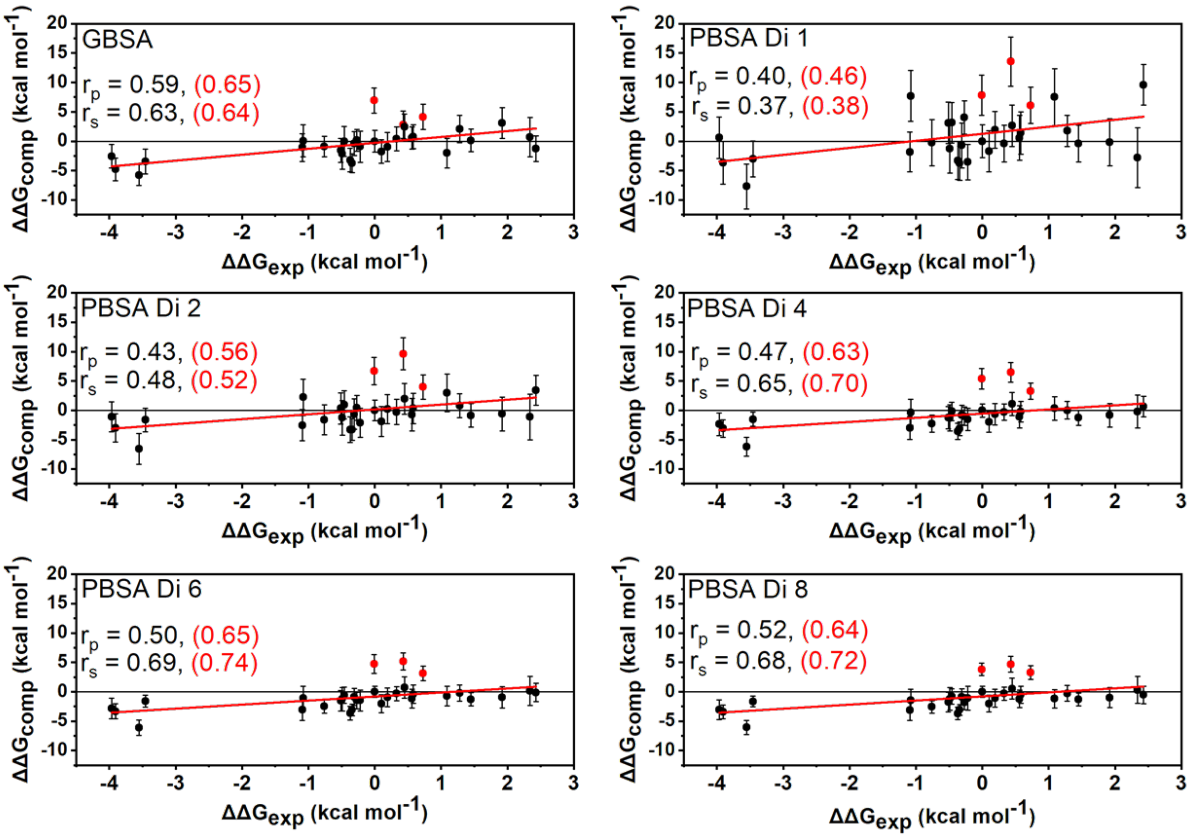
Truncated Normal Mode Analysis. Truncated normal mode analysis calculations (Trunc-NMA) were performed only on the 1G4 TCR-pHLA complexes studied in this manuscript. Residues in the pHLA that were retained were HLA residues: 5-26, 33-47, 54-101, 112-118, 123-133, 139-174, and peptide residues: 1-9. Residues in the TCR that were retained were TCRα-chain: 1-5, 23-36, 44-73, 89-105 and TCRβ-chain: 1-4, 22-34, 42-75, 89-104. Residues that were restrained during energy minimization and vibrational frequency calculations were: HLA residues: 6, 8, 10, 15-16, 23, 25-26, 33-35, 47, 54, 85-88, 93-94, 96, 98, 100-101, 112-113, 115, 117-118, 123, 125, 127-129, 139-141, 169 and172-174; TCRα-chain residues: 4-5, 23-24, 26, 35, 45-48, 60-65, 67, 67-73, 89-91, 105; and TCRβ-chain residues: 1, 3, 22, 42-43, 59, 61-64, 74-75, 89-90, 104. The "closest" command in the Ambertools18 program CPPTRAJ[6] was used to retain 1000 binding site water molecules in each frame. The 1000 closest water molecules (water position determined using the oxygen atom) to the Cα of any of central peptide residues (residues 4, 5, 6 and 7) were used for this closest water molecules calculation. We used a modified version of the Nmode program from AmberTools14, which allows for the use of the "ibelly" parameter which fixes selected atoms in place, meaning they have no (direct) contribution to the vibrational frequency calculation. The modified Nmode code (along with scripts to setup and analyse the obtained results) can be provided upon request to the corresponding authors.

Inclusion of Explicit Water Molecules for MMPB/GBSA Calculations. MMPB/GBSA calculations performed with explicit solvent (either 10, 20, 30 or 50 water molecules) were selected for using the "closest" command with CPPTRAJ.[6] Water molecules were included in the complex and receptor calculations, and receptor residues at the binding site were used to select the *x* closest water molecules. Specifically, for WT 1G4, HLA residues: 19, 62, 65-66, 68, 69, 71-73, 75, 76, 150, 151, 154, 155, 158, 159, 163, 167 and peptide residues: 4-8 were used. For WT A6, HLA residues: 62, 65, 66, 69, 72-73, 150, 155, 158, 163, 166 and peptide residues: 1-7 were used. The *x* closest water oxygen atom distances to any nitrogen or oxygen atom on the defined receptor residues were then used. The following selection expression was used in CPPTRAJ: "(:receptor_residue_selection)&(@O*|@N*)".

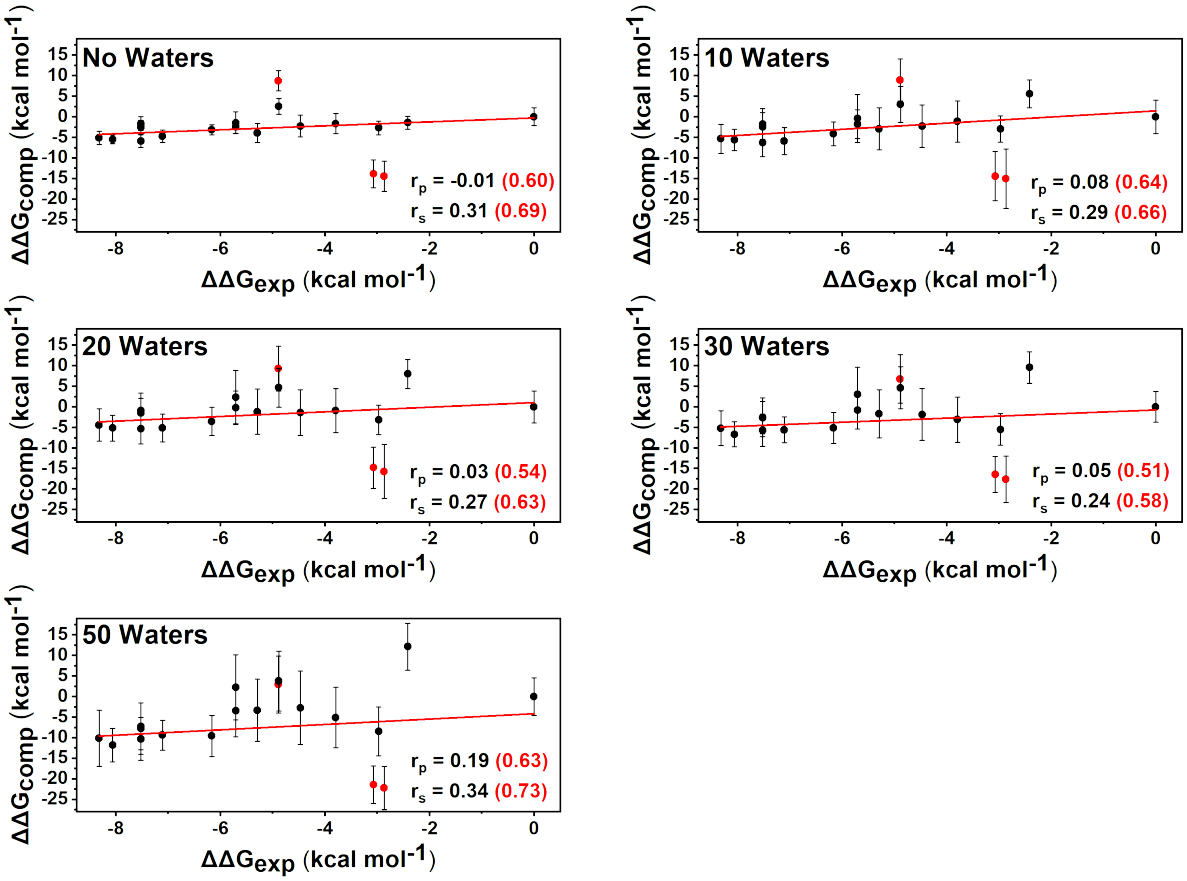
Supplementary figures



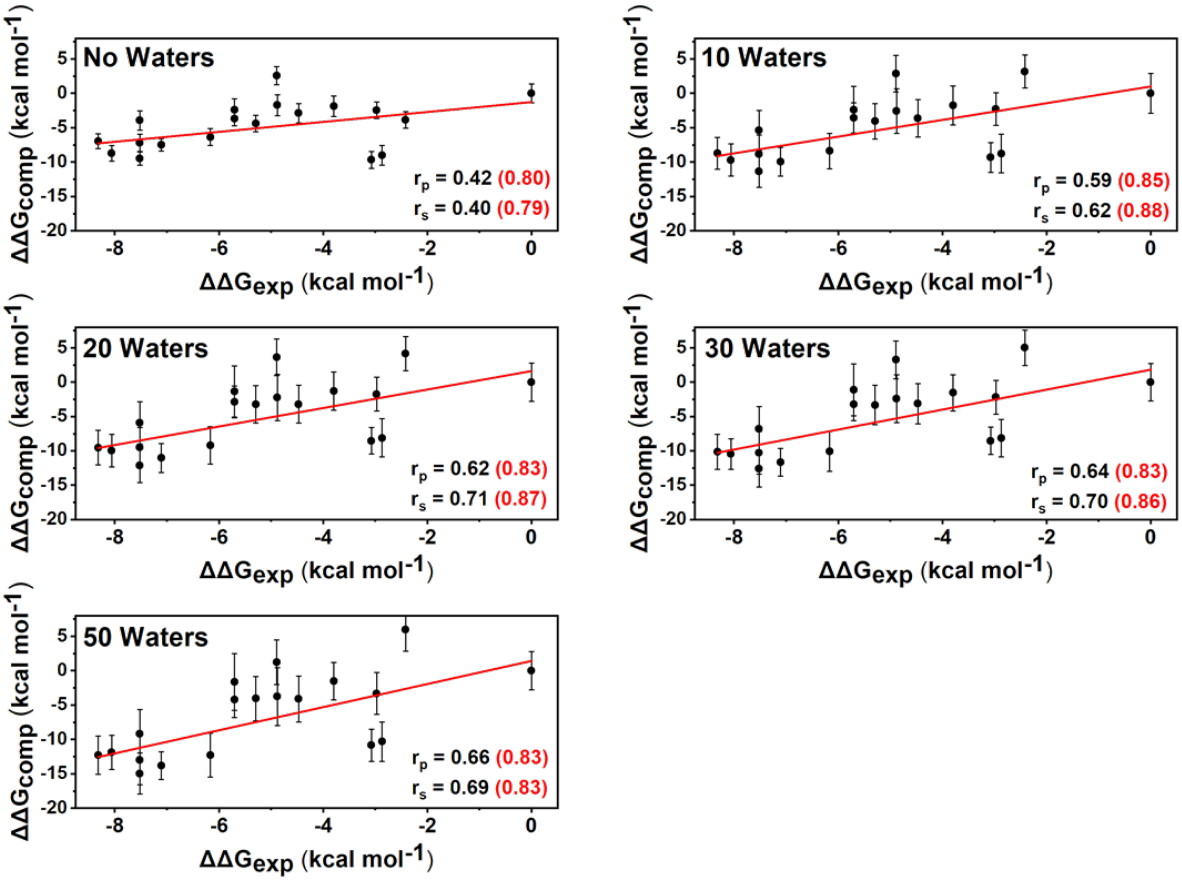
S1 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the 1G4 set of TCRs. For each panel, the MMGB/PBSA methods used are indicated in the top right. The red line corresponds to a linear fit of all the data. The three variants colored in red correspond to the outliers (NY-6, NY-33 and NY-33A) discussed in the main text. The Pearson's r value (r_p) and Spearman's rank (r_s) are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



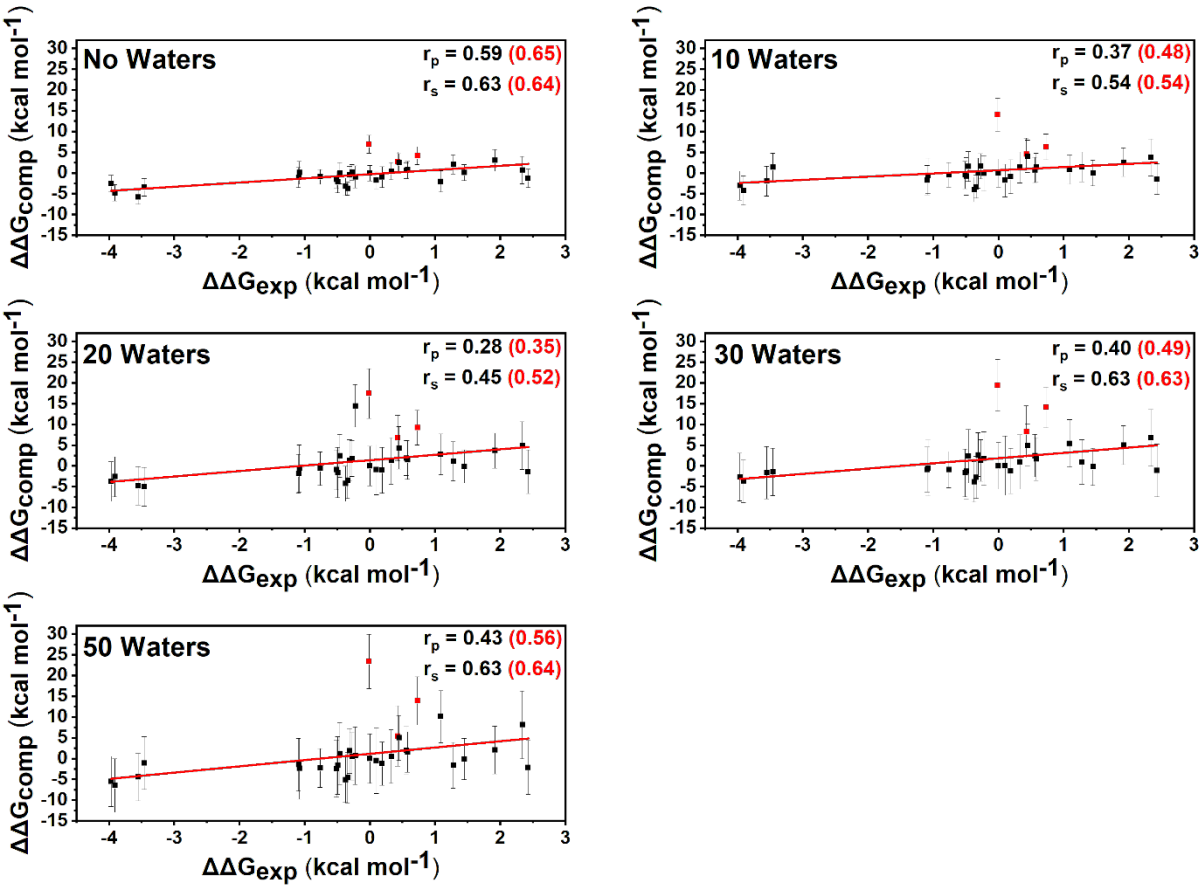
S2 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the A6 set of TCRs. For each panel, the MMGB/PBSA methods used are indicated in the top right. The red line corresponds to a linear fit of all the data. The Pearson's r value (r_p) and Spearman's rank (r_s) for each method applied are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



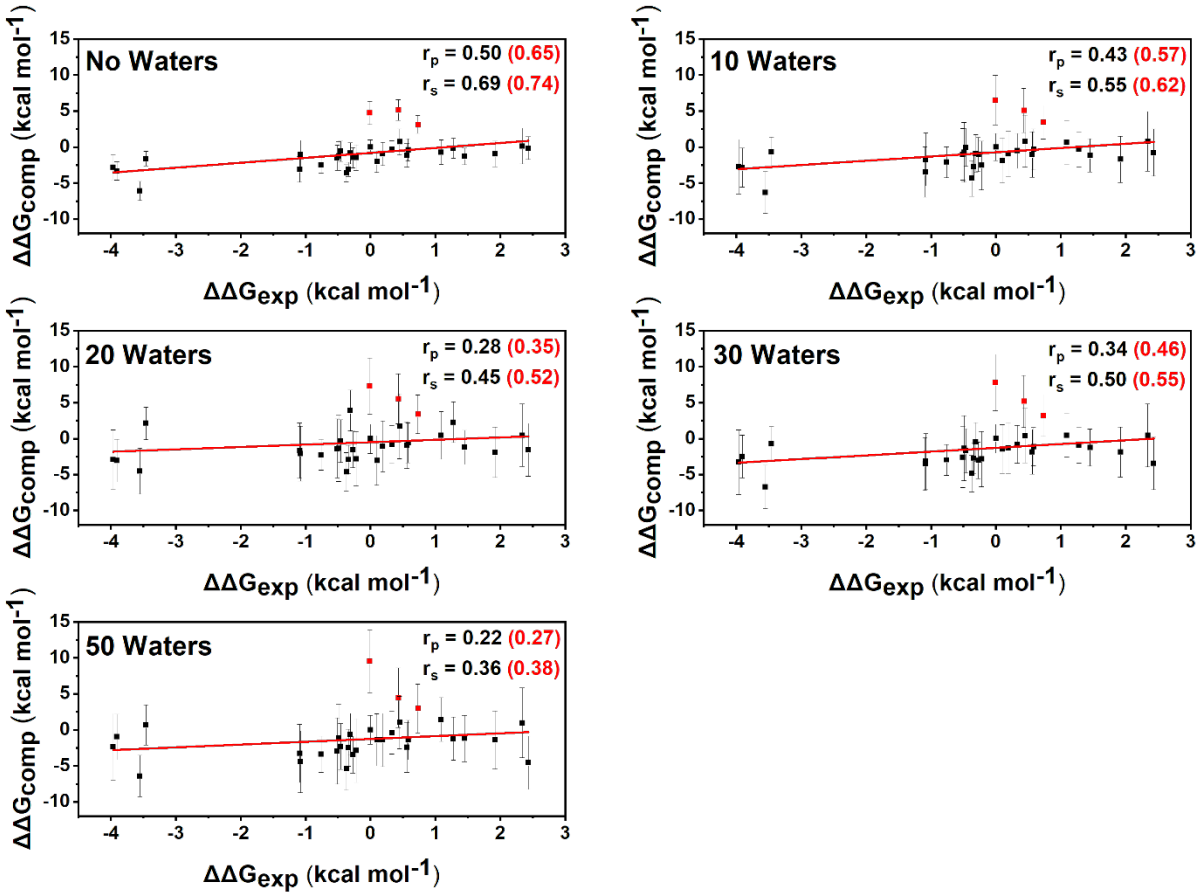
S3 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the 1G4 set of TCRs using the MMGBSA method with different amounts of explicit waters included. The number of waters used for each calculation is stated in the top left corner of each plot. The red line corresponds to a linear fit of all the data. The three variants colored in red correspond to the outliers (NY-6, NY-33 and NY-33A) discussed in the main text. The Pearson's r value (r_p) and Spearman's rank (r_s) are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



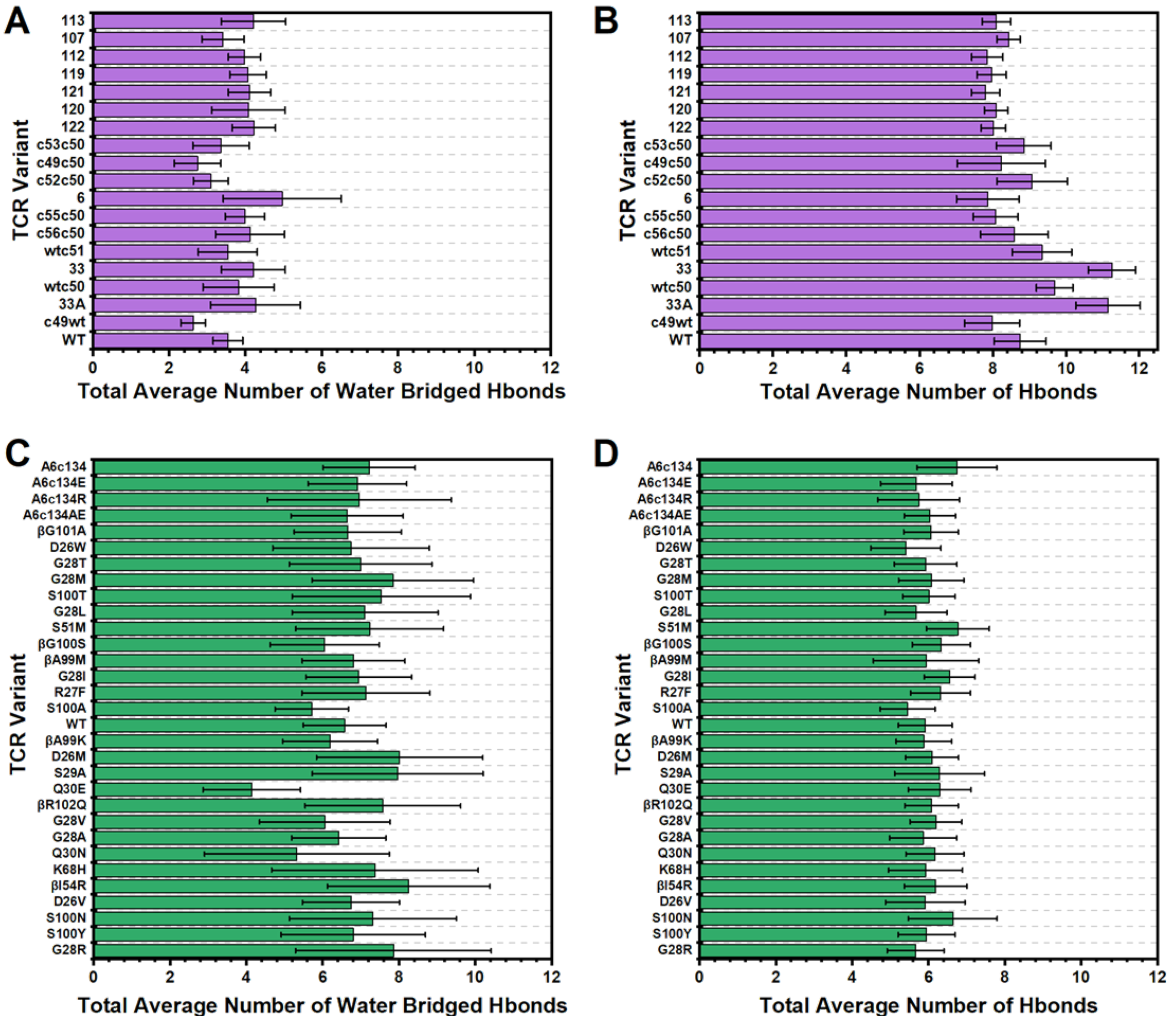
S4 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the 1G4 set of TCRs using the MMPBSA method with the internal dielectric constant set to 6 and a varied number of waters included. The number of waters used for each calculation is stated in the top left corner of each plot. The red line corresponds to a linear fit of all the data. The three variants colored in red correspond to the outliers (NY-6, NY-33 and NY-33A) discussed in the main text. The Pearson's r value (r_p) and Spearman's rank (r_s) are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



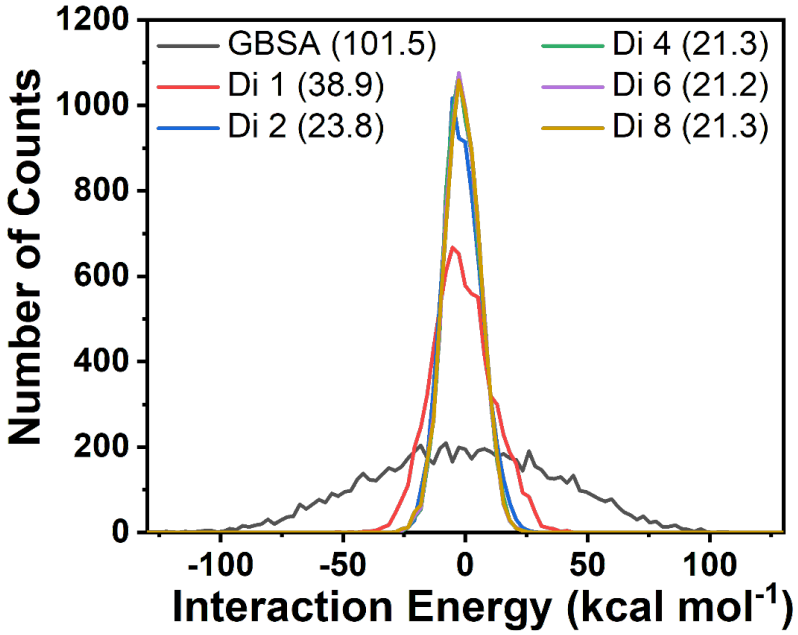
S5 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the A6 set of TCRs using the MMGBSA method and a variable amount of explicit water molecules. The number of waters used for each calculation is stated in the top left corner of each plot. The red line corresponds to a linear fit of all the data. The three variants colored in red correspond to the outliers (point mutants Q30E, S100A and S100N) discussed in the main text. The Pearson's r value (r_p) and Spearman's rank (r_s) are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



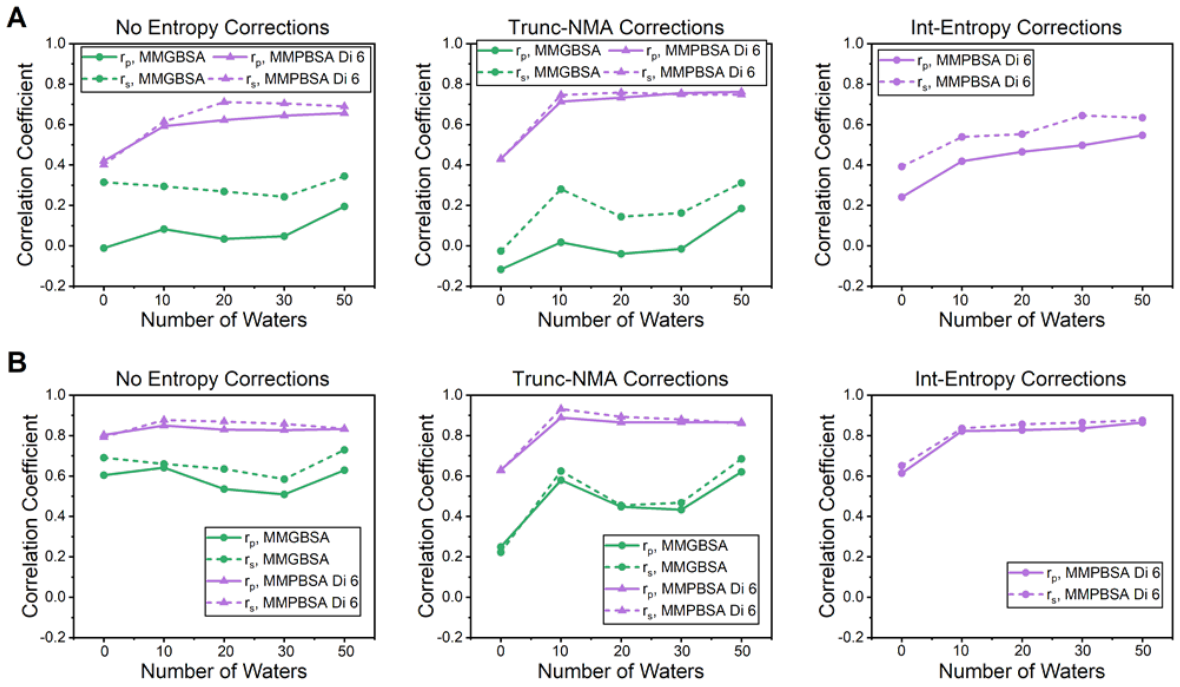
S6 Fig. Computational *versus* experimental ΔΔ*G* values obtained for the A6 set of TCRs using the MMPBSA method with the internal dielectric constant set to 6 and with a varied number of waters included. The number of waters used for each calculation is stated in the top left corner of each plot. The red line corresponds to a linear fit of all the data. The three variants colored in red correspond to the outliers (point mutants Q30E, S100A and S100N) discussed in the main text. The Pearson's r value (r_p) and Spearman's rank (r_s) are provided for the full set of data (in black) and without the three identified outliers discussed in the main text (in red and in parenthesis).



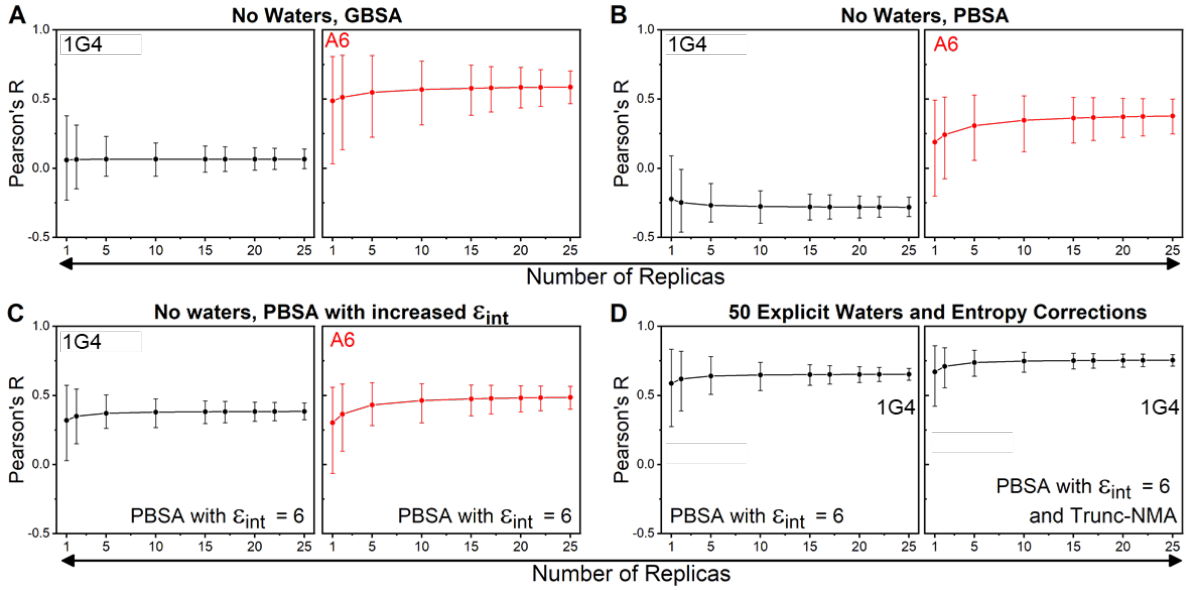
S7 Fig. Total average number of water bridged and solute-solute hydrogen bonds (Hbonds) formed between the TCR and pHLA for each TCR variant simulated. Results for both the 1G4 (**A**+**B**) and A6 (**C**+**D**) test sets are shown. Error bars are the standard deviation from the 25 replicas performed per complex. Water bridged Hbonds for the 1G4 and A6 TCRs are shown in panels **A** and **C** respectively, whilst solute-solute Hbonds are shown in panels **B** and **D** respectively. For the A6 TCRs (panels **C**+**D**), single point mutations preceded by the letter "β", mean the mutation is located on the TCRβ chain (with all others therefore located on TCRα chain). A hydrogen bond is defined as having a donor-acceptor distance less than 3.5 Å, and a donor-hydrogen-acceptor angle within 180 ± 45°. For each graph, the TCR variants are ordered according to their affinity (highest affinity at the top, lowest affinity at the bottom).



S8 Fig. Histograms of the fluctuation of gas phase interaction energy. Histograms were obtained by solving **Equation 5** in the main text for the WT-1G4 TCR-pHLA complex using the six different methods evaluated in this study (performed with no explicit water molecules included). The value in brackets corresponds to the entropy correction estimate i.e. $-T\Delta S$ (in kcal mol^{-1}) obtained from each method when using the interaction entropy (Int-Entropy) approach.



S9 Fig. Impact of entropy corrections to the 1G4 test set as determined by the Pearson's R (r_p) and Spearman's rank (r_s). The three panels in section (**A**) contain the results: No entropy corrections (left) with truncated normal mode analysis (Trunc-NMA) corrections (middle) and Interaction Entropy (Int-Entropy) corrections with all data included (*i.e.* outliers described in the main text are included). The three panels in section (**B**) show the same data but with the three outliers described in the main text excluded from the calculations.



S10 Fig. Bootstrapping to assess the impact of the number of replicas on the Pearson's R value for some of the protocols evaluated in this study. Panels **A**+**B** focus on the GBSA and PBSA approaches with no explicit waters included. Panel **C** focusses on the PBSA method with ϵ_{int} set to 6. Panel **D** focusses on the PBSA method (ϵ_{int} set to 6) with 50 explicit waters molecules included with and without the Trunc-NMA correction applied. Measurements with the 1G4 and A6 test sets are colored black and red respectively. In each panel, 1 million bootstrap resamples are used to calculate Spearman's Rank when using a differing number of replicas. Error bars shown are the 95% confidence intervals. The complete data is used in all cases (i.e. the outliers discussed above are included). Equivalent results for the Spearman's rank metric are provided in the main text (**Figure 7**).

Supplementary tables

Table S1. Histidine tautomerisation states used for 1G4 and A6 TCR-pHLA complex MD simulations.

TCR-pHLA System	HID[a] Tautomers	HIE[b] Tautomers
1G4	HLA: 3, 70, 74, 93, 114, 192, 260. β2m: 51. Peptide: N/A. TCRα: N/A. TCRβ: 152.	HLA: 145, 151, 188, 191, 197, 263. β2m: 13, 31, 84. Peptide: N/A. TCRα: 113. TCRβ: 28, 46, 135, 165, 205.
A6	HLA: 70, 74, 93, 114, 191. β2m: N/A. Peptide: N/A. TCRα: N/A. TCRβ: N/A.	HLA: 3, 145, 151, 188, 192, 197, 260, 263. β2m: 13, 31, 51, 84, Peptide: N/A TCRα: N/A TCRβ: 29, 47, 139, 156, 169, 209.

Numbering is consistent with PDB ID 2BNR[7] for 1G4 and PDB ID 1AO7[8] for A6 simulations.

[a] HID corresponds to a histidine residue which is singly protonated on its Nδ1 nitrogen.

[b] HIE corresponds to a histidine residue which is singly protonated on their Nε2 nitrogen.

Table S2. Experimental TCR-pHLA affinities for the 1G4 TCRs.

TCR ID	K_D (nM)	ΔG_{bind} (kcal mol^{-1})	$\Delta\Delta G_{bind}$ (kcal mol^{-1})	Reference
Data set 1[a]				
WT	32,000	-6.14	0	Li *et al.*[9]
6	8.4	-11.03	-4.89	Li *et al.*[9]
33	180	-9.21	-3.07	Li *et al.*[9]
33A	254	-9.01	-2.87	Li *et al.*[9]
107	0.04	-14.20	-8.06	Li *et al.*[9]
112	0.1	-13.65	-7.52	Li *et al.*[9]
113	0.026	-14.45	-8.32	Li *et al.*[9]
119	0.1	-13.65	-7.52	Li *et al.*[9]
120	0.2	-13.24	-7.11	Li *et al.*[9]
121	0.1	-13.65	-7.52	Li *et al.*[9]
122	0.98	-12.30	-6.16	Li *et al.*[9]
Data set 2[a]				
WT	15,000	-6.59	0	Dunn *et al.*[10]
c49wt	255	-9.00	-2.42	Dunn *et al.*[10]
c49c50	1	-12.29	-5.70	Dunn *et al.*[10]
c52c50	2	-11.88	-5.29	Dunn *et al.*[10]
c53c50	1	-12.29	-5.70	Dunn *et al.*[10]
c55c50	4	-11.47	-4.88	Dunn *et al.*[10]
c56c50	8	-11.06	-4.47	Dunn *et al.*[10]
wtc50	100	-9.56	-2.97	Dunn *et al.*[10]
wtc51	25	-10.38	-3.79	Dunn *et al.*[10]

TCR IDs used correspond to the same as those in the original paper indicated.

[a] For both different studies, $\Delta\Delta G$ values were determined using the WT affinity determined in the same study.

Table S3. Experimental TCR point mutation data with TCR-pHLA affinities for the A6 TCR series.

TCR I.D.	K_D (nM)	ΔG_{bind} (kcal mol^{-1})	$\Delta\Delta G_{bind}$ (kcal mol^{-1})	Reference
Data set 1 (all single point mutations)[a]				
WT	3200	-7.50	0	Haidar *et al.*[11]
(CDRα) S51M	N.P.[b]	N.P.[b]	-0.37	Haidar *et al.*[11]
(CDRα) S29A	N.P.[b]	N.P.[b]	0.33	Haidar *et al.*[11]
(CDRα) S100Y	N.P.[b]	N.P.[b]	2.34	Haidar *et al.*[11]
(CDRα) S100T	N.P.[b]	N.P.[b]	-0.49	Haidar *et al.*[11]
(CDRα) S100N	N.P.[b]	N.P.[b]	1.92	Haidar *et al.*[11]
(CDRα) S100A	N.P.[b]	N.P.[b]	-0.01	Haidar *et al.*[11]
(CDRα) R27F	N.P.[b]	N.P.[b]	-0.22	Haidar *et al.*[11]
(CDRα) Q30N	N.P.[b]	N.P.[b]	0.73	Haidar *et al.*[11]
(CDRα) Q30E	N.P.[b]	N.P.[b]	0.43	Haidar *et al.*[11]
(CDRα) K68H	N.P.[b]	N.P.[b]	1.09	Haidar *et al.*[11]
(CDRα) G28V	N.P.[b]	N.P.[b]	0.56	Haidar *et al.*[11]
(CDRα) G28T	N.P.[b]	N.P.[b]	-0.76	Haidar *et al.*[11]
(CDRα) G28R	N.P.[b]	N.P.[b]	2.43	Haidar *et al.*[11]
(CDRα) G28M	N.P.[b]	N.P.[b]	-0.51	Haidar *et al.*[11]
(CDRα) G28L	N.P.[b]	N.P.[b]	-0.46	Haidar *et al.*[11]
(CDRα) G28I	N.P.[b]	N.P.[b]	-0.27	Haidar *et al.*[11]
(CDRα) G28A	N.P.[b]	N.P.[b]	0.58	Haidar *et al.*[11]
(CDRα) D26W	N.P.[b]	N.P.[b]	-1.08	Haidar *et al.*[11]
(CDRα) D26V	N.P.[b]	N.P.[b]	1.45	Haidar *et al.*[11]
(CDRα) D26M	N.P.[b]	N.P.[b]	0.19	Haidar *et al.*[11]
(CDRβ) R102Q	N.P.[b]	N.P.[b]	0.45	Haidar *et al.*[11]
(CDRβ) I54R	N.P.[b]	N.P.[b]	1.28	Haidar *et al.*[11]
(CDRβ) G101A	N.P.[b]	N.P.[b]	-1.09	Haidar *et al.*[11]
Data set 2[a]				
WT	3200	-7.50	0	Cole *et al.*[12]
A6c134M	1900	-7.81	-0.31	Cole *et al.*[12]
A6c134S	1800	-7.84	-0.34	Cole *et al.*[12]
A6c134AE	9.4	-10.96	-3.46	Cole *et al.*[12]
A6c134E	4.4	-11.41	-3.91	Cole *et al.*[12]
A6c134R	8	-11.06	-3.55	Cole *et al.*[12]
A6c134	4	-11.47	-3.96	Cole *et al.*[12]

TCR IDs used correspond to the same as those in the original paper indicated.

[a] For both different studies, $\Delta\Delta G$ values were determined using the WT affinity determined in the same study.

[b] N.P. means not provided in the original publication (only $\Delta\Delta G$ values were provided).

Table S4. CDR Loop sequences for the 1G4 set of TCRs.

TCR I.D.	CDR2α 50-54	CDR3α 94-103	CDR2β 50-53	TCRβF3 56-70	CDR3β 94-97
WT	IQSSQ	PTSGGSYIPT	GAGI	QGEVPNGYNVSRSTT	YVGN
122	I**SPW**Q	P**LLD**G**T**YIPT	**AIQT**	QGEVPNGYNVSRST**I**	YVG**D**
121	I**SPW**Q	P**LLD**G**T**YIPT	**AIQT**	QGEVPNGYNVSRST**I**	YVGN
120	I**TPW**Q	P**LLD**G**T**YIPT	**AIQT**	QGEVPNGYNVSRST**I**	YVG**D**
119	I**TPW**Q	P**LLD**G**T**YIPT	**AIQT**	QGEVPNGYNVSRST**I**	YVGN
113	I**TPW**Q	P**LLD**G**T**YIPT	**AIQT**	**R**GEVPNGYNVSRST**I**	Y**L**GN
112	I**SPW**Q	P**LLD**G**T**YIPT	**AIQT**	**R**GEVPNGYNVSRST**I**	Y**L**GN
107	I**SPW**Q	P**FT**GG**G**YIPT	**AIQT**	QGEVPNGYNVSRSTT	YVGN
6	IQSSQ	**H**TS**N**G**YF**P**PT	GAG**T**	**R**GEVPNGYNVSRST**I**	Y**L**GN
33	IQSSQ	P**YQS**G**H**Y**M**PT	GAG**T**	**R**GEVPNGYNVSRST**I**	Y**L**GN
33A	IQSSQ	P**YQS**G**H**Y**M**PT	GAG**T**	QGEVPNGYNVSRST**I**	Y**L**GN
c49c50	I**PFW**Q	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
c49wt	I**PFW**Q	PTSGGSYIPT	GAGI	QGEVPNGYNVSRSTT	YVGN
c52c50	I**SPW**Q	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
c53c50	I**TPW**Q	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
c55c50	I**MGH**Q	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
c56c50	I**MGT**Q	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
wtc50	IQSSQ	PTSGGSYIPT	**SV**G**M**	QGEVPNGYNVSRSTT	YVGN
wtc51	IQSSQ	PTSGGSYIPT	**AIQT**	QGEVPNGYNVSRSTT	YVGN

WT positions subjected to mutagenesis are underlined. Mutations are colored in red and in bold. The unique extra proline mutation in the CDR3α of NY-6 alongside the different CDR3α loop sequence of NY-33 and NY-33A are highlighted in yellow.

Supplementary references

(1) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.

(2) Kovalenko, A.; Hirata, F. Potential of Mean Force between Two Molecular Ions in a Polar Molecular Solvent: A Study by the Three-Dimensional Reference Interaction Site Model. *J. Phys. Chem. B* **1999**, *103*, 7942–7957.

(3) Kovalenko, A.; Hirata, F.; Kinoshita, M. Hydration Structure and Stability of Met-Enkephalin Studied by a Three-Dimensional Reference Interaction Site Model with a Repulsive Bridge Correction and a Thermodynamic Perturbation Method. *J. Chem. Phys.* **2000**, *113*, 9830–9836.

(4) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal–Water Interface by the Kohn–Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **1999**, *110*, 10095–10112.

(5) Sindhikara, D. J.; Yoshida, N.; Hirata, F. Placevent: An Algorithm for Prediction of Explicit Solvent Atom Distribution-Application to HIV-1 Protease and F-ATP Synthase. *J. Comput. Chem.* **2012**, *33*, 1536–1543.

(6) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(7) Chen, J.-L.; Stewart-Jones, G.; Bossi, G.; Lissin, N. M.; Wooldridge, L.; Choi, E. M. L.; Held, G.; Dunbar, P. R.; Esnouf, R. M.; Sami, M.; Boulter, J. M.; Rizkallah, P.; Renner, C.; Sewell, A.; van der Merwe, P. A.; Jakobsen, B. K.; Griffiths, G.; Jones, E. Y.; Cerundolo, V. Structural and Kinetic Basis for Heightened Immunogenicity of T Cell Vaccines. *J. Exp. Med.* **2005**, *201*, 1243–1255.

(8) Garboczi, D. N.; Ghosh, P.; Utz, U.; Fan, Q. R.; Biddison, W. E.; Wiley, D. C. Structure of the Complex between Human T-Cell Receptor, Viral Peptide and HLA-A2. *Nature* **1996**, *384*, 134–141.

(9) Li, Y.; Moysey, R.; Molloy, P. E.; Vuidepot, A. L.; Mahon, T.; Baston, E.; Dunn, S.; Liddy, N.; Jacob, J.; Jakobsen, B. K.; Boulter, J. M. Directed Evolution of Human T-Cell Receptors with Picomolar Affinities by Phage Display. *Nat. Biotechnol.* **2005**, *23*, 349–354.

(10) Dunn, S. M.; Rizkallah, P. J.; Baston, E.; Mahon, T.; Cameron, B.; Moysey, R.; Gao, F.; Sami, M.; Boulter, J.; Li, Y.; Jakobsen, B. K. Directed Evolution of Human T Cell Receptor CDR2 Residues by Phage Display Dramatically Enhances Affinity for Cognate Peptide-MHC without Increasing Apparent Cross-Reactivity. *Protein Sci.* **2006**, *15*, 710–721.

(11) Haidar, J. N.; Pierce, B.; Yu, Y.; Tong, W.; Li, M.; Weng, Z. Structure-Based Design of a T-Cell Receptor Leads to Nearly 100-Fold Improvement in Binding Affinity for PepMHC. *Proteins Struct. Funct. Bioinforma.* **2009**, *74*, 948–960.

(12) Cole, D. K.; Sami, M.; Scott, D. R.; Rizkallah, P. J.; Borbulevych, O. Y.; Todorov, P. T.; Moysey, R. K.; Jakobsen, B. K.; Boulter, J. M.; Baker, B. M.; Li, Y.; Yi Li. Increased Peptide Contacts Govern High Affinity Binding of a Modified TCR Whilst Maintaining a Native PMHC Docking Mode. *Front. Immunol.* **2013**, *4*, 1–8.