

# Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas

Bernhard Knapp,<sup>\*,†,‡,§</sup> Luis Ospina,<sup>‡,§</sup> and Charlotte M. Deane<sup>‡,§</sup>

<sup>†</sup>Bioinformatics and Immunoinformatics Research Group, Department of Basic Sciences, International University of Catalonia, 08195 Barcelona, Spain

<sup>‡</sup>Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom

<sup>§</sup>Alliance Manchester Business School, University of Manchester, Manchester M13 9SS, United Kingdom

## Supporting Information

**ABSTRACT:** Molecular simulations are a computational technique used to investigate the dynamics of proteins and other molecules. The free energy landscape of these simulations is often rugged, and minor differences in the initial velocities, floating-point precision, or underlying hardware can cause identical simulations (replicas) to take different paths in the landscape. In this study we investigated the magnitude of these effects based on 310 000 ns of simulation time. We performed 100 identically parametrized replicas of 3000 ns each for a small 10 amino acid system as well as 100 identically parametrized replicas of 100 ns each for an 827 residue T-cell receptor/MHC system. Comparing randomly chosen subgroups within these replica sets, we estimated the reproducibility and reliability that can be achieved by a given number of replicas at a given simulation time. These results demonstrate that conclusions drawn from single simulations are often not reproducible and that conclusions drawn from multiple shorter replicas are more reliable than those from a single longer simulation. The actual number of replicas needed will always depend on the question asked and the level of reliability sought. On the basis of our data, it appears that a good rule of thumb is to perform a minimum of five to 10 replicas.



## INTRODUCTION

Molecular simulations are a computational technique<sup>1</sup> with a wide range of applications. Examples include protein folding,<sup>2</sup> protein dynamics,<sup>3</sup> RNA<sup>4</sup>/DNA<sup>5</sup> dynamics, membrane dynamics,<sup>6</sup> and the dynamics of large systems.<sup>7</sup>

Multiple simulations of identical atomic coordinates are called replicas, repeats, or independent ensembles. In this study, the term “replicas” refers to simulations of identical structures with identical parameters where only the initial velocities are created randomly according to a Maxwell distribution. Velocities are usually chosen randomly because (1) the system is assumed ergodicity of the system, (2) velocities cancel out of ensemble averages, and (3) the emerging Markovian property of complex dynamical systems makes the detail of initial velocities unimportant after very short times. Each replica simulation runs independently and is not influenced by the other replicas. Replicas of this type often produce differing trajectories<sup>8</sup> caused by the ruggedness of the energy landscape, which contains many local minima often separated by high energy barriers.<sup>9</sup>

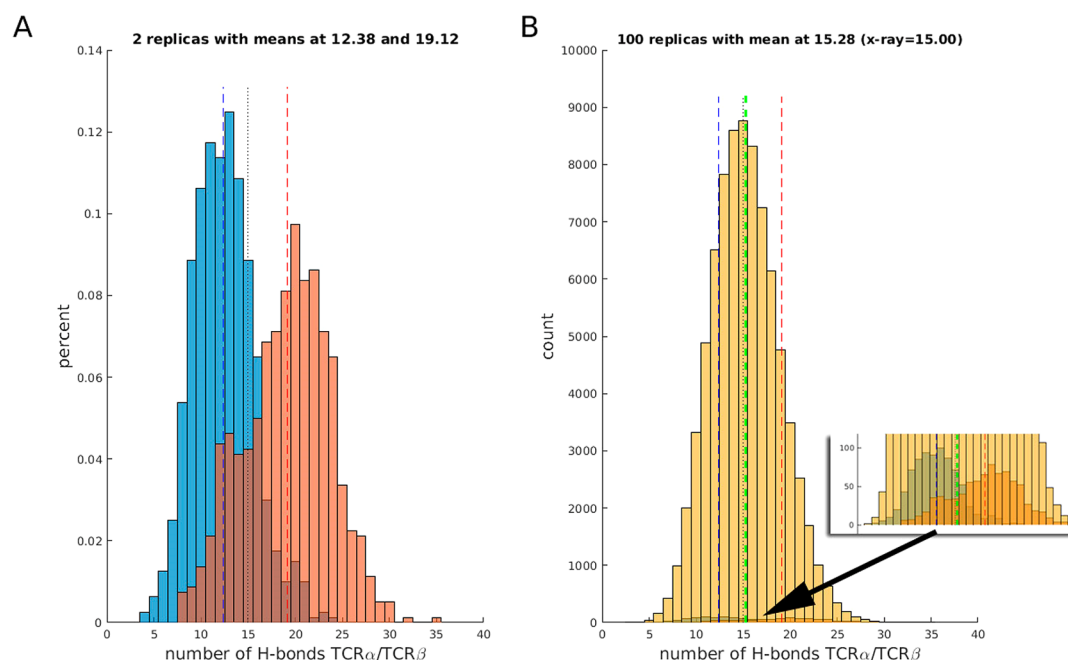
Random velocities are not the only cause of differing trajectories: replicas with identical initial velocities also produce different trajectories because of differences in floating-point precision on different machines, the number and type of processors, compiling options, system-specific random number generators,

and/or dynamic load balancing.<sup>8</sup> The central limit theorem and ergodic hypothesis for dynamical systems tell us that in principle this should not matter for infinitely long simulations. The problem, however, is that real-life simulations are not infinitely long, and quantifying the uncertainty to which this leads is challenging.<sup>10</sup>

With the assumptions of finite runtime and finite sampling, the question arises as to the simulation time and number of replicas that are necessary for reliable and reproducible conclusions to be drawn from molecular simulation studies. In the following we explain this using an example: An often seen study is the investigation of the dynamics of an X-ray structure beyond its static structure. In our example, a research group chooses the X-ray structure of a highly immunogenic T-cell receptor/peptide/major histocompatibility complex (TCRpMHC). They perform a single molecular dynamics (MD) simulation of 100 ns to investigate the molecular background of the experimentally known high immunogenicity. The group is interested in the number of H-bonds between the two TCR chains, as it has been suggested that the mechanical forces acting on the TCR could be involved in TCR triggering.<sup>11</sup> In this single simulation,

**Received:** April 25, 2018

**Published:** October 24, 2018



**Figure 1.** (A) Histograms of the number of intra-TCR-chain H-bonds based on two replicas of the same TCRpMHC structure. (B) Histogram of the intra-TCR-chain H-bonds based on 100 replicas. The two replicas of (A) are shown as tiny transparent overlays at the very bottom of the figure. The inset on the right shows a zoomed version. The areas of the two simulations of (A) occupy rather extreme parts of the search space and amount to 2% of the total yellow area (2 out of 100 replicas).

they find an average number of 19.12 intra-TCR-chain H-bonds (orange histogram in Figure 1A). In the X-ray structure, the number of H-bonds is 15. They could conclude from this analysis that the additional four transient H-bonds cause a rigidification within the TCR. Therefore, the cause of the high immunogenicity of this complex is the increased number of H-bonds. This conclusion could be further strengthened by quoting the original X-ray paper<sup>12</sup> that states that the immunological hotspot (Y7 of the peptide) binds tightly between CDR1 $\alpha$  and CDR3 $\beta$  of the TCR.

Imagine a second research group that independently and unaware of the research discussed above performs exactly the same simulation but uses a different initial seed for the velocities and/or hardware. They find an average number of 12.38 intra-TCR-chain H-bonds (blue histogram in Figure 1A). This simulation suggests that 2.6 H-bonds of the X-ray structure are present only transiently in the crystal environment and that the lower number of H-bonds observed in the simulation allows additional TCR flexibility in the interaction with CD3. This finding is proposed as the structural reason for the immunogenicity of this complex.

The two groups would have drawn two diametrically opposed conclusions based on two simulations of the same complex (replicas). An analysis of 100 replicas (yellow histogram in Figure 1B) shows that both conclusions would have been incorrect. The average number of intra-TCR-chain H-bonds over all 100 replicas is 15.28, which is almost identical to the X-ray value of 15. These 100 replicas show that the average number of intra-TCR-chain H-bonds is in reality not different from the number seen in the X-ray structure. Studies built only on the outcome of one simulation without replicas could lead to false positive conclusions and publications.

This example shows that single simulation studies may lead to false positive conclusions. However, only a few studies<sup>13,14</sup> have investigated the use of replicas, and to date no study has

quantified the magnitude of uncertainty within identical replicas at different simulation lengths. In this work, we performed a large-scale MD simulation study to quantify the differences between replica simulations. We consider two systems. The first was a small 10 amino acid “toy” example. In this system, we hoped to attain a complete exploration of all of the degrees of freedom. The second was the above-described 827 amino acid “real-world” TCRpMHC example. In this system, exploration of the entire configurational space is not possible, and therefore, we rather sought an exhaustive exploration of one or multiple local free energy basins around the folded X-ray state. For these two examples, we quantified the reliabilities of different numbers of replicas in combination with varying simulation lengths.

## METHODS

**Structures.** First we considered a small “toy” system, the structure of the chignolin mutant CLN025 (Protein Data Bank (PDB)<sup>15</sup> accession code 2RVD<sup>16</sup>). Despite its small size (only 10 amino acids), this structure shows the essential characteristics of a protein, such as thermal stability, a folding pathway, and a free energy surface.<sup>16</sup> The average folding time has been determined to be about 600 ns.<sup>2</sup> Despite its protein characteristics, we refer to this system in our study as “the peptide”.

Second we considered a large protein complex in which the Lc13 TCR is bound to the Epstein–Barr virus (EBV) peptide FLRGRAYGL loaded onto the human MHC allele HLA-B\*08:01. The structure formed by the 827 amino acids that make up this TCRpMHC complex is available via PDB accession code 1M15.<sup>12</sup> The entire system was used, including the constant regions of the TCR, as we recently showed their importance for reliable results.<sup>17</sup> This TCRpMHC structure is a good case study as it has been investigated previously.<sup>3,12</sup> We refer to this system as “TCRpMHC”. Missing side-chain atoms in the X-ray structures were modeled using SCWRL<sup>18</sup> via

the PeptX framework,<sup>19</sup> as this has been shown to yield the highest accuracy for peptide/MHC modeling.<sup>20,21</sup>

**MD Simulations.** Both systems, the peptide and TCRpMHC, were immersed in dodecahedral simulation boxes of explicit SPC water, allowing for a minimum distance of 1.5 nm between the box boundary and the protein. Random water molecules were replaced with Na<sup>+</sup> and Cl<sup>−</sup> ions to achieve a neutral charge and a salt concentration of 0.15 mol/L. Energy minimization using the steepest-descent algorithm was applied. Then the system was warmed from 0 K to 310 K using the velocity rescaling thermostat<sup>22</sup> and position restraints on all bond lengths over 500 ps. Production runs with random velocities generated at 310 K were carried out using GROMACS 4<sup>23</sup> and the GROMOS force field in the 53A6 parametrization<sup>24</sup> on the ARCUS-b cluster of the Oxford Advanced Research Computing (ARC) facility.

We performed 100 replica simulations on the peptide system, each of 3000 ns, yielding a total simulation time of 300 000 ns. These simulations total 500 times the folding time of the system, which means that it is likely that we achieved good coverage of the free energy landscape and solution space, i.e., that we sampled the set of (almost) all feasible atom arrangements.

The TCRpMHC system was simulated for 100 ns, and 100 replicas were performed, yielding a total simulation time of 10 000 ns. In this case, the simulation time did not allow complete sampling of the system. Because of the size of the system, 100 ns is a usual simulation time for current MD studies of TCRpMHC.<sup>25</sup>

**Methods of Trajectory Evaluation.** Trajectories were evaluated using the Gromacs tools *g\_hbond* and *g\_dist*<sup>23</sup> as well as the *gro2mat* package.<sup>26</sup> Simulations were manually inspected with VMD<sup>27</sup> and the *vmdICE* plugin.<sup>28</sup>

**Reliability and Reproducibility Analysis.** In this work, we analyzed the relationship among the number of replicas, simulation time, and reproducibility by comparing the similarities of groups of replicas. We performed our analysis as follows: From the set of 100 replicas, we uniformly sampled *n* replicas at random with replacement (group 1). We repeated this sampling to obtain a second set of replicas of size *n* (group 2). Then for each group we computed the average *A<sup>i</sup>* (*i* = 1, 2) of a single descriptor (e.g., the number of H-bonds or a specific distance) over all frames and replicas. We repeated this sampling procedure 5000 times and recorded the difference *d<sub>k</sub>* between group 1 and group 2 in each repetition (*k* = 1, 2, ..., 5000). We also computed and recorded the average of the 5000 *d<sub>k</sub>* values. This average corresponds to a point in Figure 2D. Each point represents a unique combination of the number of replicas *n* used within each group and the length of the trajectory (number of frames), which is truncated at different time points (*x* axis of Figure 2). This allows a visualization of the influence of the number of replicas and simulation length on the reliability and reproducibility.

## RESULTS

**The Peptide.** We analyzed the peptide system (PDB code 2RVD) using 100 replicas of 3000 ns each. An example of one replica is shown in the Appendix Movie.

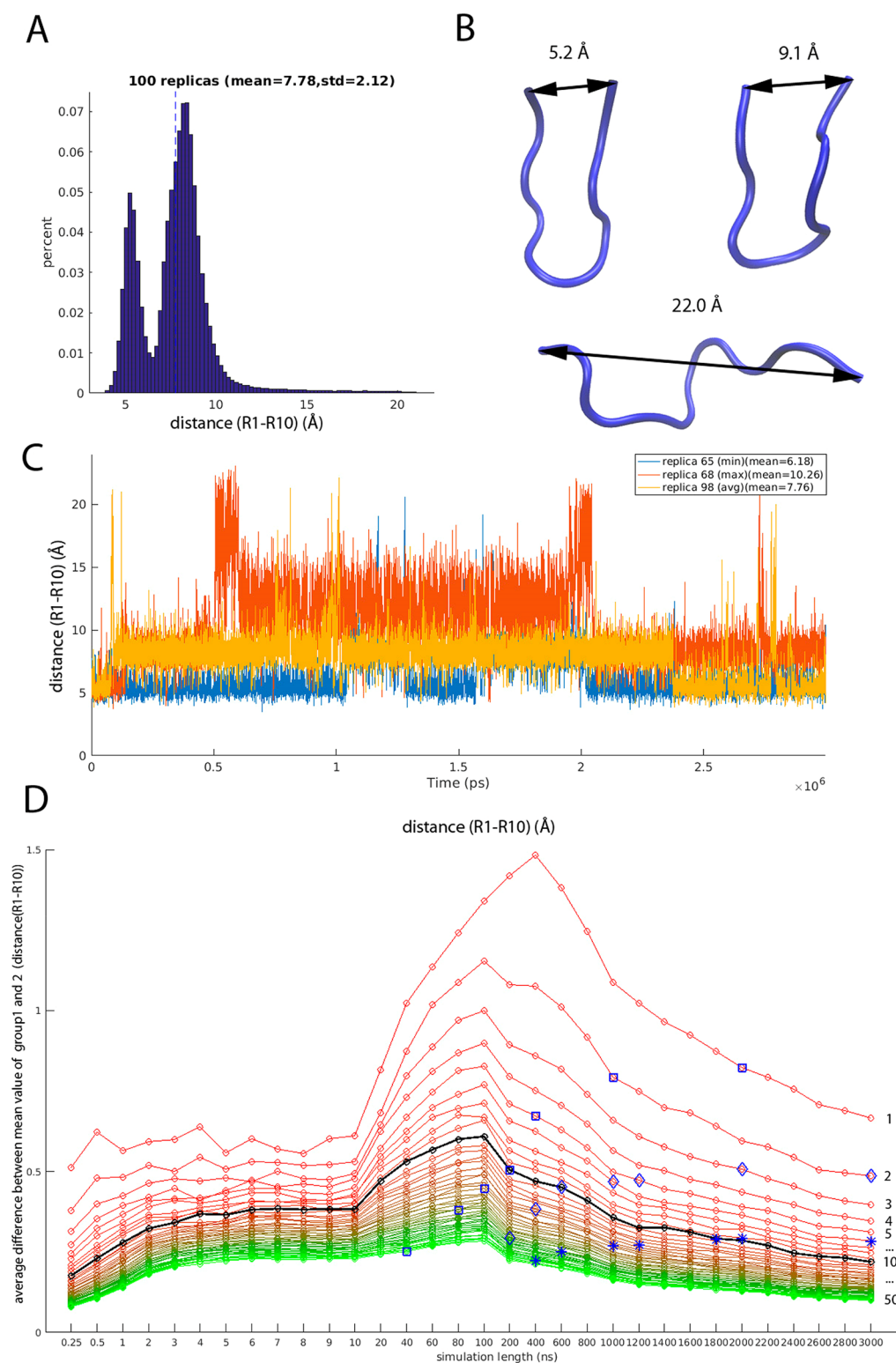
The first property we analyzed was the Euclidian distance between the termini of the peptides (the distance in space between residues Tyr1 and Tyr10). This is a simple property that discriminates between folded and unfolded states of the peptide. The distribution of the distance over all 100 replicas and time frames shows a bimodal distribution with means at 5.2 and

9.1 Å (Figure 2A). There is a slight tail to the right indicating the rare occurrence of elongated conformations. In Figure 2B we show an example conformation from each peak and a semi-elongated configuration. The overall distribution appears to be smooth, and the two peaks have nearly normal distributions around them. Figure 2C shows that not all replicas sample both states equally well. We depict the replicas with the on-average lowest, highest, and most average distance. Replica 65 spends most of the simulation time in a closed conformation (average distance 6.18 Å), while replica 68 spends more time in (semi)extended conformations (average distance 10.26 Å). Replica 98 has a mean value closest to the overall mean of all 100 replicas and also follows a similar distribution (average distance 7.76 Å).

While the most extreme examples among the 100 replicas differ greatly (e.g., Figure 1A and Figure 2C), the more interesting question is how much replicas differ on average from each other and how this difference changes with increased simulation time and/or more replicas per group. In Figure 2D we analyze this dependency (see Reliability and Reproducibility Analysis). Short simulations (<10 ns) tend to agree with each other. For example, two simulations of 10 ns differ on average by only 0.61 Å (top line in Figure 2D). However, this is not due to the good sampling quality but rather is due to undersampling and exploration of just the local neighborhood solution space (e.g., side-chain conformations) instead of the global solution space (e.g., unfolding and folding events of the backbone). This undersampling is reflected by the sharp increase in the average difference between groups between 10 and 400 ns. For example two simulations of 400 ns differ on average by 1.48 Å. Once the simulations are longer than 400 ns, the difference between groups starts to fall rapidly until about 1200 ns, after which the fall continues more slowly up to 3000 ns. If we move from the analysis of one replica per group to 10 (black line in Figure 2D), we find a median drop of 63.7% in the group difference. The curve is smoother and the decrease in group difference starts 200 ns earlier. If we then consider 50 replicas in each group, the median drop in difference is 83.7%. Even with 50 replicas per group and 3000 ns simulations, the difference between groups is not zero (green circle at the bottom right in Figure 2D). These differences can also be considered in the light of computational cost: using 10 replicas achieves 76.1% of the improvement of using 50 replicas at only 20% of the computational cost. The average standard deviation (over all 5000 permutations and 21 simulation lengths) of the replica group difference in peptide termini distance (in Å) drops considerably with increasing number of replicas from 1.23 (one replica) to 0.49 (five replicas), 0.32 (10 replicas), 0.18 (25 replicas), and 0.12 (50 replicas) (Appendix Figure 2).

Figure 2D shows the variation for a various simulation lengths and numbers of replicas. With these data we can also calculate the optimal way to invest computational resources. If it is assumed that budget restrictions allow only a certain amount of simulation time to be computed, is it better to run multiple replicas or longer simulations? In Figure 2D, the blue squares mark different possibilities assuming fixed computational resources of 2000 ns, the blue diamonds 6000 ns, and the blue stars 18 000 ns. In all three cases, the investment of computational time into a larger number of shorter replicas yields better results (lower difference in mean). For example, using one replica of 2000 ns per group yields a mean difference of 0.82 Å between groups, while using five replicas of 400 ns per group lowers the difference to 0.68 Å and using 20 replicas of 100 ns





**Figure 2.** Distance between peptide termini residues 1 and 10,  $d(R1-R10)$ . (A) Histogram of  $d(R1-R10)$  over all time frames of all 100 replicas. (B) Graphical illustration of three different states of the peptide: closed, open, and semi-elongated. (C)  $d(R1-R10)$  over time (3  $\mu$ s) shown for three example replicas. Replica 65 has on average over all time frames the lowest  $d(R1-R10)$ , while replica 68 has the highest. Replica 98 has an average  $d(R1-R10)$  very similar to the overall mean of all 100 replicas (7.76 vs 7.78). (D) Mean difference in  $d(R1-R10)$  between groups of replicas in relation to simulation time. For readability, the  $x$ -axis scale is not equally distributed. The analysis is shown for one replica (reddest line at the top) to 50 replicas (greenest line at the bottom). The lines change their color gradually from red to green representing the use of one to 50 replicas. The thick black line represents 10 replicas. Examples of using fixed amounts of computation time are indicated by symbols: blue squares mark a total computational investment of 2000 ns, blue diamonds 6000 ns, and blue asterisks 18 000 ns. The same plot with error bars for selected replica numbers is shown in [Appendix Figure 2](#).

lowers it to 0.44 Å. A complementary type of analysis that indicates the same results as the Figure 2D is shown in Appendix Figures 4 and 5.

In Appendix Figure 4 we show the robustness of our approach using a slightly different methodology: instead of using two groups of equal sizes, we use one group that consists of all 100 replicas and a second group of  $n$  replicas, where  $n$  runs from 1 to 50. In this way we illustrate the difference between the assumed truth (100 replicas) and  $n$  replicas. While the scales of the  $y$  axes in Figure 2 and Appendix Figure 4 are different, the overall shapes are highly similar, showing the robustness of our method.

In Appendix Figure 5 we show the same data, but we use the number of replicas (instead of the simulation length) for the  $x$  axis and the simulation time (instead of the number of replicas) for the color code. This figure emphasizes the convergence with increasing number of replicas, but the “peak” at around 100 ns in Figure 2 is only visible indirectly by overlapping lines. It can be seen that there is a strong decrease in average difference when more replicas are used at a fixed investment of 2000 ns but less difference at a fixed investment of 18 000 ns. This shows that performing more replicas instead of longer simulations is especially important if limited overall computational resources are available.

The distribution of the number of H-bonds over all 100 replicas and time frames is shown in Figure 3A. This is a more global property of the peptide. A nearly normal distribution with a mean of 6.49 H-bonds is observed. There is a slight tail to the left indicating few conformations with few or zero H-bonds. In Figure 3B we show three example conformations: a semi-elongated conformation with zero H-bonds and two closed conformations with two and eight H-bonds, respectively. Both open and closed conformations can have few or many H-bonds, but there is a moderate negative correlation ( $r = -0.504$ ) between the total number of H-bonds and the distance between the peptide termini.

Once again it is possible to find extreme examples within our 100 replicas that deviate from the overall mean of all replicas to a significant extent: 5.53 H-bonds for replica 68 and 6.97 H-bonds for replica 65 versus the overall mean of 6.49 H-bonds (Figure 3C). A detailed analysis of replica differences shows a picture similar to that for the peptide termini distance (Figure 2D). The main difference is that for very short simulations ( $\leq 5$  ns) there is no high agreement. This is due to the high on and off rates of H-bonds, where angle changes of a few degrees between the donor and acceptor can influence the presence/absence of a H-bond. Between around 5 and 40 ns an area of “misleading” good agreement between replicas is present (similar to the area at the beginning of the peptide termini analysis) that is probably caused by the lack of sampling of backbone configurations. The peak of disagreement between replicas is once again located between around 40 and 200 ns, with a rapid fall in average difference between groups to about 1000 ns and a slower fall afterward. Extending the analysis to more replicas, we find median drops in group difference of 64.5% for 10 replicas (black line in Figure 3D) and 83.9% for 50 replicas, but of using 10 replicas already achieves 76.7% of the improvement of using 50 replicas at 20% of the computational cost. The average standard deviation (over all 5000 permutations and 21 time cuts) of replica group difference in H-bonds drops considerably with the increasing number of replicas from 0.59 (one replica) to 0.23 (five replicas), 0.15 (10 replicas), 0.09 (25 replicas), and 0.06 (50 replicas) (Appendix Figure 3).

The analysis of efficient resource usage for H-bond reproducibility shows results similar to those for the peptide termini distance: using a larger number of shorter replicas yields better results (blue markers in Figure 3D). For example, using one replica of 2000 ns per group yields a mean difference of 0.32 H-bonds between groups, while using five replicas of 400 ns per group lowers the difference to 0.27 H-bonds and using 20 replicas of 100 ns yields a mean difference of 0.20 H-bonds.

**TCRpMHC.** We performed a total of 100 replicas, each of 100 ns, on the TCRpMHC protein complex (see Methods).

First we analyzed the Euclidian distance between the center of mass of the C-terminal peptide residue Leu9 and the center of mass of CDR3 $\beta$ . This is a local property of the TCRpMHC system. A high degree of flexibility in this distance is expected because terminal peptide residues are known to often unstably bind to the MHC.<sup>29–31</sup> The distribution of the distance over all 100 replicas and times shows a nearly normal distribution with a mean of around 13 Å (Figure 4A). In Figure 4B we show an example configuration with a distance of 13.5 Å. Figure 4C shows extreme examples that deviate from the overall mean value of 13 Å (10.5 Å for replica 58 and 15.6 Å for replica 25).

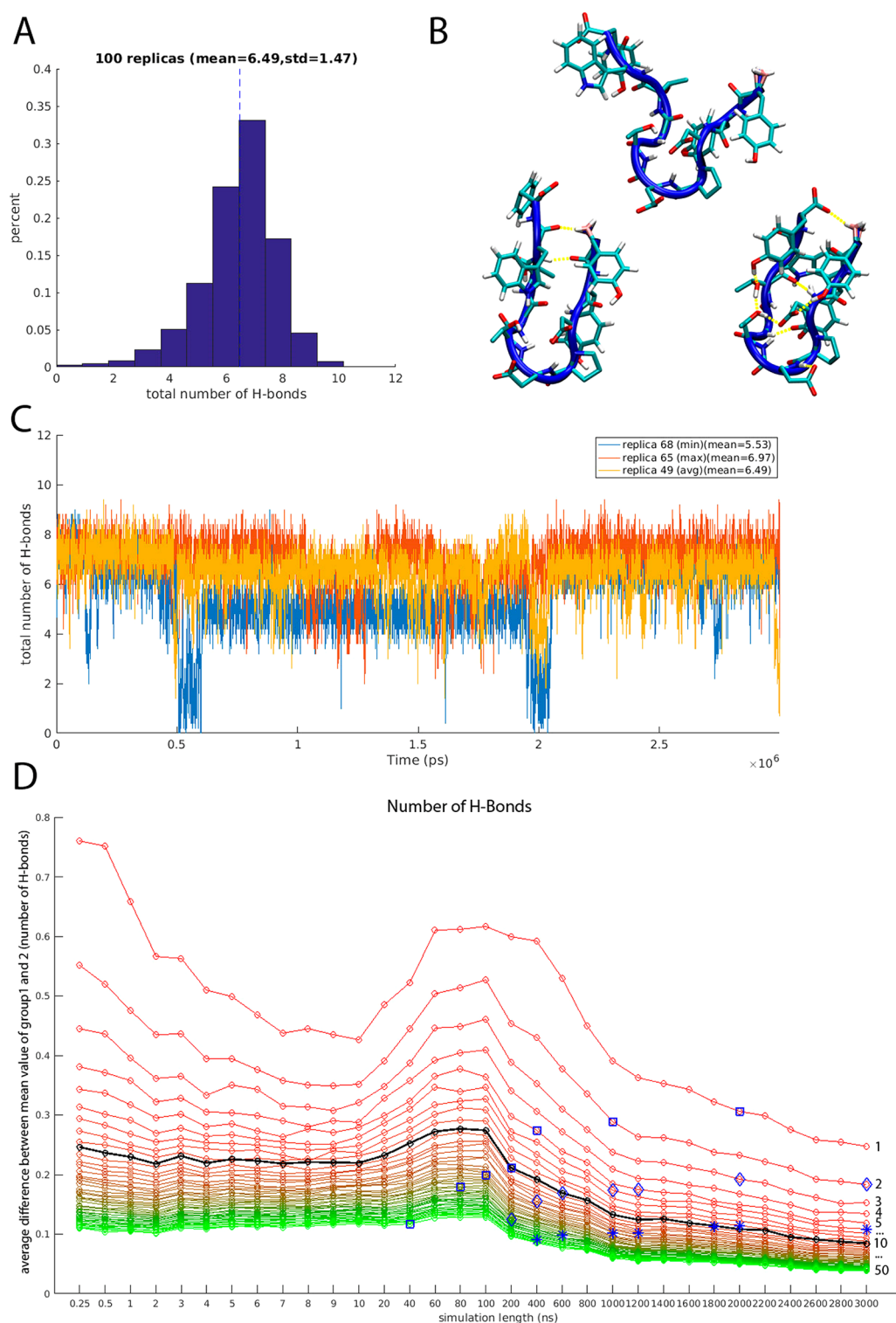
For one replica per group, the analysis of replica differences (Figure 4D) shows a quick increase in the average group difference up to about 4 ns and a slower but constantly present increase afterward. The “jump” between 10 and 20 ns is an intended optical effect as the  $x$ -axis step size changes from 1 to 10 ns. With five to 50 replicas per group, only a marginal increase in group difference is observable after 20 ns.

The use of more replicas per group reduces the average group difference: we find median drops in group difference of 64.9% for 10 replicas (black line in Figure 4D) and 84.1% for 50 replicas, but using 10 replicas already achieves 77.1% of the improvement of using 50 replicas at 20% of the computational cost. The average standard deviation (over all 5000 permutations and 21 time cuts) of the replica group difference in distance (in Å) drops considerably with increasing number of replicas from 0.63 (one replica) to 0.25 (five replicas), 0.18 (10 replicas), 0.11 (25 replicas), and 0.08 (50 replicas) (Appendix Figure 6).

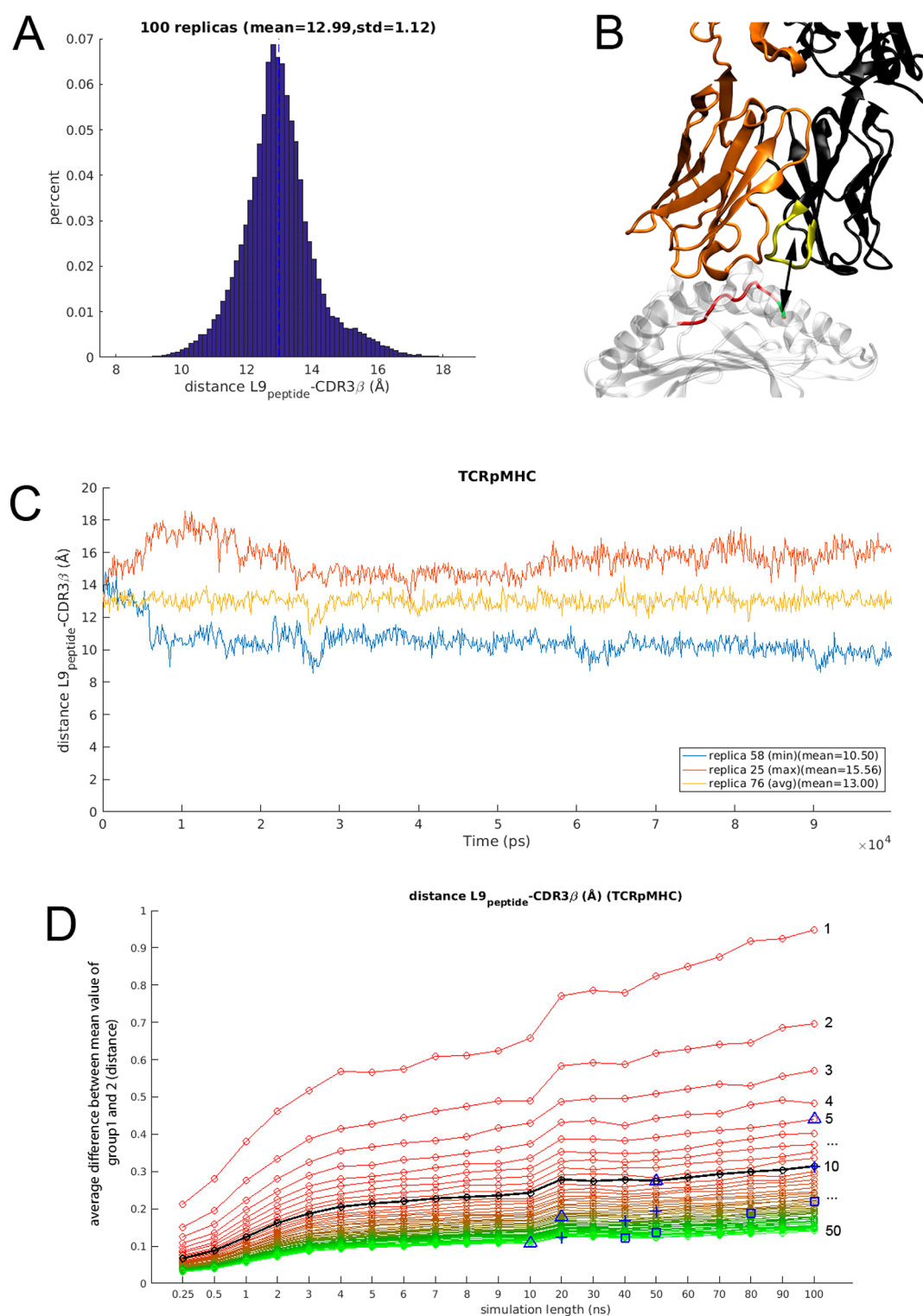
For this larger system, the analysis of efficient resource usage for reproducibility shows results similar to those for the peptide system: Using a larger number of shorter replicas yields better results (blue markers in Figure 4D). For example, using five replicas of 100 ns yields a mean difference of 0.44 Å between groups, while using 10 replicas of 50 ns yields a mean difference of 0.28 Å.

Finally, we analyzed the number of H-bonds between the TCR  $\alpha$ - and  $\beta$ -chains based on our 100 TCRpMHC replica simulations. This is the same descriptor that we used as motivation in the Introduction (Figure 1), and it takes into account many amino acids of the TCR. A nearly normal distribution of the number of H-bonds is shown in Figure 5A, and Figure 5B illustrates an example configuration with 15 H-bonds present. As expected, the number of intra-TCR-chain H-bonds is not correlated with the distance between Leu9 of the peptide and CDR3 $\beta$  ( $r = 0.042$ ). The most extreme examples of replicas were found to have on average 12.38 (replica 49) and 19.12 (replica 38) intra-TCR-chain H-bonds (Figure 5C).

The analysis of replica differences shows a more stable picture than the Leu9–CDR3 $\beta$  distance analysis (Figure 5D). After a slight increase in group difference over the first nanosecond the average does not change much over the remaining 99 ns. However, the use of more replicas reduces the average group difference: we find median drops in group difference of 68.4%



**Figure 3.** Total number of H-bonds within the peptide. (A) Histogram of the number of H-bonds over all time frames of all 100 replicas. (B) Graphical illustration of three different states facilitating zero, two, and eight H-bonds. The H-bonds are illustrated as dotted yellow lines. (C) Number of H-bonds over time (3  $\mu$ s) shown for three example replicas. Replica 68 has on average over all time frames the lowest number of H-bonds, while replica 66 has the highest. Replica 49 has an average number of H-bonds almost identical to the overall mean of all 100 replicas (6.49 in both cases). As H-bonds can change their state from present to nonpresent very quickly, the plot uses a running average over a window size of 500 ps for readability. The histogram uses the original data. (D) Mean difference in the number of H-bonds between groups of replicas in relation to simulation time. For readability, the x-axis scale is not equally distributed. The analysis is shown for one replica (reddest line at the top) to 50 replicas (greenest line at the bottom). The lines change their color gradually from red to green representing the use of one to 50 replicas. The thick black line represents 10 replicas. Examples of using fixed amounts of computation time are indicated by symbols: blue squares mark a total computational investment of 2000 ns, blue diamonds 6000 ns, and blue asterisks 18 000 ns. The same plot with error bars for selected replica numbers is shown in [Appendix Figure 3](#).

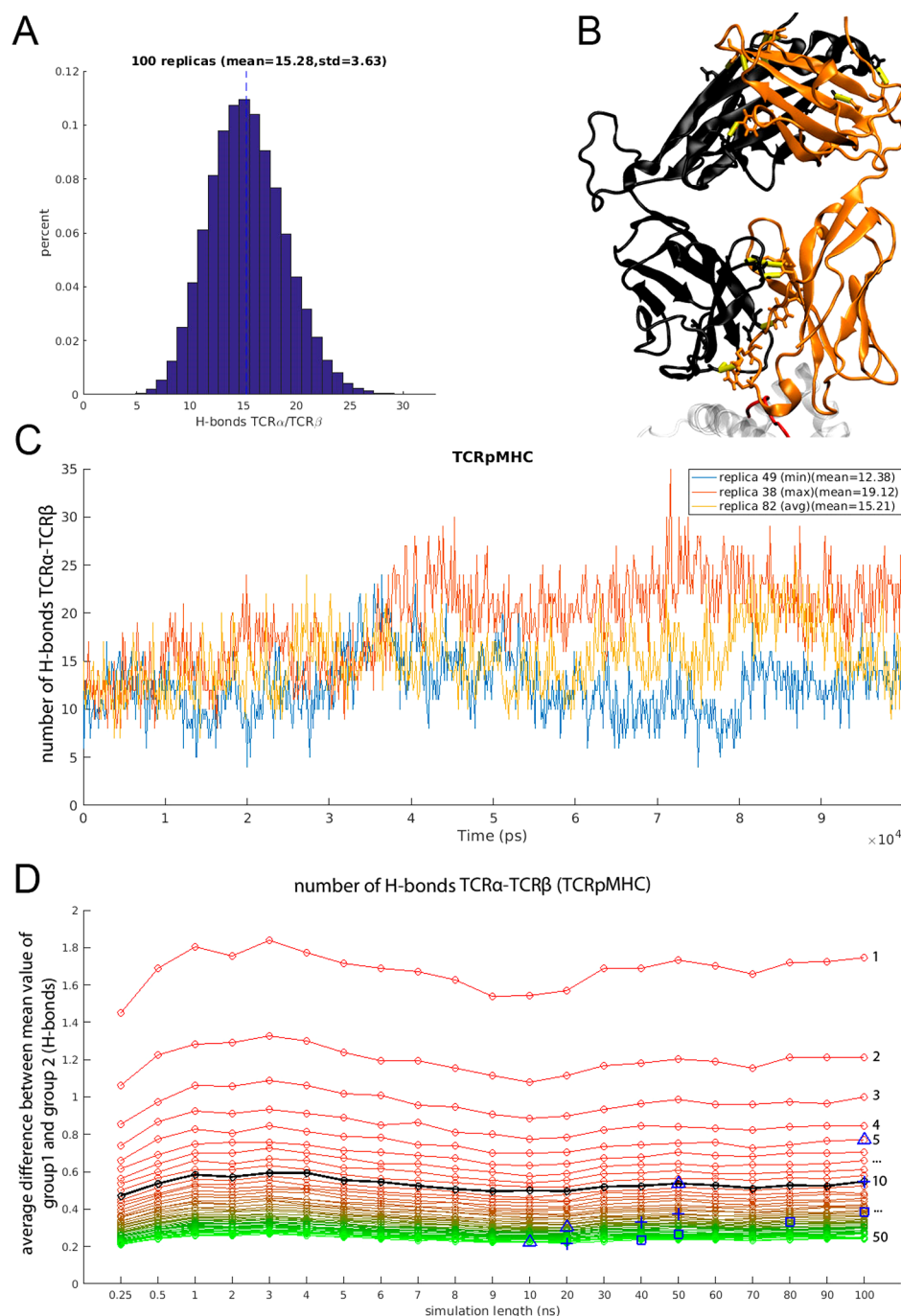


**Figure 4.** Distance between the peptide C-terminus (L9) and CDR3 $\beta$ . (A) Histogram of the distance over all time frames of all 100 replicas. (B) Graphical illustration of the distance between L9 (green) and CDR3 $\beta$  (yellow) as part of the MHC (transparent white), peptide (red), TCR  $\alpha$ -chain (orange), and TCR  $\beta$ -chain (black). (C) Distance over time (100 ns) shown for three example replicas. Replica 58 has on average over all time frames the lowest distance, while replica 25 has the highest. Replica 76 has an average distance very similar to the overall mean of all 100 replicas (13.00 vs 12.99). (D) Mean difference in distance between groups of replicas in relation to simulation time. For readability, the x-axis scale is not equally distributed. The analysis is shown for one replica (reddest line at the top) to 50 replicas (greenest line at the bottom). The lines change their color gradually from red to green representing the use of one to 50 replicas. The thick black line represents 10 replicas. Examples of using fixed amounts of computation time are indicated by symbols: blue triangles mark a total computational investment of 500 ns, blue pluses 1000 ns, and blue squares 2000 ns. The same plot with error bars for selected replica numbers is shown in [Appendix Figure 6](#).

for 10 replicas (black line in [Figure 5D](#)) and 85.7% for 50 replicas, but using 10 replicas already achieves 79.6% of the

improvement of using 50 replicas at 20% of the computational cost. The average standard deviation (over all 5000 permutations





**Figure 5.** Total number of H-bonds between the TCR  $\alpha$ - and  $\beta$ -chains based on TCRpMHC replicas. (A) Histogram of the number of H-bonds over all time frames of all 100 replicas. (B) Graphical illustration of the H-bond interface between the two TCR chains. The H-bonds are illustrated as yellow cylinders and the donor and acceptor residues as licorice representation. (C) Number of H-bonds over time (100 ns) shown for three example replicas. Replica 49 has on average over all time frames the lowest number of H-bonds, while replica 38 has the highest. Replica 82 has an average number of H-bonds highly similar to the overall mean of all 100 replicas (15.21 vs 15.28). A 3D illustration of the last simulation frames of the most extreme examples is provided in [Appendix Figure 7](#). (D) Mean difference in H-bond number between groups of replicas in relation to simulation time. For readability, the x-axis scale is not equally distributed. The analysis is shown for one replica (reddest line at the top) to 50 replicas (greenest line at the bottom). The lines change their color gradually from red to green representing the use of one to 50 replicas. The thick black line represents 10 replicas. Examples of using fixed amounts of computation time are indicated by symbols: blue triangles mark a total computational investment of 500 ns, blue pluses 1000 ns, and blue squares 2000 ns. The same plot with error bars for selected replica numbers is shown in [Appendix Figure 8](#).

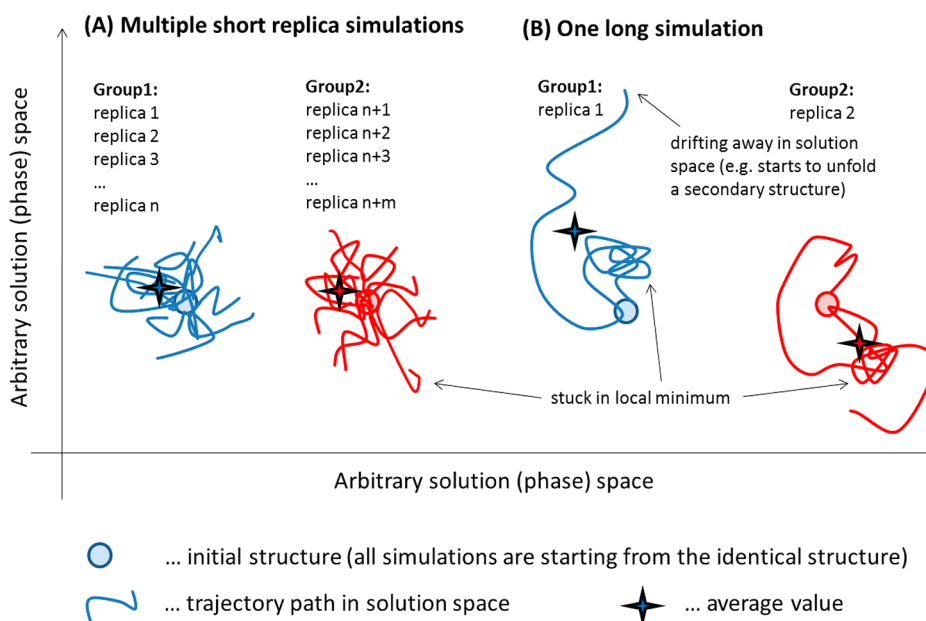
and 21 time cuts) of the replica group difference in H-bonds drops considerably with increasing number of replicas from 1.28 (one replica) to 0.57 (five replicas), 0.40 (10 replicas), 0.25 (25 replicas), and 0.18 (50 replicas) ([Appendix Figure 8](#)).

Once again, using a larger number of shorter replicas yields better results (blue markers in [Figure 5D](#)). For example, using

five replicas of 100 ns yields a mean difference of 0.77 H-bonds between groups, while using 10 replicas of 50 ns yields a mean difference of 0.54 H-bonds.

The same type of reliability analysis based on group differences is shown in [Appendix Figure 9](#) for other descriptors: radii of gyration of CDR3 $\alpha$  and CDR3 $\beta$ , backbone root-mean-square





**Figure 6.** Schematic explanation why the averages of multiple replicas tend to agree more than single longer simulations. (A) Even if a few replicas get stuck in local minima, enough other replicas continue to sample the remaining solution space. Independent of how the replicas are split between groups 1 and 2, the difference in average descriptors (e.g., H-bonds) is relatively low. (B) The difference in average descriptors of two single (long) simulations can be unexpectedly high, as simulations can get stuck in a local minimum for a considerable time or drift away in solution space. Therefore, multiple replicas are essential for reliable results.

deviation, distance between CDR3 $\alpha$  and CDR3 $\beta$ , solvent-accessible surface area (SASA) of the peptide within the TCRpMHC, overall SASA of TCRpMHC, and the numbers of peptide/MHC, TCR/MHC, TCR/peptideMHC, and TCR/peptide H-bonds.

## DISCUSSION

Reproducibility is a central theme in science.<sup>32</sup> The use of replicas to increase reproducibility and reliability is a standard method in many areas of research, including clinical trials,<sup>33</sup> gene expression analysis,<sup>34</sup> and experimental design.<sup>35,36</sup>

The law of large numbers<sup>37</sup> states that performing an experiment multiple times will bring the average result closer to the true outcome. The variance of the mean of  $n$  replicas is  $n$  times smaller than the variance of the variable measured in each independent replica.<sup>36</sup>

However, in the context of molecular simulations, performing replicas is computationally expensive. It is tempting to perform one longer simulation, and for many readers a paper that describes a single 2000 ns simulation is more impressive than a paper describing 20 replicas of 100 ns each. It is therefore perhaps not surprising that replicas are not regularly used in current molecular (dynamics) simulation studies (e.g., refs 38–44). The results of the analysis in this paper suggest that in terms of reproducibility and reliability, multiple replicas should be preferred over single long simulations. These findings agree with observations made in molecular docking approaches.<sup>45,46</sup>

Another reason for the frequent use of single simulations might be that even single long simulations struggle to reach biologically relevant time scales. For example, although a 100 ns TCRpMHC simulation is the current state of the art for this type of system,<sup>25</sup> it is unlikely that such a simulation will provide insights into global rearrangements. Our results suggest that after 100 ns this 827 amino acid system is in a similar phase of sampling as our 10 amino acid peptide system after 5 ns

(compare Figure 4D with the 0 to 5 ns portion of Figure 2D). In both cases, the local neighborhood has been explored but no fundamental changes have occurred. It is likely that the curves for the TCRpMHC simulations in Figures 4 and 5 would form peaks similar to those in Figures 2 and 3 if sufficiently long simulations were carried out. How long the simulations of TCRpMHC complexes would need to be extended to show global rearrangements similar to those for our peptide system is hard to predict. Most likely the time scale would be far beyond the microsecond scale, as even 1000 ns simulations of TCRpMHC still do not show global changes.<sup>29,47</sup>

The length of the simulations (once they have reached a minimum length and left local solution space) is not as important for reliability as the number of replicas. This agrees with previous studies showing that very short simulations of only few nanoseconds can effectively explore the local neighborhood around X-ray structures, leading to high correlations between predicted and experimental binding affinities using 50 replicas of only 4 ns each.<sup>47,48</sup> It could be argued that better agreement between multiple shorter simulations is not surprising because the simulations have the same starting configuration and are likely to have a similar start of the trajectory. While this might be partly true, it can be seen from the peptide simulations (Figure 2) that even three 1000 ns simulations against three 1000 ns simulations on average show better agreement than one 3000 ns simulation against one 3000 ns simulation. It is unlikely that this effect is due to failure of the 1000 ns simulation to leave the initial local solution space. A possible explanation why long simulations do not necessarily agree with each other is that they can get trapped in local minima for a considerable amount of time or drift away in solution space by unfolding structure elements that should not be unfolded (Figure 6). This explanation model agrees with previous studies showing that long simulations can get trapped in local minima.<sup>49,50</sup>

For shorter simulations (e.g., 100 ns) of larger systems (e.g., 800 amino acids), one might expect differences between identical simulations, but the extent to which we observed this was surprising. We found an average difference in the number of intra-TCR-chain H-bonds between replica simulations of  $1.75 \pm 1.26$  (Figure 5 and Appendix Figure 8), while the total average number of H-bonds is about 15.28 (11.5% difference or 19.7% including the upper boundary of the standard deviation). This means that if a study based on a single 100 ns simulation finds an increase/decrease by about three H-bonds (20% of the total H-bonds), there is no strong reason to believe that this increase/decrease is true.

This uncertainty can be brought down considerably by using five replicas of the same length ( $0.77 \pm 0.58$  H-bonds; 5.0% or 8.8%), 10 replicas ( $0.55 \pm 0.41$ ; 3.6% or 6.3%), 25 replicas ( $0.34 \pm 0.26$ ; 2.2% or 3.9%), or 50 replicas ( $0.24 \pm 0.18$ ; 1.6% or 2.7%), depending on the level of precision sought. If it is acceptable to be wrong by about one H-bond then five replicas will probably be sufficient; if an error of only half a H-bond is acceptable, 25 replicas might be the better choice. In our own experience, multiple simulations per comparison group have been shown to be beneficial for the reproducibility of a wide range of applications, such as H-bond networks<sup>51</sup> and free energy calculations.<sup>47</sup>

Even in very different approaches, such as hierarchical Monte Carlo simulations, the challenge of non-reproducible trajectories persists (see Figure 5 in ref 29). In previous work,<sup>29</sup> we used the Mosaics package,<sup>54</sup> which has been successfully applied before for a wide variety of challenges, including receptor signaling,<sup>55</sup> prediction of nucleosome occupancy,<sup>56</sup> empty MHC binding groove dynamics,<sup>57</sup> energy contributions of epigenetic DNA marks,<sup>58</sup> RNA mutations,<sup>4</sup> and macromolecule refinement.<sup>59</sup> We employed a coarse-grained model based on a three-point representation of each amino acid ( $\alpha$ -carbon, carbonyl oxygen, and side-chain centroid) in combination with a knowledge-based potential.<sup>52</sup> Movements were based on hierarchically grouped regions to restrict unnecessary degrees of freedom while allowing essential degrees. Potential chain breaks were resolved using a stochastic chain closure algorithm.<sup>53</sup> We used repeated simulated annealing cycles to allow for a rapid local and global search for energetically favorable states. Monte Carlo approaches like this are able to sample a larger part of the free energy landscape, but replicas still differ considerably from each other, emphasizing the need for multiple replicas in molecular simulation approaches.

Another important question involves the minimum simulation length needed in order to avoid missing potentially important transitions? For our small peptide system we can approximate an answer on the basis of Figure 2D: once replicas have left local sampling after about 20 ns, we become aware of the undersampling problem as reflected in the peak at around 100 ns. After this peak the difference lines start to flatten out, and we find the area around 1000 ns to be a good recommendation for the minimum simulation length of the peptide system. However, this does not mean that shorter simulations are not useful in general, since free energy calculations, for example, can reproduce experimental data very well using a large number of very short simulations.<sup>47,48</sup>

The actual choice between simulation length and number of replicas (given a fixed amount of available computational resources) will always depend on the question asked and the classical statistical dilemma of whether *false negative* results (i.e., many short simulations do not see relevant motion because they are too short) or *false positive* results (i.e., one or few very long

simulation show interesting results, but these results are outliers, and the average over many very long simulations would have shown a different picture, as discussed in the Introduction) are more acceptable.

We conclude that opting for multiple replicas over single molecular simulations allows for more reproducible results and thereby increases the reliability of molecular simulation papers. We strongly recommend that scientists perform multiple replicas before drawing conclusions from molecular simulations.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00391.

Movie of peptide replica 18 (MPG)

Appendix Figures 1–9 (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: science.bernhard.knapp@gmail.com.

### ORCID

Bernhard Knapp: 0000-0002-5714-7105

Charlotte M. Deane: 0000-0003-1388-2252

### Author Contributions

Conceived and designed the experiments: B.K., C.M.D. Performed the experiments: B.K. Analyzed the data: B.K., L.O. Provided computational resources: C.M.D. Wrote the paper: B.K. Revised the paper critically: C.M.D., L.O.

### Funding

This work was funded by the 2020 Science Programme (UK Engineering and Physical Sciences Research Council (EPSRC) Cross-Discipline Interface Programme, EP/I017909/1), EPSRC and Medical Research Council Systems (MRC) Approaches to Biomedical Science Centre for Doctoral Training (EP/L016044/1), and a Ministerio de Economía, Industria y Competitividad (MINECO), a la Agencia Estatal de Investigación (AEI) y al Fondo Europeo de Desarrollo Regional (FEDER) (BIO2017-83650-P). The authors acknowledge the use of the Oxford Advanced Research Computing (ARC) facility in carrying out this work.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank an anonymous reviewer for providing a more detailed conceptual background that we have incorporated into the Introduction and Discussion.

## ■ REFERENCES

- (1) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, 9 (9), 646–652.
- (2) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, 334 (6055), 517–520.
- (3) Knapp, B.; Dunbar, J.; Deane, C. M. Large Scale Characterization of the LC13 TCR and HLA-B8 Structural Landscape in Reaction to 172 Altered Peptide Ligands: A Molecular Dynamics Simulation Study. *PLoS Comput. Biol.* **2014**, 10 (8), e1003748.
- (4) Krawczyk, K.; Sim, A. Y.; Knapp, B.; Deane, C. M.; Minary, P. Tertiary Element Interaction in HIV-1 TAR. *J. Chem. Inf. Model.* **2016**, 56 (9), 1746–1754.
- (5) Perez, A.; Luque, F. J.; Orozco, M. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* **2012**, 45 (2), 196–205.

- (6) Wan, S.; Flower, D. R.; Coveney, P. V. Toward an atomistic understanding of the immune synapse: Large-scale molecular dynamics simulation of a membrane-embedded TCR-pMHC-CD4 complex. *Mol. Immunol.* **2008**, *45* (5), 1221–1230.
- (7) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497* (7451), 643–646.
- (8) Abraham, M. Reproducibility. <http://www.gromacs.org/Documentation/Terminology/Reproducibility> (accessed April 25, 2018).
- (9) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (10) Grossfield, A.; Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* **2009**, *5*, 23–48.
- (11) van der Merwe, P. A.; Dushek, O. Mechanisms for T cell receptor triggering. *Nat. Rev. Immunol.* **2011**, *11* (1), 47–55.
- (12) Kjer-Nielsen, L.; Clements, C. S.; Purcell, A. W.; Brooks, A. G.; Whisstock, J. C.; Burrows, S. R.; McCluskey, J.; Rossjohn, J. A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* **2003**, *18* (1), 53–64.
- (13) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., III Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850* (5), 1041–1058.
- (14) Musiani, F.; Rossetti, G.; Capece, L.; Gerger, T. M.; Micheletti, C.; Varani, G.; Carloni, P. Molecular dynamics simulations identify time scale of conformational changes responsible for conformational selection in molecular recognition of HIV-1 transactivation responsive RNA. *J. Am. Chem. Soc.* **2014**, *136* (44), 15631–15637.
- (15) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids. Res.* **2000**, *28* (1), 235–242.
- (16) Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K. Crystal structure of a ten-amino acid protein. *J. Am. Chem. Soc.* **2008**, *130* (46), 15327–15331.
- (17) Knapp, B.; Dunbar, J.; Alcalá, M.; Deane, C. M. Variable Regions of Antibodies and T-cell Receptors may not be Sufficient in Molecular Simulations Investigating Binding. *J. Chem. Theory Comput.* **2017**, *13* (7), 3097–3105.
- (18) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Struct., Funct., Genet.* **2009**, *77* (4), 778–795.
- (19) Knapp, B.; Giczi, V.; Ribarics, R.; Schreiner, W. PeptX: Using Genetic Algorithms to optimize peptides for MHC binding. *BMC Bioinf.* **2011**, *12*, 241.
- (20) Knapp, B.; Omasits, U.; Schreiner, W. Side chain substitution benchmark for peptide/MHC interaction. *Protein Sci.* **2008**, *17* (6), 977–982.
- (21) Knapp, B.; Omasits, U.; Frantal, S.; Schreiner, W. A critical cross-validation of high throughput structural binding prediction methods for pMHC. *J. Comput.-Aided Mol. Des.* **2009**, *23* (5), 301–307.
- (22) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
- (23) Hess, B.; Kutzner, C.; vanderSpoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (24) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets S3A5 and S3A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676.
- (25) Knapp, B.; Demharter, S.; Esmailbeiki, R.; Deane, C. M. Current Status and Future Challenges in T-cell receptor/peptide/MHC Molecular Dynamics Simulations. *Briefings Bioinf.* **2015**, *16* (6), 1035–1044.
- (26) Dien, H.; Deane, C. M.; Knapp, B. Gro2mat: A package to efficiently read Gromacs output in Matlab. *J. Comput. Chem.* **2014**, *35* (20), 1528–1531.
- (27) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38.
- (28) Knapp, B.; Lederer, N.; Omasits, U.; Schreiner, W. vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. *J. Comput. Chem.* **2010**, *31* (16), 2868–2873.
- (29) Knapp, B.; Demharter, S.; Deane, C. M.; Minary, P. Exploring peptide/MHC detachment processes using Hierarchical Natural Move Monte Carlo. *Bioinformatics* **2016**, *32* (2), 181–186.
- (30) Knapp, B.; Deane, C. M. T-Cell Receptor Binding Affects the Dynamics of the Peptide/MHC-I Complex. *J. Chem. Inf. Model.* **2016**, *56* (1), 46–53.
- (31) Knapp, B.; Omasits, U.; Bohle, B.; Maillere, B.; Ebner, C.; Schreiner, W.; Jahn-Schmid, B. 3-Layer-based analysis of peptide-MHC-interaction: in silico prediction, peptide binding affinity and T cell activation in a relevant allergen-specific model. *Mol. Immunol.* **2009**, *46* (8–9), 1839–1844.
- (32) Ioannidis, J. P. Why most published research findings are false. *PLoS. Med.* **2005**, *2* (8), e124.
- (33) Rohrig, B.; du Prel, J. B.; Wachtlin, D.; Kwicien, R.; Blettner, M. Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.* **2010**, *107* (31–32), 552–556.
- (34) Novak, J. P.; Sladek, R.; Hudson, T. J. Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* **2002**, *79* (1), 104–113.
- (35) Kshivsagar, A. Book reviews : Hinkelman K, Kempthorne O 1994: Design and analysis of experiments. Volume I: Introduction to experimental design. New York: John Wiley and Sons, Inc. 495 pp. US \$59.95; UK £49.50. ISBN 0 471 55178 3. *Stat. Methods Med. Res.* **1996**, *5* (1), 101–102.
- (36) Hogg, R. V.; McKean, J.; Craig, A. T. *Introduction to Mathematical Statistics*, 6th ed.; Pearson Education, 2005.
- (37) Bernoulli, J. *Ars Conjectandi*; Impensis Thurnisiorum, fratrum, 1713.
- (38) Mirza, M. U.; Rafique, S.; Ali, A.; Munir, M.; Ikram, N.; Manan, A.; Salo-Ahen, O. M.; Idrees, M. Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Sci. Rep.* **2016**, *6*, 37313.
- (39) Yang, X.; Wang, Z.; Xiang, Z.; Li, D.; Hu, Z.; Cui, W.; Geng, L.; Fang, Q. Peptide probes derived from pertuzumab by molecular dynamics modeling for HER2 positive tumor imaging. *PLoS Comput. Biol.* **2017**, *13* (4), e1005441.
- (40) Zhang, Y.; Niu, Y.; Tian, J.; Liu, X.; Yao, X.; Liu, H. The glycan-mediated mechanism on the interactions of gp120 with CD4 and antibody: insights from molecular dynamics simulation. *Chem. Biol. Drug Des.* **2017**, *90* (6), 1237–1246.
- (41) Das, J.; Aleksandrov, A. A.; Cui, L.; He, L.; Riordan, J. R.; Dokholyan, N. V. Transmembrane helical interactions in the CFTR channel pore. *PLoS Comput. Biol.* **2017**, *13* (6), e1005594.
- (42) Lugsanangarm, K.; Nueangaudom, A.; Pianwanit, S.; Kokpol, S.; Nunthaboot, N.; Tanaka, F.; Taniguchi, S.; Chosrowjan, H. Dynamics of the protein structure of T169S pyranose 2-oxidase in solution: Molecular dynamics simulation. *Proteins: Struct., Funct., Bioinf.* **2017**, *85* (10), 1913–1924.
- (43) Yokoyama, M.; Fujisaki, S.; Shirakura, M.; Watanabe, S.; Odagiri, T.; Ito, K.; Sato, H. Molecular Dynamics Simulation of the Influenza A(H3N2) Hemagglutinin Trimer Reveals the Structural Basis for Adaptive Evolution of the Recent Epidemic Clade 3C.2a. *Front. Microbiol.* **2017**, *8*, 584.
- (44) Knapp, B.; Fischer, G.; Van Hemelen, D.; Fae, I.; Maillere, B.; Ebner, C.; Schreiner, W.; Bohle, B.; Jahn-Schmid, B. Association of HLA-DR1 with the allergic response to the major mugwort pollen allergen: molecular background. *BMC Immunol.* **2012**, *13* (1), 43.
- (45) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4:



Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.

(46) Morris, G. M. Personal communication, 2018.

(47) Wan, S.; Knapp, B.; Wright, D.; Deane, C.; 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.* **2015**, *11* (7), 3346–3356.

(48) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V. Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *J. Chem. Theory Comput.* **2014**, *10* (3), 1228–1241.

(49) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E., III Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10* (1), 492–499.

(50) Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. ORAC: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J. Comput. Chem.* **2010**, *31* (5), 1106–1116.

(51) Zhang, H.; Lim, H. S.; Knapp, B.; Deane, C. M.; Aleksic, M.; Dushek, O.; van der Merwe, P. A. The contribution of major histocompatibility complex contacts to the affinity and kinetics of T cell receptor binding. *Sci. Rep.* **2016**, *6*, 35326.

(52) Minary, P.; Levitt, M. Probing protein fold space with a simplified model. *J. Mol. Biol.* **2008**, *375* (4), 920–933.

(53) Minary, P.; Levitt, M. Conformational optimization with natural degrees of freedom: a novel stochastic chain closure algorithm. *J. Comput. Biol.* **2010**, *17* (8), 993–1010.

(54) Sim, A. Y.; Levitt, M.; Minary, P. Modeling and design by hierarchical natural moves. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (8), 2890–2895.

(55) Moraga, I.; Wernig, G.; Wilmes, S.; Gryshkova, V.; Richter, C. P.; Hong, W. J.; Sinha, R.; Guo, F.; Fabionar, H.; Wehrman, T. S.; Krutzik, P.; Demharther, S.; Plo, I.; Weissman, I. L.; Minary, P.; Majeti, R.; Constantinescu, S. N.; Piehler, J.; Garcia, K. C. Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands. *Cell* **2015**, *160* (6), 1196–1208.

(56) Minary, P.; Levitt, M. Training-free atomistic prediction of nucleosome occupancy. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (17), 6293–6298.

(57) Demharther, S.; Knapp, B.; Deane, C. M.; Minary, P. Modeling Functional Motions of Biological Systems by Customized Natural Moves. *Biophys. J.* **2016**, *111* (4), 710–721.

(58) Krawczyk, K.; Demharther, S.; Knapp, B.; Deane, C. M.; Minary, P. In silico structural modeling of multiple epigenetic marks on DNA. *Bioinformatics* **2018**, *34* (1), 41–48.

(59) Zhang, J.; Minary, P.; Levitt, M. Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (25), 9845–9850.