



PHD

Utilising Molecular Dynamics Simulations to Understand and Engineer T-Cell Receptors
Alternative Format Thesis

Crean, Rory

Award date:
2020

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Utilising Molecular Dynamics Simulations to Understand and Engineer T-Cell Receptors

Rory Maurice Crean

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2019



COPYRIGHT

Attention is drawn to the fact that copyright of this thesis/portfolio rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis/portfolio has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as licenced, permitted by law or with the consent of the author or other copyright owners, as applicable.

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has / has not been incorporated into a submission for another degree

A handwritten signature in black ink, appearing to be 'A', is centered within a grey rectangular box.

This thesis may be made available for consultation within
the University Library and may be photocopied or lent to other libraries
for the purposes of consultation.

A handwritten signature in black ink, appearing to be 'A', is centered within a grey rectangular box.

Acknowledgements

First, I would like to thank my supervisor Dr Chris Pudney, who has supported and guided me throughout my entire PhD, training me on all the different aspects of becoming a scientist. Arguably of even more importance are the catchphrases that are suitable for just about every occasion in life (“just say yes” and “Gucci” to name just two). And finally, for supporting me to arguably explore an area of science ever so slightly different to what my PhD proposal might have suggested I would be doing...

This leads me on perfectly to thank my co-supervisor, Dr Marc van der Kamp who happily took me under his wing and shared so much knowledge and enthusiasm for computational chemistry with me. Marc and his group have always been so welcoming and helpful, and even went as far as to make my “adoption” official by signing the papers to make me a University of Bristol student (not just for the extra computing resources of course...).

I must thank all the lab members of both past (1.28 and 1.33) and present (0.39 and 0.40) for advice, support and friendship. A special shout out must go to my friend, collaborator and PhD boss Dr Hannah Jones, who not only taught me a lot of “Nice!” and “interesting sminteresting” science but also supported and encouraged me throughout my PhD.

I would also like to thank the exploratory research group at Immunocore for being so welcoming of someone with such a limited grasp of immunology (I didn’t even know my CD4 from my CD8). A special thanks must go to my Immunocore supervisor Dr David Cole for sharing his passion and knowledge of immunology and ensuring every meeting would be filled with laughter.

I would also like to thank Prof Shina Lynn Kamerlin and the Kamerlin Lab group for a wonderful and intense learning experience (with the exception of the Swedish language, varför? varför inte!), and the opportunity to come back and learn even more.

Last but certainly not least, I must thank my family and friends for all their support, humour and for offering escapes from the at times all-consuming PhD. In particular, I must thank my Mum and Dad because I am so very lucky to have such great parents that have never wavered in wanting the best for me and doing all they can to support me.

Table of Contents

Acknowledgements	3
List of Figures.....	12
List of Tables	22
List of Equations	23
Abstract.....	25
Chapter 1: Introduction and Background.....	26
1.1 The Human Immune System.....	26
1.2 Antigen Binding by The Adaptive Immune system.....	27
1.3 Antigen Presentation to T-Cell Receptors.....	28
1.4 Structural Characterisation of the pHLA Complex.....	29
1.5 How do Naturally Occurring TCRs Achieve Specificity?	31
1.6 Structural Features of the TCR-pHLA Complex	32
1.7 The Application of T-cells and TCRs as Therapeutics.	34
1.7.1 The Advantages and Challenges Associated with Soluble, High Affinity TCRs as Therapeutics	37
1.8 The Application of Computation to Understand and Engineer TCRs.	38
1.8.1 Methods to Calculate Protein-Protein Binding Free Energies.	38
1.8.2 Prior Attempts to Perform Rational TCR Engineering	39
1.9 Aims and Objectives	40
1.9.1 Objective 1: The role of Peptide Cargo in Allosterically Regulating the HLA	40
1.9.2 Objective 2: Engineering Principles for TCR-pHLA Specificity	40
1.9.3 Objective 3: Engineering Principles for TCR-pHLA Affinity.....	41
Chapter 2: Theory of Methodology.....	42
2.1. A Molecular Mechanics Force Field.....	42
2.2 Geometry Optimisation	44
2.3 Molecular Dynamics Simulations	45
2.3.1 Periodic Boundary Conditions	46
2.3.2 Pressure and Temperature Regulation	47
2.3.3 Dealing with Long Range Electrostatics Interactions	48
2.3.4 A General MD Simulation Protocol.....	49
2.4 Allosteric Analysis of MD Simulations	51
2.4.1 Dynamic Cross Correlation Matrices (DCCM)	51

2.4.2 Community Network Analysis (CNA).....	51
2.5 Molecular Mechanics Poisson-Boltzmann/Generalised Born Surface Area Calculations	52
2.5.1 Theoretical basis.	52
2.5.2 Calculation of Internal Energy, Electrostatic and van der Waals Terms	53
2.5.3 Calculation of the Solvation Free Energy	53
2.5.4 Calculation of Solute Entropy.....	54
2.5.5 Practical Considerations When Performing MMPB/GBSA Calculations	55
Chapter 3: Peptide Dependant Allosteric Regulation of the pHLA Complex.....	57
Peptide cargo tunes a network of correlated motions in human leukocyte antigens...	59
Abstract.....	60
Introduction	60
Results and Discussion	62
<i>Pressure-temperature matrices expose differing thermodynamic contributions to pHLA...</i>	62
<i>Molecular dynamics simulations identify both local and distal changes in flexibility for different peptide cargo.</i>	65
<i>Identification of correlated motions between the peptide and HLA.....</i>	70
<i>Peptide dependent tuning of the allosteric communication network.</i>	71
Conclusions	73
Materials and Methods	74
<i>Protein expression and refolding.</i>	74
<i>Pressure/temperature dependent fluorimetry</i>	74
<i>MD simulations.</i>	74
<i>Structure Equilibration Procedure</i>	75
<i>MD Trajectory Analysis</i>	76
Supporting Information for Peptide cargo tunes a network of correlated motions in human leukocyte antigens	78
Chapter 4: Design Principles for Engineering Highly Specific TCRs.....	84
Specificity of bispecific TCRs and antibodies targeting peptide-HLA.....	86
Abstract	87
Introduction	87
Results	89
<i>Structural analysis of pHLA-targeting reagents</i>	89
<i>Alanine scan analysis reveals distinct molecular recognition patterns</i>	93

<i>TCR-mimic antibodies bind to several commonly expressed self-peptides</i>	96
<i>Deep sequencing of peptides from randomised pHLA phage-libraries demonstrates the binding degeneracy of pHLA-targeting reagents</i>	98
<i>Molecular dynamics simulations reveal peptide-selectivity is associated with distinctive energetic modes of binding</i>	101
<i>Re-directed T cell killing of antigen-positive and negative cell lines using pHLA-targeting bispecifics</i>	105
<i>The NYBR1 TCR exhibits no cellular off-target reactivity and utilises a broad, peptide side-chain centric binding mode</i>	106
Discussion	109
Methods.....	112
<i>TCR engineering</i>	112
<i>Construct design, protein expression and purification</i>	112
<i>Analysis of on-target and off-target T cell reactivity via re-direction using anti-pHLA/anti-CD3 bispecific reagents</i>	113
<i>MAGPIX peptide screening</i>	113
<i>SPR Single cycle kinetic analysis</i>	114
<i>Generation of scHLA libraries and panning</i>	114
<i>Deep sequencing of pHLA libraries</i>	115
<i>X-ray crystallography</i>	115
<i>MD Simulations and MMPBSA calculations</i>	116
<i>Structure Equilibration Procedure.</i>	117
<i>Selection of explicit waters for MMPBSA calculations</i>	117
Supporting Information for: Specificity of bispecific TCRs and antibodies targeting peptide-HLA.....	119
Chapter 5: Drivers for High Affinity TCR-pHLA Interactions: Insights from Structure and Dynamics	128
Drivers for high affinity TCR-pHLA interactions: insights from structure and dynamics	130
Introduction	131
.....	133
Results and Discussion.....	133
Structural comparisons of WT and affinity matured TCRs show preservation of the WT binding mode.	133
Contact analysis of molecular dynamics simulations.	135
Reductions in flexibility are not required for affinity maturation.....	136

Energetic hot spots are preserved over the course of affinity maturation.	140
Mutations can directly or indirectly drive affinity enhancement.	141
Conclusions	146
Methods	146
<i>Molecular Dynamics Simulations</i>	146
<i>MD Trajectory and Crystal Structure Analysis</i>	147
<i>MMGBSA Methodology</i>	148
<i>MMGBSA Procedure</i>	148
Author Contributions	149
Acknowledgments.....	149
Conflict of interest	149
Supporting Information Methods	151
<i>MD Equilibration Procedure</i>	151
<i>RMS Fitting Procedure</i>	151
Supporting Information Tables	152
Supporting Information Figures	159
Chapter 6: Conclusions and Future Work	164
Appendix Chapter 1: Understanding the Role of Heat Capacity in Enzyme Catalysis	169
Abstract:	171
Results and discussion.....	174
Hydride transfer in ssGDH is rate determining.....	174
Heat capacity changes during enzyme reaction.	176
Change in $\Delta CP \ddagger$ with substrate isotopic substitutions.	177
Effect of increasing isotopic substitution on $\Delta CP \ddagger$	178
Conclusions	182
Materials and methods.....	183
ssGDH expression and purification.	183
Substrates and coenzymes.....	184
Enzyme assays.	184
Molecular dynamics simulations.	184
QM Cluster Model.	185
Supporting information	187

Supporting Information Materials and Methods	188
Appendix Chapter 2: Probing the Importance of the Environment in Regulating Functionally Important Enzyme Dynamics.....	191
Abstract:	193
Results and Discussion.....	195
Evidence for a resting-state semiquinone in MAO-B.	195
Influence of the membrane environment on MAO-B turnover.....	198
Computational evidence for a new entrance to the MAO-B active site mediated by the protein-membrane interaction.	199
Evidence of asymmetry in MAO-B from MD simulation.	201
Mutagenesis reveals long-range networks of motion are important in MAO-B.....	203
Conclusions	206
Materials and Methods	207
MAO-B expression and purification.....	207
Nanodisc preparation.	208
Enzyme assays.	208
Redox assays.....	208
EPR.	208
Computational materials and methods.....	208
Supporting information	211
Supporting Materials and Methods	212
FIRST calculations.....	212
Molecular Dynamics (MD) Simulations.....	213
References	215

Abbreviations and symbols

APC – antigen presenting cell

BSASA – buried solvent accessible surface area

BZA – benzylamine

C α – alpha carbon.

CAR – chimeric antigen receptor

CG – conjugate gradient

CM – continuum model

CNA – community network analysis

DCCM – dynamic cross correlation matrix

EPR – electron paramagnetic resonance

ER – endoplasmic reticulum

FAD – flavin adenine dinucleotide

FEL – free energy landscape

GB – generalised Born

H-bonds – Hydrogen bonds

HA – High Affinity

HLA – human leukocyte antigen

IE – interaction entropy

IMMTACs – Immune mobilising monoclonal TCRs against cancer

KIE – kinetic isotope effect

KYN – kynuramine

MAO – monoamine oxidase

mAbs – monoclonal antibodies

MD – molecular dynamics

ML – machine learning

MMFF – molecular mechanics force field

MMGBSA – molecular mechanics generalised Born surface Area
MMPBSA – Molecular Mechanics Poisson Boltzmann Surface Area
NMA – normal mode analysis
NVE – (Constant) Number (N) volume (V) and Energy (E)
NVT – (Constant) Number (N) volume (V) and Temperature (T)
NPT – (Constant) Number (N) Pressure (P) and Temperature (T)
PB – Poisson Boltzmann
PBCs – periodic boundary conditions
PCA – principal component analysis
PES – Potential Energy Surface
pHLA – peptide-human leukocyte antigen
PME – particle mesh Ewald
pMHC – peptide-major Histocompatibility Complex
POPC – palmitoylcholinephosphatidylcholine
POPE – palmitoylcholinephosphatidylethanolamine
QHA – quasi harmonic analysis
QM – Quantum Mechanics
ROS – reactive oxygen species
RMSD – root-mean-square deviation
RMSF – root mean square fluctuation
RU – response unit
SASA – solvent accessible surface area
SD – steepest descent
SPR – surface plasmon resonance
TAPBPR – TAP-binding protein-related
TCR – T-cell receptor
TILs –tumour-infiltrating lymphocytes
vdWs – van der Waals

WT – wild-type

ΔC_p^\ddagger – heat capacity of catalysis

ΔG_{bind} – change in binding free energy

ΔH – change in enthalpy

ΔS – change in entropy

ΔV – change in activation volume

$\Delta\beta$ – change in compressibility

$\Delta\alpha$ – change in expansivity

$C_{i,j}$ – cross correlation between atoms i and j

E_{int} – total internal energy

E_{el} – total electrostatic energy

E_{vdw} – total van der Waals energy

E_{bond} – total bonding energy

E_{angle} – total angle energy

E_{dih} – total dihedral energy

E_{imp} – total improper torsions energy

G_{npol} – non-polar solvation free energy

G_{pol} – polar solvation free energy

G_{solv} – solvation free energy

k_b – Boltzmann constant

q_i – charge of atom i .

r_{ij} – distance between atoms i and j

v_i – velocity of atom i

$\sigma(r)$ – charge density

σ_{ij} – distance between atoms i and j at which the intermolecular energy is 0.

ε_{ij} – well depth

$\varepsilon(r)$ – position dependant dielectric constant

List of Figures

Figure 1: Diagrams of the structural configurations of antibodies and TCRs. The transmembrane region is indicated as optional in antibodies as they can be expressed as either soluble or membrane bound proteins.....27

Figure 2: Schematic representation of antigen processing, presentation and T-cell binding pathways for HLA Class I and II. In the HLA Class I pathway, the antigen is first degraded in the proteasome (1), then bound to a HLA molecule (if it can bind) to form a pHLA complex (2), before being presented on the cell surface for both TCR and CD8 co-receptor recognition (3). In HLA Class II, antigens from the extracellular milieu are internalised (1), degraded by the proteasome (2), then peptides which are able to bind to a HLA Class II molecule (3) are presented on the cell surface for both TCR and CD4 co-receptor recognition (4).....29

Figure 3: (A) Schematic diagrams of the HLA Class I and II complexes. All domains which make up both complexes are labelled, with domains that are part of the same protein coloured the same. The black dotted lines represent the location of the membrane. (B) Representative structure of HLA Class I molecule with peptide bound. (C) Top down view of the peptide binding site of HLA Class I. Residues which make up the A-pocket (peptide N-terminal binding site) and F-pocket (peptide C-terminal binding site) are coloured in green and orange respectively. The α -helices which form the groove are labelled on the figure. For panels B+C, protein structure taken from PDB 3UTQ.⁹⁸³⁰

Figure 4: Positive and negative thymic selection of T-cells by the Thymus. Naive T-cells are produced in the thymus and their affinity towards self-peptides is evaluated. If a peptide binds too strongly (scenario 1) or too weakly (scenario 3) the T-cell is subjected by apoptosis (cell death). Only in scenario 2 is the T-cell (and its corresponding TCR) allowed to mature.....32

Figure 5: (A) Exemplar structure of the TCR-pHLA complex, with the six CDR loops and different portions of the overall scaffold coloured differently for clarity. (B+C) Zoom in on the TCR-pHLA binding site from two different orientations (same colouring as in A). Protein structure taken from PDB: 2P5E.⁴⁵.....34

Figure 6: Schematic diagram of the various types of therapeutics available that utilise T-cells to kill diseased/cancerous cells. The dotted lines indicate binding between different molecules. In examples 1+2 the antigen presented is a peptide from a pHLA molecule, whilst in 3+4 it is a protein expressed on the surface of the cell. A description of each of the therapy type is provided in the text. This figure was adapted from one produced by Immunocore Ltd. Original figure is available at: <https://www.immunocore.com/technology>.....36

Figure 7: The common terms used in a molecular mechanics force field (MMFF). All terms are summed together to give the total (potential) energy of the system. For the first four terms kr , $k\theta$, $k\varphi$, and $k\omega$ are force constants dependant on the distance (r), angle (θ), dihedral (φ) and torsion angle (ω) respectively. For the van der Waals (vdW) term, ϵ_{ij} is the well depth (minimum potential energy value possible), σ_{ij} is the distance between atoms i and j at which the potential energy is 0, and r_{ij} is the distance between atoms i and j respectively. For the electrostatic term, q_i is the charge of atom i and ϵ_0 is the dielectric constant.....	43
Figure 7: The common terms used in a molecular mechanics force field (MMFF). All terms are summed together to give the total (potential) energy of the system. For the first four terms kr , $k\theta$, $k\varphi$, and $k\omega$ are force constants dependant on the distance (r), angle (θ), dihedral (φ) and torsion angle (ω) respectively. For the van der Waals (vdW) term, ϵ_{ij} is the well depth (minimum potential energy value possible), σ_{ij} is the distance between atoms i and j at which the potential energy is 0, and r_{ij} is the distance between atoms i and j respectively. For the electrostatic term, q_i is the charge of atom i and ϵ_0 is the dielectric constant.....	44
Figure 8: Exemplar Leonard-Jones Potential describing the strength of a vdWs interactions between two interacting particles. The locations of σ (distance at which the intermolecular potential energy is 0, i.e. the vdWs radius) and ϵ (well depth) are indicated on the figure. As can be seen from the graph, after a short distance of $\sim 6 \text{ \AA}$, the potential energy is already near 0. For the equation used to calculate the Leonard-Jones potential, see Figure 7	49
Figure 9: Schematic Representation of the thermodynamic cycle used in MMPB/GBSA calculations. (The top three molecules are depicted inside a blue background to indicate they are solvated, whilst the bottom three are depicted as being in the gas phase). The model system shown is a TCR-pHLA complex, with the pHLA shown as the receptor and the TCR shown as the ligand.....	53
Figure 10: (A) Structural representation of the 1E6 TCR-pHLA interaction with the inset showing a zoom in on the TCR-pHLA binding site. The conserved GPD motif in the peptide and TCR residues that interact with these residues are shown as sticks. For the 4 peptides that also interact with the TCR via their position 1 residue (X), the side-chain of this residue is also indicated with orange sticks. (B) Peptide sequences and their respective pHLA melting temperatures (T_m) as determined by CD spectroscopy (Reported previously, Cole <i>et al.</i> 2016).....	63
Figure 11: Combined p/T matrices for each pHLA studied. Solid spheres represent the experimental data, transformed to ΔG by (Eq 12). The coloured surfaces are the resulting fit of these values to (Eq 13). Each panel is labelled as the specific pHLA complex.	65

Figure 12: Differences in flexibility at the peptide binding groove for all 6 pHLA complexes investigated. (A-B). Change in C α RMSF (Average – pHLA) for the α 1 and α 2 domain (A) and peptide (B), meaning a positive Δ RMSF value indicates an increase in rigidity for that pHLA complexes residue relative to the average. (C) Δ RMSF values as shown in A-B colour mapped on the pHLA structure (HLA as cartoon, peptide as ball-and-stick), with blue indicating increasing rigidity, and red indicating increasing flexibility (again relative to the average RMSF value for that residue). Heat mapping is scaled from -0.5 – 0.5 Å for all complexes. The black dots towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample *t*-test (*p* < 0.05). A colour bar is also included for reference.67

Figure 13 Differences in flexibility at the peptide binding groove for all 6 pHLA complexes investigated. (A-B) Change in C α RMSF (Average – pHLA) for the α 3 domain (A) and β 2m (B), meaning a positive Δ RMSF value indicates an increase in rigidity for that pHLA complexes residue relative to the average. (C) Δ RMSF values as shown in A-B colour mapped on the pHLA structure, with blue indicating increasing rigidity, and red indicating increasing flexibility (again relative to the average RMSF value for that residue). Heat mapping is scaled from -0.5 – 0.5 Å for all complexes. The locations of the α 3 and β 2m solvent exposed loops are indicated throughout the figure. The black dots towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample *t*-test (*p* < 0.05). A colour bar is also included for reference.69

Figure 14: Dynamic cross correlation matrices (DCCMs) for all 6 pHLA complexes simulated. On the Y-axis is each residue of the peptide, which is plotted against all other residues (377 total) on the HLA. The matrices are colour mapped according to the degree of correlated motion between the two residues, with a value of +1 meaning perfectly correlated and –1 meaning perfectly anticorrelated motion. Complete DCCM plots of the pHLA against the pHLA are provided in **Supplementary Figure S3**.71

Figure 15: Peptide dependent tuning of the allosteric communication network. Community networks determined for all pHLA complexes studied. Networked communities are shown as coloured spheres, with the radii of the sphere indicating the number of residues within the community. Edges between the nodes/communities represent communication pathways between the nodes, with the thickness of the edge indicating the degree of correlation between the two communities (thicker = greater correlation). All pHLA complexes are shown from the same orientation, such that N-terminus of the peptide is in the foreground. The results here are provided in tabulated form in **Supplementary Table S2**.72

Figure 16: Structural analysis of pHLA-targeting reagents. Structures of the TCRs and TCR-mimics, in complex with pHLA, were analysed to determine the structural mechanism underpinning their binding characteristics. (A) TCRs (blue ribbon) or TCR-mimics (orange ribbon) binding to peptide (red sticks) and HLA (green ribbon) compared to the binding range employed by all published wild-type TCR structures (grey cartoon). Yellow arrows indicate unconventional binding modes. (B) Coloured as in A, top down view of TCR, or TCR-mimic binding to pHLA. Black circles represent the center of binding. Yellow arrows indicate unconventional binding modes. (C) Structural analyses of TCRs versus TCR-mimics binding to pHLA. Bonds were assessed using the program contact (CCP4), implementing a 3.4 Å cut-off for H-bonds and a 4 Å cut-off for Van der Waals interactions. Any peptide residue with at least one interaction with the TCR is documented, with peptide residues with >10 contacts shown in bold, and the residue making the most contacts underlined. Any HLA residue with >5 interactions with the TCR is documented, with HLA residues with >10 contacts shown in bold. Buried surface area (BSA) Å² was determined using ePISA. 91

Figure 17: Alanine scan analysis reveals distinct molecular recognition patterns. The contribution of peptide side chains to binding specificity was analysed using alanine scan mutagenesis (by surface plasmon resonance). Binding affinities of the TCRs and TCR-mimic antibodies were determined using single cycle kinetic analysis. Bar graphs show binding affinity as a % relative to the binding affinity to the index peptide. (A) A2-SLL affinity-enhanced TCRs, (B) A2-SLL TCR-mimics, (C) A1-EVD affinity-enhanced TCRs, (D) Hyb3, (E) A2-RMF affinity-enhanced TCRs, and (F) ESK-1..... 95

Figure 18: Deep sequencing of peptides from randomised pHLA phage-libraries demonstrates the binding degeneracy of pHLA-targeting reagents. Sequence logos (Icelogo software) and heat maps were generated from NGS sequencing of pan 3 data identifying the distribution of amino acid identities per position of the peptide selected by A2-SLL-reactive affinity-enhanced TCRs and TCR mimic antibodies. The abundance of an amino acid is shown by intensity of colour. Outlined boxes identify the amino acids of the cognate antigen SLL. Data shown is the average of two experimental repeats. (A) 1G4_α5β100 (B) 1G4_α5β51, (C) 1G4_α58β61, (D) 3M4E5, (E) 3M4E5_T2, and (F) 3M4E5_T3. 100

Figure 19: Molecular dynamics simulations reveal broad side chain contacts with the peptide drive specificity. Relative number of Hydrogen bonds (H-bonds) and Van der Waals (vdWs) interactions formed between either the main or side chain of each peptide residue to the TCR/TCR-mimic over the course of our MD simulations. Total side versus main chain ratios for H-bonds and vdWs interactions are shown, with the larger value (side or main for each category) scaled to 100 % (absolute values for all contacts are provided in **Supp Figure S2**). (A) 1G4_α58β61-A2-SLL, (B) MAG-IC3-A1-EVD, (C) WT1_α7β2-A2-RMF, (D) 3M4E5-A2-SLL, (E) Hyb3.3-A1-EAD, (F) ESK-1-A2-RMF. 102

Figure 20: Binding free energy decomposition analysis of TCR and TCR-mimic-pHLA interactions. Per-residue decomposition of the binding free energy obtained from our MMPBSA calculations, to identify energetic hotspots for each TCR or TCR-mimic interaction with cognate pHLA. A top down view of each pHLA is shown, with the peptide depicted as sticks and the HLA as a surface. Colour mapping of the decomposition results for each residue was performed across the entire binding interface and used to indicate which residues across this interface favour (blue) or disfavour (red) binding (with white indicating no preference). (Bar graphs for all decomposition results are provided in **Supp Figure S3**). (A) 1G4_α58β61-A2-SLL, (B) MAG-IC3-A1-EVD, (C) WT1_α7β2-A2-RMF, (D) 3M4E5-A2-SLL, (E) Hyb3.3-A1-EAD, (F) ESK-1-A2-RMF.104

Figure 21: Redirected T cell killing of antigen positive and negative cell lines using pHLA-targeting bispecifics. The activity of the ImmTAC molecules and the TCR-mimic-anti-CD3 fusions was tested against a range of antigen positive and antigen negative cell lines (tumour and healthy cells) using IncuCyte killing assays. Data are plotted using area under the curve analysis. (A) IMC-1G4_α58β61, 3M4E5_T2-anti-CD3 and 3M4E5_T3-anti-CD3 T cell re-direction against HLA-A*02:01+/NY-ESO-1+ (NCI-H1755) and HLA-A*02:01+/NY-ESO-1- (HEP-G2, Ren8 and HISMC) cells lines. (B) IMC-MAG-IC3 and Hyb3.3-anti-CD3 T cell re-direction against HLA-A*01:01+/MAGE-A3+ (HCC1428), HLA-A*01:01+/MAGEA1+ (HCC1428 and NCI-H1703) and HLA-A*01:01+/MAGE- (COLO205 and HISMC) cells lines.....106

Figure 22: The NYBR1 TCR exhibits no cellular off-target reactivity and utilises a broad, peptide side-chain centric binding mode. (A) The activity of IMC-NYBR1 was tested against a range of HLA-A*02:01+/NYBR1+ (CAMA1 and CAMA1 β2m) and HLA-A*02:01+/NYBR1- cell lines (MDA-MB-231, HA13, HAo14, HDMEC2, Ren8, CM12 HCC1419, NCI-H661, SNU475 and SNU398) using IFNγ ELISpot (bar graphs) and IncuCyte killing assays (area under the curve analysis) in two donors. (B) Above: Side on view of the structure of the NYBR1 TCR (blue ribbon) in complex with A2 (green ribbon) -SLS (red sticks). Below: Top down view of the NYBR1 TCR-A2-SLS interaction. Black circles represent the center of binding. The table shows a structural analysis of the NYBR1-A2-SLS complex. Bonds were assessed using the program contact (CCP4), implementing a 3.4Å cut-off for H-bonds and a 4Å cut-off for vdWs interactions. Buried surface area (BSA) Å² was determined using ePISA. (C) The contribution of peptide side chains to binding specificity was analysed using alanine scan mutagenesis (by surface plasmon resonance). Binding affinities of the NYBR1-A2-SLS interaction was determined using single cycle kinetic analysis. Bar graphs show binding affinity as a % relative to the binding affinity to the index peptide. (**Figure Legend Continued on the next page**).....108

Figure 23: (A) Exemplar structure of a TCR-pHLA complex, with a different portions of the TCR and pHLA labelled. (B) Zoom in on the binding TCR-pHLA showing the 6 CDR loops responsible for binding the pHLA. (C) Table of all TCR-pHLA complexes investigated in this study. Note the α and β -chain framework (FW α and FW β) are non-hypervariable loops that flank the CDR loops. (A complete list of all mutations made onto each CDR loop can be found in **Table S1**). 130

Figure 24: (A-D) Aligned structures of the binding interface for all four sets of TCR-pHLA complexes investigated: (A) – NY-ESO, (B) – MEL5, (C) – DMF5 and (D) – Tax A6. In all cases, the WT TCR is coloured in white. (E) Structural analyses of all TCR-pHLAs complexes under investigation. The d_{TRangle} is a measure of the TCR variable domains orientational similarity for 2 different structures (see Methods).²⁵¹ For this, we compare the WT to each HA TCR. The docking angle (which describes the angle the TCR engages the pHLA) and d_{TRangle} were determined using the STCRDAB server. Buried solvent accessible surface area (BSASA) was determined using the LCPO algorithm within CPPTRAJ. Hydrogen bonds (H-bonds) are defined for donor-acceptor distances ≤ 3.5 Å and donor-hydrogen-acceptor angles 135 - 225° . Van der Waals (vdWs) interactions are considered for all heavy atoms within 4 Å of each other. 130

Figure 25: Total average number of Hydrogen bonds (A) and vdWs (B) contacts between the TCR and pHLA from 10 independent 100 ns long MD simulations for each system (using the time from 10-100 ns to measure contacts). Different TCRs are grouped into their own section, with all WT TCRs coloured in green and all affinity matured TCRs coloured purple. 130

Figure 26: Differences in flexibility between the affinity matured TCRs variable regions and their counterpart WT TCRs. (A-D) ΔRMSF values (HA variant RMSF–WT RMSF) for all NY-ESO TCRs, with panels A and B corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels C and D corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. (E-H) ΔRMSF values (HA variant RMSF–WT RMSF) for the three pairs of HA and WT TCRs, with panels E and F corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels G and H corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. A more negative ΔRMSF value indicates increased rigidity for the HA variant relative to the WT. Complete RMSF plots for all TCRs simulated are provided in **Figures S1-S4**. 130

Figure 27: Experimental vs computational $\Delta\Delta G_{\text{bind}}$ values obtained from our MMGBSA calculations for all TCR-pHLA systems studied. For the three pairs of TCRs with a single WT and affinity matured TCR, lines are drawn to guide the eye. Error bars plotted are the standard deviation obtained from the 25 replicas performed per complex (see Methods). 130

Figure 28: Changes in the energetic footprint between the WT and affinity matured TCRs. For all TCR-pHLA complexes, the HLA (top) and TCR (bottom) structures are plotted as surfaces with the peptide shown in both structures as sticks. All plots are colour mapped according to the MMGBSA per residue decomposition results, going from blue (favours binding) to white (no preference) to red (disfavours binding). Separate scaling is used for each of the 4 sets of TCRs. All pHLA and TCR structures are shown in the same orientation, such that the peptide N-terminus is left and the C-terminus right. Several mutations sites are indicated on the high affinity variants (purple labels: CDR α mutations; green labels: CDR β mutations).130

Figure 29: NY-ESO variants show largely additive energetic effects upon affinity maturation. **A.** Per-residue ΔG differences between the high affinity (HA) variants and WT-NY-ESO (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR are coloured blue if < -0.5 kcal mol $^{-1}$ (favourable for binding) and red if > 0.5 kcal mol $^{-1}$ (unfavourable for binding), with values in between coloured green. **(B–G)** Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-NY-ESO structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant. Figures are divided to focus on the different regions of the TCR subjected to affinity maturation (CDR2 α , CDR3 α , CDR2 β and CDR3 β), and subdivided when mutations are not consistent between HA TCRs. (c58c61/2 means both c58c61 and c58c62 are shown).130

Figure 30: Changes in Energetics at the TCR-pHLA Interface upon affinity maturation. **(A)** Per-residue ΔG differences between the three high affinity (HA) variants and their counterpart WT TCRs (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR pair are coloured blue if < -0.5 kcal mol $^{-1}$ (favourable for HA) and red if > 0.5 kcal mol $^{-1}$ (unfavourable for HA), with all values in-between coloured green. **(B–G)** Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-TCR structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant respectively. Figures are divided up to show the regions which show the major changes upon affinity maturation.130

Figure 31: **(A)** Exemplar structure of a TCR-pHLA complex, with a different portions of the TCR and pHLA labelled. **(B)** Zoom in on the binding TCR-pHLA showing the 6 CDR loops responsible for binding the pHLA. **(C)** Table of all TCR-pHLA complexes investigated in this study. Note the α and β -chain framework (FW α and FW β) are non-hypervariable loops that flank the CDR loops. (A complete list of all mutations made onto each CDR loop can be found in **Table S1**).....133

Figure 32: Differences in the average number of H-bonds, vdWs contacts and the BSASA from our MD simulations. All WT TCRs are coloured in green and all affinity matured TCRs are coloured in purple. Bars for the H-bonds and vdWs contacts are divided in two based on HLA-TCR interactions (darker colour and hashed bars) and peptide-TCR interactions (lighter colour). The totals obtained are from 10 independent 100 ns long MD simulations of each TCR-pHLA complex (using the last 90 ns of each simulation). Error bars plotted for **A** and **B** are the standard deviation of the averages from the 10 replicas. Error bars plotted for **C** are the standard deviation obtained from combining all snapshots from all replicas together..... 136

Figure 33: Differences in flexibility between the affinity matured TCRs variable regions and their counterpart WT TCRs. (A–D) Δ RMSF values (HA variant RMSF – WT RMSF) for all NY-ESO TCRs, with panels **A** and **B** corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels **C** and **D** corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. (E–H) Δ RMSF values (HA variant RMSF – WT RMSF) for the three pairs of HA and WT TCRs, with panels **E** and **F** corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels **G** and **H** corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. A more negative Δ RMSF value indicates increased rigidity for the HA variant relative to the WT. The points towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample *t*-test ($p < 0.05$). Crosses are used to indicate the locations of mutated regions of the TCRs (using the same colouring as line graph). Complete RMSF plots for all TCRs simulated are provided in **Figures S1-S4**..... 138

Figure 34: Experimental vs computational $\Delta\Delta G_{\text{bind}}$ values obtained from our MMGBSA calculations for all TCR-pHLA systems studied. For the three pairs of TCRs with a single WT and affinity matured TCR, lines are drawn to guide the eye. Error bars plotted are the standard deviation obtained from the 25 replicas performed per complex (see Methods)..... 140

Figure 35: NY-ESO variants show largely additive energetic effects upon affinity maturation. **A.** Per-residue ΔG differences between the high affinity (HA) variants and WT-NY-ESO (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR are coloured blue if < -0.5 kcal mol $^{-1}$ (favourable for binding) and red if > 0.5 kcal mol $^{-1}$ (unfavourable for binding), with values in between coloured green. **(B–G)** Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-NY-ESO structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant. Figures are divided to focus on the different regions of the TCR subjected to affinity maturation (CDR2 α , CDR3 α , CDR2 β and CDR3 β), and subdivided when mutations are not consistent between HA TCRs. (c58c61/2 means both c58c61 and c58c62 are shown)..... 142

Figure 36: Changes in Energetics at the TCR-pHLA Interface upon affinity maturation. (A) Per-residue ΔG differences between the three high affinity (HA) variants and their counterpart WT TCRs (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR pair are coloured blue if $< -0.5 \text{ kcal mol}^{-1}$ (favourable for HA) and red if $> 0.5 \text{ kcal mol}^{-1}$ (unfavourable for HA), with all values in-between coloured green. (B–G) Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-TCR structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant respectively. Figures are divided up to show the regions which show the major changes upon affinity maturation.145

Figure 37: A, Representative structures of glucose (green) and xylose (blue) from our MD simulations demonstrating they have the same binding interface with NADP^+ . **B,** Normalized histograms (bin width 0.1 \AA) of the hydrogen transfer distance of glucose and xylose from MD simulations of ssGDH. **C,** QM cluster model created of glucose in complex with NADP^+ , with asterisks indicating atoms fixed throughout the optimisation process. (D–G) Reaction mechanism obtained from the QM model, starting from the reactant (D), to the deprotonated reactive intermediate (E), the transition state (F), and finally the product. (G).175

Figure 38: The temperature-dependence of NADP^+ reduction by glucose (A) and xylose (B). Solid and dashed lines show the fitted to Eq 16 for the protiated and deuterated Glucose/xylose (D^1), respectively. (C), the resulting KIE extracted from the $\ln k_{\text{cat}}$ ($\ln \text{min}^{-1}$) in panels A and B. The solid line is the modelled KIE based on the parameters extracted from Eq 16 (Solid lines in panels A and B).176

Figure 39: Numerical model showing how the magnitude of the glucose 1° KIE versus temperature is affected by differences in the isotope effect on ΔCP^\ddagger , ΔH^\ddagger and ΔS^\ddagger values used for modelling given in Table 2.178

Figure 40: The effect of isotopic labelling on $\Delta\Delta CP^\ddagger$. (A) The temperature-dependence of k_{cat} for each isotopologue of glucose. Solid lines are fits to Eq 16. (B) Resulting KIEs extracted from the fits in panel A. C, correlation between the increase in molecular mass (through isotopic substitution) and the extracted magnitude of $\Delta\Delta CP^\ddagger$. The solid line is a simple rectangular hyperbola and is to aid the eye only to illustrate the trend. D, structures of each isotopologue used in the panel A.181

Figure 41: (A) The absorbance spectra of MAO-B after treatment with BZA over time. Inset, the effect on the 415 nm peak over time. (B) Fluorescence excitation/emission matrix resolves oxidized and semiquinone flavin states (highlighted in solid black boxes). Scale bar is relative intensity Conditions: 50 mM HEPES 0.5 % w/v Triton X-100, 20 °C. For absorption experiments 40 mM BZA, anaerobic conditions as Materials and Methods.196

Figure 42: X-band cw-EPR spectra of WT (top, black) and Y398F (bottom, black) MAO-B, with their respective fitted simulations (red). EPR microwave frequency = 9.3916 GHz (WT) & 9.3926 GHz (Y398F), microwave power = 0.2 mW, modulation amplitude = 0.5 mT, temperature = 16 K. 197

Figure 43: The temperature-dependence of MAO-B with BZA (**A**) and KYN (**B**) with reduced Triton and nanodisc environments, fit to the MMRT equation. Conditions, 50 mM HEPES 0.5 % w/v reduced Triton X-100 pH 7.5. Reduced Triton X-100: 1.5 mM BZA, 0.75 mM KYN. Nanodiscs: 3 mM BZA, 0.66 mM KYN. 200

Figure 44: MAO-B in POPC/POPE membrane environment. The substrate binding area is shown in light green and light blue (residues 80-210, 286-390) for each monomer. The active site area is shown in dark green and dark blue (residues 4-79, 211-285, and 391-452). The C-terminal tail and transmembrane helix are shown in red (residues 453-520). The binding site gating loop is shown in magenta (residues 99-112). Images **A** and **B** inset denote the two main entrances (or tunnels) to the binding site (denoted Entrance A and Entrance B throughout the manuscript), with a representative tunnel (navy spheres) shown for both entrances. Key residues which describe the location of the entrance are shown as sticks and labelled. Entrance B is newly identified here, Entrance A has been previously described.³⁴⁵ 202

Figure 45: The summed tunnel frequency from the substrate binding site of MAO-B to solvent, identified by Caver.³⁶³ BZA₀ – No BZA is bound in either monomer. BZA₁ – BZA is bound only in monomer 1. All are from triplicate MD runs. **B**, The average maximum bottleneck (Å) from the substrate binding site of MAO-B to the solvent, as identified by Caver. **C + D**, All tunnels identified by Caver over all three MD simulation repeats for BZA₀ (**C**) and BZA₁ (**D**). FAD and BZA (when present) are shown as yellow sticks. Tunnels are colored according to entrance/exit pathway used, with magenta and green representing Entrances A and B respectively. Blue tunnels indicate pathways which go through neither of the two main entrances described. 203

Figure 46: (**A–D**) Distance between the N5 (FAD) and CH2 group of BZA for wtMAO-B and all three enzyme variant simulations. A black dotted line indicates the start of each new trajectory (all runs performed in triplicate). Additional measurements for all BZA containing simulations can be found in **Figures S16** and **S17**. (**E+F**), Normalized histograms (bin width 0.25 Å) of principle principal component 2 (PC2) for all BZA1 simulations of the bound (**E**) and unoccupied monomers (**F**). (**G**) ‘Porcupine’ plot of PC2, with arrows indicating the direction of the PC2 eigenvector and arrow size indicating the magnitude of the corresponding eigenvalue, for all C α atoms with eigenvalues greater than 4 Å. The gating loop residues (99 – 112) are coloured in red, and the approximate location of the bilayer is indicated with a black dotted line. (A mobility plot of PC2 is provided in **Figure S18A**.) key for catalysis (see *Supporting Information Materials and Methods* and **Table S5**). 205

List of Tables

Table 1: Parameters extracted from surface fits shown in Figure 11 . T_m values previously published. ²³	65
Table 2: Data collection and refinement statistics for WT1_α7β2-A2-RMF and NYBR1-A2-SLS structures. One crystal was used for solving each structure. Figures in brackets refer to the highest resolution bin.	92
Table 3: MAGPIX analysis to investigate a panel of commonly expressed HLA-A*0201-restricted self-peptides. Peptides generating a signal above background (3 times median intensity of all bead regions bound to native helper-phage) were classified as positive binders, and binding expressed as a percentage of signal obtained from binding to index peptide in each case. Percentage binding for each interaction is reported as the average of several experimental repeats.	97
Table 4: MAGPIX analysis to investigate a panel of commonly expressed HLA-A*0101-restricted self-peptides. Peptides generating a signal above background (3 times median intensity of all bead regions bound to native helper-phage) were classified as positive binders, and binding expressed as a percentage of signal obtained from binding to index peptide in each case. Percentage binding for each interaction is reported as the average of several experimental repeats	98
Table 5: Structural analyses of all TCR-pHLAs complexes under investigation. The $d_{TRangle}$ is a measure of the TCR variable domains orientational similarity for 2 different structures (see Methods). ⁴¹ For this, we compare the WT to each HA TCR. The docking angle (which describes the angle the TCR engages the pHLA) and $d_{TRangle}$ were determined using the STCRDAB server. Buried solvent accessible surface area (BSASA) was determined using the LCPO algorithm within CPPTRAJ. Hydrogen bonds (H-bonds) are defined for donor-acceptor distances ≤ 3.5 Å and donor-hydrogen-acceptor angles 135-225°. Van der Waals (vdWs) interactions are considered for all heavy atoms within 4 Å of each other.	134
Table 6. Kinetic parameters extracted from fits of Eq 15 and Eq 16 to the temperature-dependence data shown in Figure 31A	179
Table 7. Spectral Parameters of EPR Data Extracted by Simulation and Fitting.	198

List of Equations

$$(1) \quad F(t) = ma = -\frac{\partial PE}{\partial(r)} = m \frac{\partial^2 r}{\partial t^2}$$

$$(2) \quad \frac{\partial^2 r}{\partial t^2} \approx \frac{r(t + \Delta t) + r(t - \Delta t) - 2r(t)}{\Delta t^2} = \frac{F(t)}{m}$$

$$(3) \quad r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{\Delta t^2 F(t)}{m}$$

$$(4) \quad KE = \frac{3}{2} Nk_b T = \left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle$$

$$(5) \quad P = \frac{1}{3V} \left\langle 3Nk_b T + \sum_{i=1}^N f_i r_i \right\rangle$$

$$(6) \quad C_{i,j} = \langle \Delta r_i \cdot \Delta r_j \rangle / (\langle \Delta r_i^2 \rangle \cdot \langle \Delta r_j^2 \rangle)^{1/2}$$

$$(7) \quad \Delta G_{bind} = \Delta E_{int} + \Delta E_{el} + \Delta E_{vdw} + \Delta G_{pol} + \Delta G_{npol} - T \Delta S$$

$$(8) \quad G_{pol} = \frac{1}{2} \int_0^s \sigma(r) \cdot V(r) dr$$

$$(9) \quad \sigma(r) = \frac{1}{4\pi} \left(1 - \frac{1}{\epsilon(r)} \right) \nabla V(r) \cdot n(r)$$

$$(10) \quad G_{pol} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon(r)} \right) \sum_i^{atoms} \sum_j^{atoms} \frac{q_i q_j}{\sqrt{R_{ij}^2 + r_i \cdot r_j \cdot e^{(-R_{ij}^2/4r_i \cdot r_j)}}$$

$$(11) \quad \frac{F_i}{\sum F} (p, T) = \frac{K(p, T)}{1 + K(p, T)}$$

$$(12) \quad \ln K = \frac{-\Delta G}{RT}$$

$$(13) \quad \Delta G_{P,T} = \Delta G_0 + \Delta V_0(P - P_0) + \Delta \alpha'^{(P-P_0)(T-T_0)} + \frac{\Delta \beta'}{2}(P - P_0)^2 - \Delta S_0(T - T_0) - \Delta C_P \left[T \left(\ln \left(\frac{T}{T_0} \right) - 1 \right) + T_0 \right]$$

$$(14) \quad d_{\text{TRangle}} = \sqrt{\sum (\theta_{i,WT} - \theta_{i,HA})^2}$$

$$(15) \quad k = \left(\frac{k_B T}{h} \right) e^{\frac{\Delta S^\ddagger}{R}} e^{-\frac{\Delta H^\ddagger}{RT}}$$

$$(16) \quad \ln k = \ln \frac{k_B T}{h} - \left[\frac{\Delta H_{T_R}^\ddagger + \Delta C_P^\ddagger(T - T_R)}{RT} \right] + \left[\frac{\Delta S_{T_R}^\ddagger + \Delta C_P^\ddagger(\ln T - \ln T_R)}{R} \right]$$

Abstract

Approximately 90% of all therapeutic targets in the human proteome operate solely inside the cell, making them unavailable for recognition by antibodies which instead bind antigens presented on the exterior of the cell surface. To target the 90%, the human immune system utilises a class of binding proteins known as T-cell receptors (TCRs). These TCRs recognise peptide fragments sourced from proteins produced inside the cell that have subsequently been degraded and transported to the cell surface by the human leukocyte antigen (HLA, pHLA with peptide bound). TCRs are membrane bound and attached to T-cells and use their six complementarity-determining region (CDR) loops to bind antigenic pHLA molecules (i.e. peptides that come from protein antigens). TCR binding to pHLA molecules induces an immune response from the T-cell, which ultimately leads to the antigen presenting cells death. The capability of TCRs to identify antigens which are not naturally expressed on the cell surface (unlike antibodies) has helped drive the development of a new class of therapeutics that consist of a soluble, bispecific TCR engineered to bind a specific antigenic pHLA for the treatment of various diseases (such as cancers and viral infections). Natural TCRs bind with characteristically poor affinities ($\sim\mu\text{M}$) and half-lives ($\sim\text{seconds}$), which are undesirable properties for a therapeutic. The CDR loops on TCRs are therefore normally subjected to affinity maturation to produce TCRs with affinities in $\sim\text{pM}$ range for their target pHLA. This does however carry a significant risk in terms of safety, as the very large majority of peptides presented by HLA molecules are sourced from endogenous (i.e. healthy) proteins and must not be bound by the TCR in order to avoid the production of an autoimmune response on the healthy cells.

These requirements for a highly specific TCR that binds with high affinity to its target pHLA is the primary motivation for this thesis, as herein, fundamental engineering principles for generating TCRs with these properties are determined and protocols to evaluate these properties are developed and demonstrated. This insight is obtained through combinations of structural analysis, molecular dynamics simulations and free energy calculations, providing an atomistic description of how this has occurred for several TCRs. Furthermore, we characterise how different peptide cargo can tune the molecular flexibility of the entire pHLA molecule, including regions distal from the HLA binding site. These findings suggest peptide dependant tuning of the HLA molecule may play a role in regulating the functional outcome of an immune response.

Ultimately, this work and the principles identified herein will aid in the rational design of high affinity and high specificity TCRs as therapeutics for various diseases.

Chapter 1:

Introduction and Background

The main thrust of this thesis focusses on using engineered T-cell Receptors (TCRs) as therapeutics, with a focus towards both gaining key insights into design principles for affinity and specificity, as well as characterising the allosteric regulation of the TCRs therapeutic target, the peptide-human leukocyte antigen (pHLA). As each results chapter in this thesis contains its own introduction, this section will provide a broader overview of the topics covered. Herein, I will first describe TCRs and pHLAs in their natural context (as vital parts of the immune system) before moving on to discuss the application of TCRs as therapeutics and the challenges associated with this.

1.1 The Human Immune System

All organisms contain an immune system which acts to protect the host against disease. In humans, the immune system is highly complex, with many mechanisms in place to protect against pathogens.¹ This system can be broken down into multiple “layers” of protection, the first of which can be described as simple physical barriers that prevent entry into the host such as the skin. Other examples of this layer include coating the airways with mucus to catch microorganisms as they are inhaled, or stomach acid, which provides a hostile environment (low pH) for most microorganisms. The next layer is the “innate immune system” which is found in all plants and animals, as well as most bacteria.² The innate immune response is fast acting and non-specific (in the sense that the response to different pathogens is the same).

The final layer of the immune system is the “adaptive (or acquired) immune system”.³ The adaptive immune system is significantly slower acting than the innate immune response, but highly specific towards the given pathogen (or pathogen infected cells). The adaptive immune system recognizes “non-self” (i.e. not natively produced) molecules such as proteins or lipids which are referred to as antigens. A key benefit of the adaptive immune system is in the generation of “memory cells”, which are produced during the first exposure to a given antigen. These memory cells can lie dormant in the body for several years unless they re-identify their target antigen from the host becoming re-infected. Upon exposure to the given antigen these memory cells can trigger an effective adaptive immune response significantly more rapidly when compared to the first exposure event. Interestingly, this “training” of the adaptive immune system to recognize and then remember a given antigen is the basic process by which vaccines work.⁴

1.2 Antigen Binding by The Adaptive Immune system

Two related classes of proteins are used by the adaptive immune system to identify and target antigens.³ The first are antibodies (also known as immunoglobulins), which are produced by B-cells as either membrane bound (protruding out of a given B-cell and into the extracellular surface) or as soluble proteins. Human antibodies adopt a “Y” shaped quaternary structure (see **Figure 1**), whereby there are two antigen binding domains located on the ends of the forks of the Y (known as the antigen-binding fragment regions, Fab regions), whilst the bottom of the Y (known as the fragment crystallizable region, Fc region) is used to communicate with other parts of the immune system. Antibodies bind antigens through six specialised loops, known as complementarity-determining regions (CDRs). These CDR loops are “hypervariable” in sequence composition which is essential for providing coverage against the wide variety of possible epitopes (antigen binding sites).³ The second class of targeting protein used by the adaptive immune system is the T-cell receptor (TCR), which are produced by T-cells as membrane bound proteins (see **Figure 1**). These proteins share common structural features with the Fab region of an antibody, including in the use of six CDR loops to bind their target antigens. TCRs consist of two different chains (α and β or a less common variant made of γ and δ chains) which both contribute equally to forming the antigen binding site by providing three CDR loops per chain to bind the epitope.

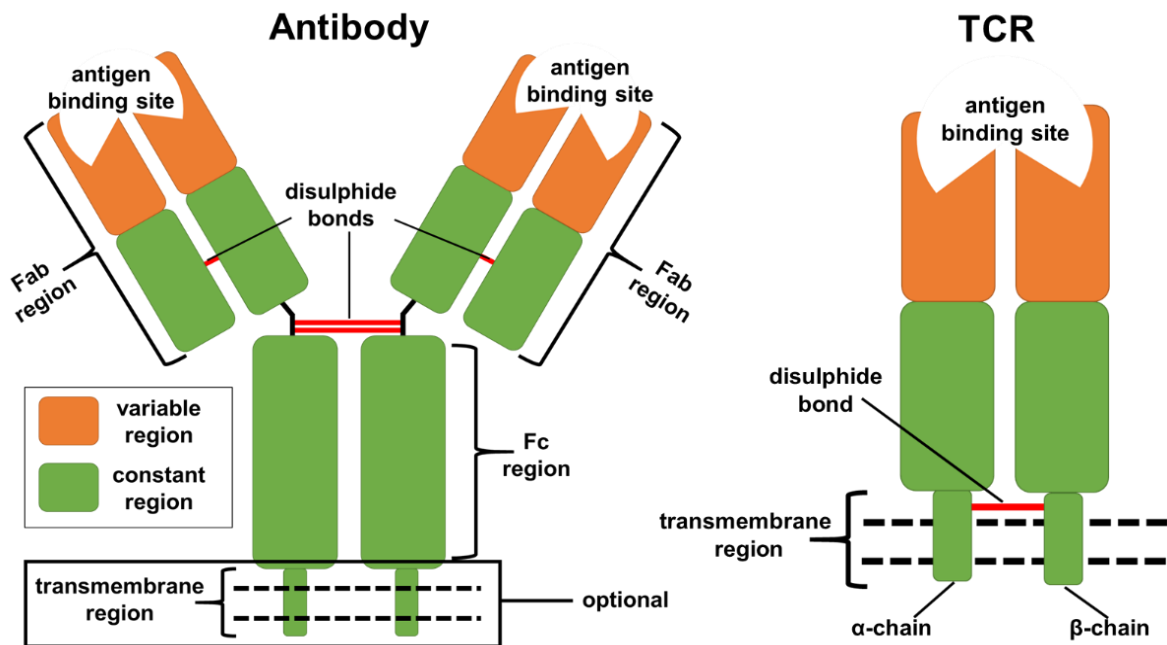


Figure 1: Diagrams of the structural configurations of antibodies and TCRs. The transmembrane region is indicated as optional in antibodies as they can be expressed as either soluble or membrane bound proteins.

A key difference between TCRs and antibodies is in the nature of the epitopes they target. Whilst antibodies can target a range of different molecular surfaces found on proteins (or other molecules like lipids and sugars), TCRs bind only a small fragment of an antigen. These small fragments are presented to the TCR by a multi-domain protein complex known as the major histocompatibility complex (MHC), with the TCR binding to both the antigen fragment and the MHC.⁵ The following subsection will now describe the mechanism by which MHC molecules present antigens for TCR recognition.

1.3 Antigen Presentation to T-Cell Receptors

As discussed briefly above, TCRs differ from antibodies in that they target only small fragments of antigens presented on the cell surface. These antigens (which are most often peptides, but can also be lipids or metabolites⁶) are displayed on the cell surface by binding to the major histocompatibility complex (MHC). Please note that the MHC is a general term used to describe this molecule which is found in all vertebrates.⁴ The human leukocyte antigen (HLA) is the human form of the MHC and the term “HLA” will now be used throughout to avoid any confusion.

HLA molecules bind peptides available inside the cell to form the peptide-HLA (pHLA) complex. Once a pHLA complex is produced, it is transported to the cell surface and presented facing outwards into the extracellular milieu, allowing for TCR recognition.⁷ It should be noted at this point that HLA molecules bind to and present both endogenous (self-produced and therefore “healthy”) and antigenic peptides. It is primarily the role of the TCR to bind (and therefore trigger an immune response) to only antigenic peptides and not the self-peptide-HLA (self-pHLA) molecules which are also presented on the surface of the cell (discussed further in **Section 1.5**).

There are two major classes of HLA molecules (Class I and II) and both perform different roles in the immune system. HLA class I molecules present peptide fragments on the surface of their cells, with these peptides sourced from proteins produced inside the cell (see **Figure 2**). This process occurs in all nucleated cells (cells with a nucleus), with the first step being the degradation of proteins into small peptides via the proteasome.⁸ The peptide fragments which bind to the HLA strongly enough are then transported and presented onto the cell surface for TCR recognition. Unlike Class I, HLA Class II molecules source peptides for presentation from outside the cell.⁹ Proteins outside the cell surface are internalised and degraded into peptides. The peptides produced that can bind HLA Class II are then transported to the cell surface for TCR recognition. HLA Class II expression and presentation is performed only by certain immune cells, which include dendritic and B-cells.⁹ This is due to the different roles the two HLA Classes have in the immune system. HLA Class I molecules are recognised by TCRs located on cytotoxic T-cells, which kill the cell (by the release of cytotoxic compounds) presenting the HLA

Class I molecule.⁸ HLA Class II molecules are instead recognised by TCRs present on T-helper cells.⁹ Unlike in Class I, antigen recognition (in this case by the T-helper cells) does not kill the HLA Class II presenting cell. T-helper cells are instead responsible for up/down regulating the local immune response. To ensure the correct T-cell type binds the correct HLA Class, a co-receptor present on the T-cell must also bind the pHLA in order to trigger an immune response.⁴ Cytotoxic T-cells contain a CD8 co-receptor molecule which can only bind to HLA Class I molecules⁸, whereas T-helper cells contain a CD4 co-receptor molecule which can only bind to HLA Class II molecules.⁹ These co-receptors are both located far away from the TCR/peptide binding site.

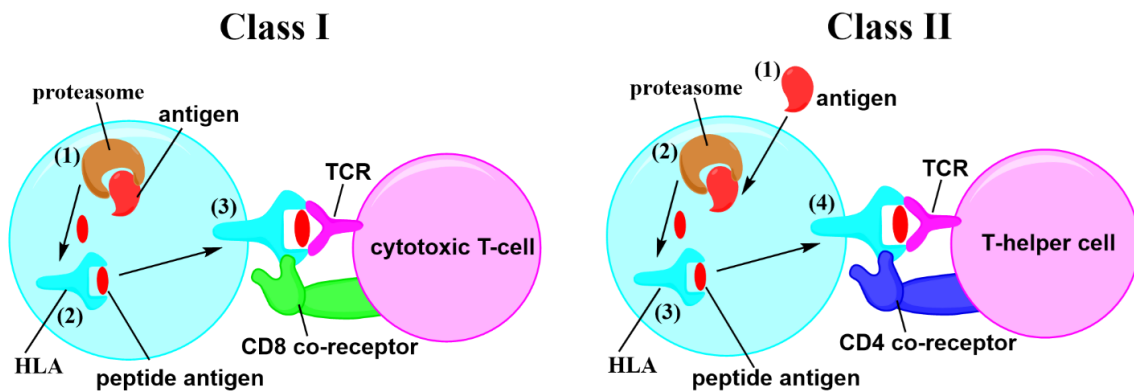


Figure 2: Schematic representation of antigen processing, presentation and T-cell binding pathways for HLA Class I and II. In the HLA Class I pathway, the antigen is first degraded in the proteasome (1), then bound to a HLA molecule (if it can bind) to form a pHLA complex (2), before being presented on the cell surface for both TCR and CD8 co-receptor recognition (3). In HLA Class II, antigens from the extracellular milieu are internalised (1), degraded by the proteasome (2), then peptides which are able to bind to a HLA Class II molecule (3) are presented on the cell surface for both TCR and CD4 co-receptor recognition (4).

1.4 Structural Characterisation of the pHLA Complex

HLA Class I and II molecules are composed of two different protein subunits (**Figure 3A**).⁹ In HLA Class I, a single protein chain forms both the peptide binding groove and the connection to the interior of the cell (via a transmembrane helix). This protein chain is in complex with a small (100 residue) β 2-microglobulin (β 2m) domain, which binds well away from the peptide binding site. In contrast, HLA Class II consists of two protein subunits of approximately equal size, with both domains connected to the interior of the cell (again via a transmembrane helix). In HLA Class II, both domains also play a roughly equal role in forming the peptide binding groove (**Figure 3A**). In both HLAs, the binding groove consists of a β -sheet base with a flanking α -helix either side to form the cavity. As

the results chapters of this thesis focus on HLA Class I molecules (due to their increased relevance as therapeutic targets, see **Section 1.3**), the remainder of this section will focus on characterising the structural features of HLA Class I in more detail.

HLA Class I Molecules tend to bind peptides of sequence lengths between 8–12, primarily through interactions between the HLA and the first and last few residues of the peptide.¹⁰ The regions of the HLA responsible for binding the N and C-terminal residues of the peptide are referred to as the A and F-pockets respectively (See **Figure 3B+C**). This “anchoring” effect of the N and C-terminal portions of the peptide often results in these residues playing a less significant role in TCR recognition, as they are relatively buried into the HLA’s peptide binding pockets. The central portion of the peptide however tends to “bulge out” of the binding pocket, making these residues primarily responsible for regulating TCR affinity to the pHLA.

There are 177 isoforms of HLA Class I molecules¹, and the combination of A and F-pockets and peptide N and C-terminal residues will largely control whether a particular peptide will bind to a given HLA isoform, and therefore can be presented to a TCR. Extensive studies have been performed to characterise the peptide sequence space that individual HLA isoforms are capable of binding.^{11–13} A commonality identified is the occurrence of peptide “anchors”, which are peptide residues that have to be one of very few amino acids in order to enable binding to the HLA isoform. Generally, HLA Class I molecules possess two “primary” and two “secondary” anchors (a primary anchor has a stronger effect than a secondary anchor), with one of each located in both the N and C-terminal residues.⁸

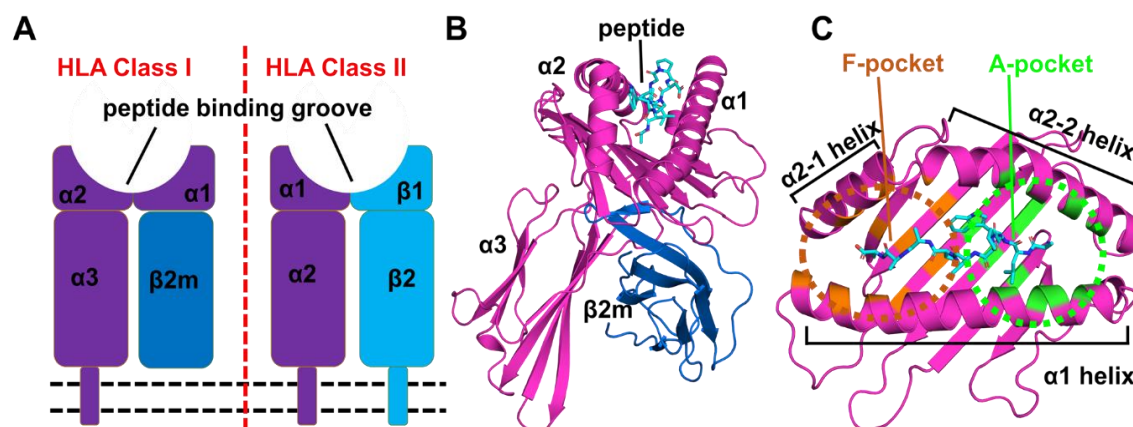


Figure 3: (A) Schematic diagrams of the HLA Class I and II complexes. All domains which make up both complexes are labelled, with domains that are part of the same protein coloured the same. The black dotted lines represent the location of the membrane. (B) Representative structure of HLA Class I molecule with peptide bound. (C) Top down view of the peptide binding site of HLA Class I. Residues which make up the A-pocket (peptide N-terminal binding site) and F-pocket (peptide C-terminal binding site) are coloured in green and orange respectively. The α -helices which form the groove are labelled on the figure. For panels B+C, protein structure taken from PDB 3UTQ.⁹⁸

1.5 How do Naturally Occurring TCRs Achieve Specificity?

Given only the 20 standard amino acids, there are theoretically $\sim 1 \times 10^{13}$ (20^{10}) possible different peptide sequences for a peptide of length 10 amino acids long. Furthermore, the very large majority of peptides that will be presented by the HLA will be from endogenous (native and healthy) proteins (and therefore should not be recognized by a TCR). This provides the $\sim 2 \times 10^7$ mature TCRs produced by the average human host with a seemingly impossible challenge.¹⁴ That is, they must be both highly cross reactive (to ensure coverage of the vast sequence space possible for antigenic peptides) whilst simultaneously being non-cross reactive to the vast number of self-pHLA complexes that will also be presented on healthy cells, to ensure an autoimmune response does not occur. Prior studies have shown that individual mature TCRs are indeed able to bind (at least) somewhere between the high thousands to millions of non-native pHLAs.^{15,16}

To overcome the aforementioned challenge, the immune system utilises a process known as “thymic selection”, in which a naive pool of $\sim 10^{11}$ T-Cells (each with different TCRs) are selected against several criteria to produce the $\sim 2 \times 10^7$ mature T-cells.^{14,17} The first step of thymic selection (see **Figure 4**) is “positive selection”, in which naive T-cells are tested for their ability to bind self-pHLA molecules.⁶ T-cells that are unable to bind self-pHLA are subjected to apoptosis (cell death). This step is important to remove TCRs that would likely be of limited value as they would not be able to recognise any pHLA. The T-cells that survive positive selection then undergo “negative selection”, in which T-cells that bind too strongly to self pHLA are subjected to apoptosis (to remove TCRs that could cause autoimmune problems by inducing an immune response towards healthy cells).⁶

One consequence of thymic selection is that the TCRs produced tend to bind their targets with relatively weak affinities (K_D normally $\sim \mu M$).¹⁸ These relatively weak affinities and subsequently weak half-lives (normally in the seconds¹⁸) mean multiple TCRs on a single T-cell need to bind multiple pHLA on a single (diseased) cell in order to generate a strong enough signal for an immune response to be generated.¹⁹ This requirement for multiple TCRs to bind a diseased cell provides an unfortunate “escape mechanism” often used by diseased cells to avoid destruction. That is, the diseased cells down regulate the production of pHLA on the cell surface (previous studies have observed as few as 10 antigen specific pHLA presented per cell for several different cancers), making it very challenging for the “critical mass” of simultaneous TCR binding events to occur that would allow for the diseased cells eradication.¹⁹

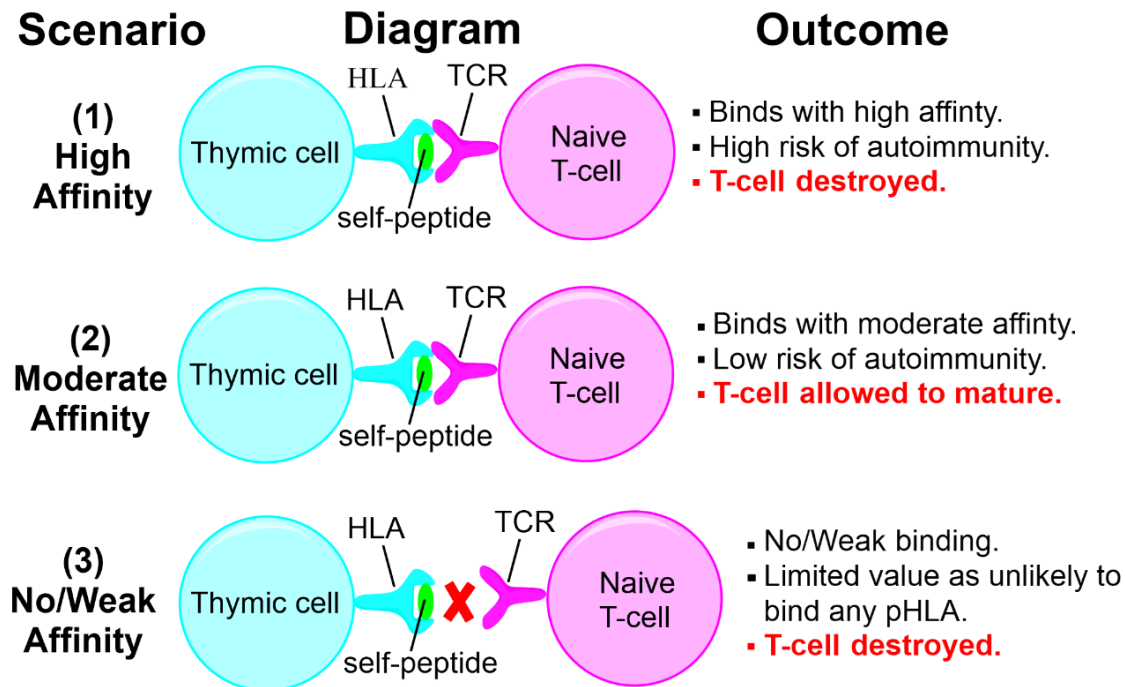


Figure 4: Positive and negative thymic selection of T-cells by the Thymus. Naive T-cells are produced in the thymus and their affinity towards self-peptides is evaluated. If a peptide binds too strongly (scenario 1) or too weakly (scenario 3) the T-cell is subjected by apoptosis (cell death). Only in scenario 2 is the T-cell (and its corresponding TCR) allowed to mature.

1.6 Structural Features of the TCR-pHLA Complex

TCRs bind pHLA through their six hypervariable loops (see **Figure 5**), with both the peptide and the two alpha helices of the HLA ($\alpha 1$ and $\alpha 2$) that form the peptide binding groove responsible for interactions with the TCR. TCRs tend to dock diagonally over the pHLA complex such that the CDR3 loops make contacts primarily with the peptide (as well as some HLA residues) and the CDR1+2 loops primarily engage the HLA residues on the alpha helices (see **Figure 5B+C**). The orientation at which the TCR docks can be characterised by determining the “docking angle” of the TCR with respect to the pHLA.⁵ This measurement determines the angle between the major axis of the peptide and the vector between the (conserved) interface cysteine residues on the variable domains, with the canonical range of docking angles observed for (crystallised) TCRs between 40 to 85°.⁵

An approach first used to characterise the orientation of the two antibody variable domains with respect to one another,²⁰ has been modified for use with TCRs²¹ (known as “Abangle” for antibodies and “TRangle” for TCRs). This approach gives six

measurements (five angles and one distance) to describe the differences in variable domain orientations. Interestingly, clustering of the available structures of antibodies and TCRs based on their Abangle/TRangle parameters results in them being clustered into separate groups. Furthermore, by computationally modifying TCRs to adopt standard antibody variable domain orientations, it was shown that the TCR could no longer bind the pHLA with a canonical (i.e. TCR-like) docking angle.²¹ This may therefore suggest that engineering antibodies for the application of pHLA recognition could be more challenging than engineering TCRs, as the TCR scaffold has arguably evolved to bind with an orientation optimal for peptide discrimination.²¹

In **Section 1.5**, the biological mechanism for thymic selection was described, in which the TCRs produced are simultaneously cross-reactive, yet “specific” in the sense that they do not bind endogenous pHLA molecules presented on healthy cells. This does not explain however the physical chemical mechanism of how TCRs are able to do this. Whilst still not fully understood, arguments have been proposed in the literature to rationalise these observations. The major argument is that TCRs have “hotspot” peptide and HLA residues which are primarily responsible for driving binding affinity, meaning if all or most of these residues are present, the TCR is likely to bind.²² Recent work has identified a single TCR that can bind over a million different peptide sequences presented on the same HLA isoform.¹⁶ Several peptide sequence motifs were identified from this, with the most populated being: xOxGPDxxxO”, whereby “O” is a hydrophobic amino acid and “x” can be any amino acid.¹⁶ A follow-up study in which eight of the identified peptides in the above described motif were crystallised identified the importance of the central “GPD” residues in binding the TCR.²³ In all eight cases, between 41-50% of all X-ray contacts formed between the TCR and pHLA were from just these three residues.²³

Recent work by Riley *et al.* demonstrated that a single TCR was able to bind two distinctive classes of peptide sequences by adopting essentially the same TCR conformation for the two different peptides.²⁴ In this study, yeast display of many pHLA complexes was used to identify 53 different peptide sequences that could bind the same TCR (to form the TCR-pHLA complex). Analysis of the peptides sequences showed two distinct classes of peptide that could be differentiated by the presence of either a “DRG” or “GIG” motif in the center of the peptide, with each of these motifs “bulging” out of the HLA binding groove to engage the TCR. Several of the above identified peptides were crystallised, which interestingly showed that the TCR used essentially the same orientation to bind both types of motifs (all atom RMSD difference in the CDR loops for the TCR was < 1 Å for all examples). In contrast, the two classes of peptides adopted vastly different conformations, (with backbone RMSD differences of 3.5 Å).²⁴ This demonstrates that not only can the CDR loops of the TCR make use of conformational flexibility to enable binding of different pHLA surfaces (not seen in this case but seen in others²⁵), but that different peptides can adopt different conformations in the HLA groove to facilitate TCR binding.

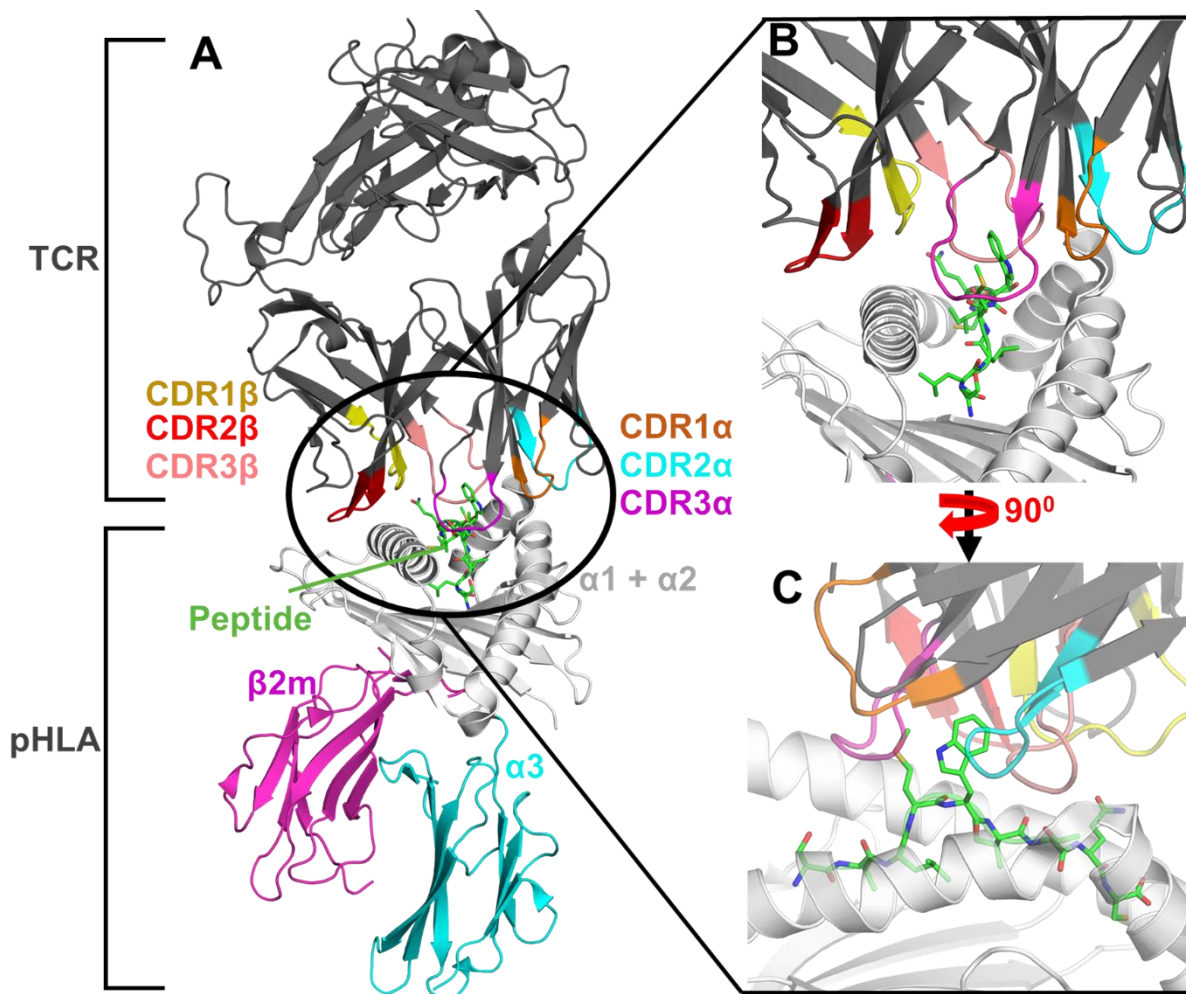


Figure 5: (A) Exemplar structure of the TCR-pHLA complex, with the six CDR loops and different portions of the overall scaffold coloured differently for clarity. (B+C) Zoom in on the TCR-pHLA binding site from two different orientations (same colouring as in A). Protein structure taken from PDB: 2P5E.⁴⁵

1.7 The Application of T-cells and TCRs as Therapeutics.

This section provides an overview of the different types of therapeutics either available today or currently under investigation that utilise TCRs or T-cells to combat disease. These therapeutics can be broken down into four major categories (see **Figure 6**), by partitioning based on the receptor molecule used to recognise the antigen and on the solubility of the drug.

The first category in **Figure 6** are so called IMMTACs (Immune mobilising monoclonal T-cell receptors Against Cancer), which are soluble, bispecific (can bind two different epitopes) and affinity enhanced (K_D for target pHLA \sim pM) TCRs.²⁶ As these TCRs are

engineered to be soluble (and therefore are not attached to a T-cell) they contain a co-receptor which binds with picomolar affinity to CD3 receptors on T-cells.²⁶ This enables them to recruit T-cells to destroy the antigen presenting cell.

The second category in **Figure 6** are tumour-infiltrating lymphocytes (TILs), which can be used in an approach known as adoptive cell therapy (ACT). In ACT, cells from the immune system are given to a patient to kill diseased (often cancerous) cells. In TIL therapy, T-cells located near tumours are taken from the patient and then multiplied in a lab before reinjecting the T-cells back into the patient.²⁷ The basic idea behind this process is that the hosts immune system is too weak to fight the tumour but has already produced the necessary T-cells to do so. Therefore, supplementing with additional anti-tumour T-cells can artificially “increase the strength of the immune system” and ultimately assist in fighting the disease. TIL therapy is similar to category 1 in that a TCR is used for pHLA recognition, however in this case the TCR used is membrane bound and (normally) of low affinity ($\sim\mu\text{M}$). A major challenge associated with TIL therapy is the risks of toxicity tend to be far greater than for soluble therapeutics due to the significantly longer long-half lives of T-cells in the body (compared to soluble proteins like TCRs).²⁸ Whilst one could engineer the TCRs in TILs to be of high affinity, this is not commonly performed because of the increased risk of toxicity.

The third category in **Figure 6** are chimeric antigen receptor T-cells (CAR T-cells).²⁹ CAR T-cells are T-cells produced by either the patient or a donor which are then engineered to produce specific antigen receptors (normally an antibody or antibody fragment) to bind a given antigen. The major differences between this category and category 2 is in the use of targeting protein (not a TCR) and the nature of the target antigen (pHLA for category 2 compared to a whole protein bound to the exterior of the cell surface for category 3).²⁹

Finally, category 4 in **Figure 6** refers to the many types of soluble antibody therapeutics (or other engineered protein scaffolds that could be used in its place).³⁰ As with category 1, soluble therapeutics need a way to destroy the antigen presenting cell (and not just bind to it). Two commonly used strategies are to (1) add a co-receptor binding site that can recruit another part of the immune system to kill the cell (e.g. a T-cell as seen in **Figure 6**) or (2) attach a drug molecule to the antibody via a partially chemically labile linker that will be released at the site of the diseased cell to kill it (commonly known as antibody-drug conjugates).³¹

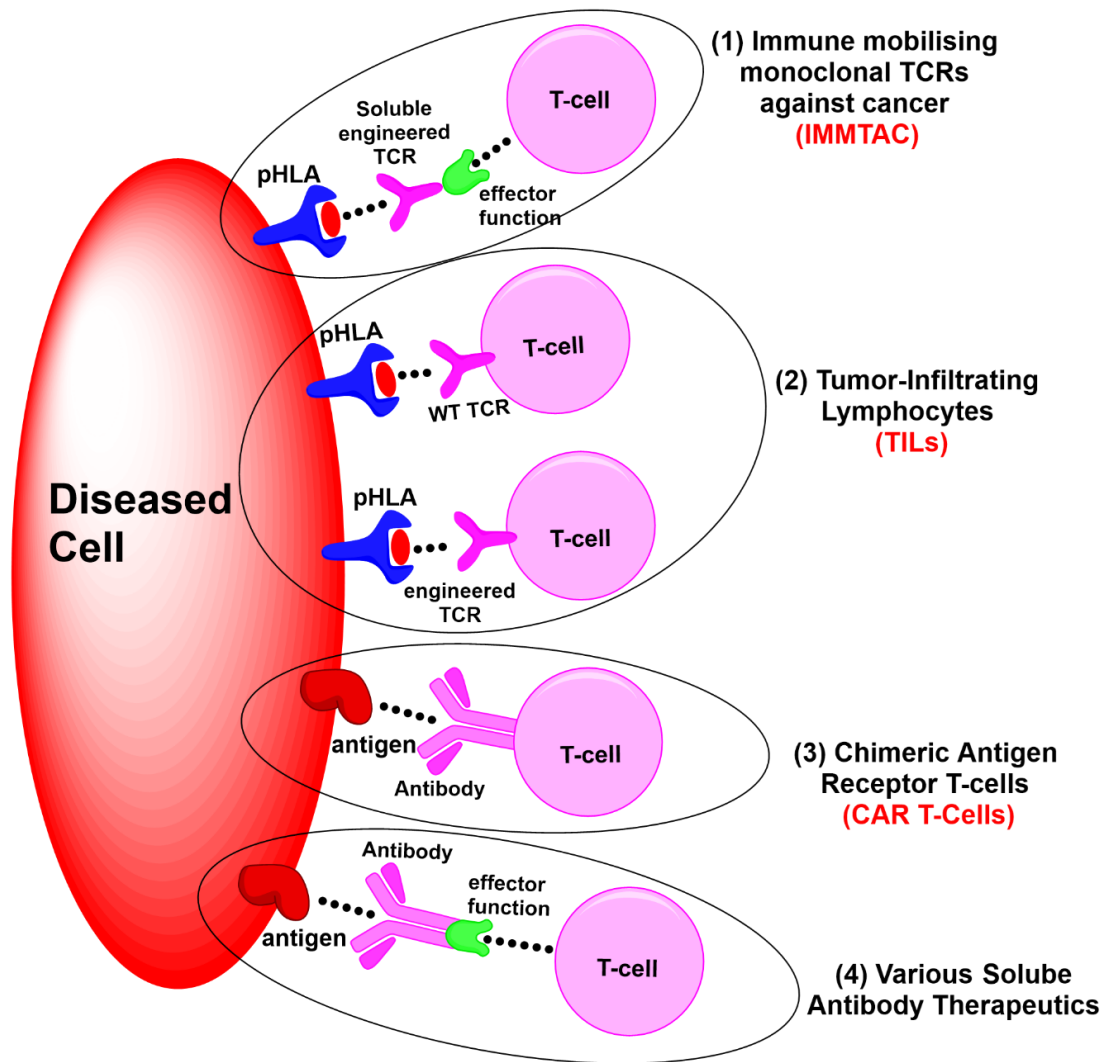


Figure 6: Schematic diagram of the various types of therapeutics available that utilise T-cells to kill diseased/cancerous cells. The dotted lines indicate binding between different molecules. In examples **1+2** the antigen presented is a peptide from a pHLA molecule, whilst in **3+4** it is a protein expressed on the surface of the cell. A description of each of the therapy type is provided in the text. This figure was adapted from one produced by Immunocore Ltd. Original figure is available at: <https://www.immunocore.com/technology>

1.7.1 The Advantages and Challenges Associated with Soluble, High Affinity TCRs as Therapeutics

Chapters 4 and **5** in this thesis focus on gaining insight into fundamental design principles for achieving high specificity and high affinity in TCR based therapeutics. These chapters are performed in collaboration with scientists from Immunocore, a company that develops IMMTACs (category 1 in **Figure 6**). This subsection will therefore focus on the advantages and challenges associated with the use of IMMTAC style therapeutics.

Targeting pHLA molecules (instead of surface expressed proteins like antibodies or most engineered protein scaffolds do) provides access to the internal proteome and a greater number of potential targets. It is estimated that ~90% of all potential disease targets are located inside the cell (and these targets can be presented by HLA Class I molecules to TCRs) as opposed to the ~10% available to drugs that can only recognise antigens on the cell surface.³²

The use of a WT-TCR (sourced from either the patients or a donor) with ~ μ M affinity has arguably a reduced risk of toxicity in comparison to an affinity matured nM-pM TCR (as the affinity maturation process may have had the unintended consequence of producing a TCR with reasonable affinity for an self-pHLA molecule). As discussed in **Section 1.5** however, the low affinity of WT-TCRs can provide diseased cells the opportunity to escape TCR recognition by down regulating the expression of antigenic pHLA. Previous studies have shown soluble TCRs with picomolar affinity (and subsequently much longer half-lives) can kill cells with as low as 5-10 copies of the pHLA target present on an individual cell.³³ If one can therefore produce affinity matured TCRs that have a large enough therapeutic window, then one would expect them to work more effectively than a WT-TCR as a therapeutic. A further point to consider is the reduction in the dose required, which is likely to be approximately inversely proportional to the affinity for the antigenic pHLA.

Whilst questions have been raised about the fundamental structural/orientational differences between TCRs and antibodies²¹ (see **Section 1.6**) and whether these differences make TCRs better suited than antibodies (or other proteins scaffolds) for specific pHLA recognition, there is no reason to think one could not use an antibody instead, and indeed several examples exist in the literature of generating antibodies to do exactly that.³⁴⁻³⁶

Evaluating the specificity of a high affinity TCR (or any engineered protein) as a potential therapeutic often requires a large number of cell-based assays, which are both time consuming and expensive. This bottleneck could of course be elevated in the future by an improved understanding of what makes a specific TCR, and better computational protocols to predict the specificity of candidate TCRs more accurately.

1.8 The Application of Computation to Understand and Engineer TCRs.

The following section aims to provide insight into how computation has been used to study and, in some cases, engineer TCRs. Please note that the theoretical aspects of the computational methods used in this thesis are covered in detail in **Chapter 2**.

1.8.1 Methods to Calculate Protein-Protein Binding Free Energies.

The large number of methods available to calculate protein-protein (or protein-ligand) binding free energies can be separated out into “pathway” and “end-state” calculation approaches. Common pathway methods include free energy perturbation or thermodynamic integration and work based on the principle of defining a reaction coordinate (or multiple reaction coordinates) that describe the transition from one state to another, in order to calculate the free energy difference between the states. These approaches can utilise alchemical (unphysical) pathways to describe the change in free energy and thermodynamic cycles can be constructed to produce either absolute or relative binding free energies (i.e. ΔG or $\Delta\Delta G$ respectively).³⁷

End-state free energy methods instead calculate the free energy of the two desired states separately (i.e. bound and unbound) and then directly subtract their free energies from one another. Although in this case one explicitly obtains a ΔG value, the approximations and errors associated with end-state methods mean they are commonly used to predict $\Delta\Delta G$ values or to rank a group of similar ligands instead. End-state approaches can be performed on a single or multiple conformation(s), with ranking of the protein-ligand complexes normally performed using empirical “scoring functions”. Scoring functions are based around countering the number and strength of various types of interactions (e.g. hydrophobic, hydrogen bonds etc...) between the receptor and ligand to determine the strength of the overall interaction.³⁸ The coefficients used for each interaction term are normally determined by non-linear regression. An alternative to using a scoring function is to use a MMFF to calculate the electrostatic and vdWs interactions between the receptor and ligand and then model the solvation effects with an implicit solvent model such as the Generalized Born (GB) or Poisson-Boltzmann (PB) equation. This approach is known as the MMPB/GBSA (Molecular Mechanics Generalized Born/ Poisson-Boltzmann Surface Area) method (described in detail in **Section 2.5**).³⁹ One could also consider describing some or all of the atoms with a quantum mechanical (QM) model, potentially offering increased accuracy. Considering the overall size of the TCR-pHLA complex (not only in terms of total numbers of atoms, but also that the binding interface is very large), approaches that use QM are unlikely to be worth the additional computational cost (most QM methods computing times scale significantly greater than linearly with an increasing numbers of atoms³⁸). Further, for the same amount of computing resources, if one were

to use a QM method to describe a TCR-pHLA interface, the amount of sampling possible (both in terms of number of conformations per complex and number of mutations tested) would be significantly reduced as compared to an MMFF method.

A completely different way to calculate the binding free energy could be to utilise machine learning (ML) approaches. ML methods normally require a large amount of (reliable) experimental data to generate models that are predictive. ML approaches have had reasonable success in predicting which peptide sequences can bind a given HLA isoform.⁴⁰ This is arguably because as long as the HLA A and F-pocket residues are compatible with the peptide N and C-terminal residues, binding is reasonably likely to occur (see **Section 1.4**).

1.8.2 Prior Attempts to Perform Rational TCR Engineering

In comparison to the available literature and software programmes/protocols available for antibody modelling and design (for example RosettaAntibody⁴¹ and the MOE antibody modeler suite⁴²), TCRs have been subjected to a relatively limited amount of study. Further, most publications that have generated high affinity TCRs appear to have done so through high throughput (semi-)random mutagenesis protocols.^{26,43-45} Two examples do exist in the literature however in which molecular docking was used to generate higher affinity TCRs.^{46,47} In the first of these two publications, single point mutations were predicted for the TCR α -chain towards binding the Tax peptide/HLA-A2 complex, with four point mutations combined to obtain a mutant with an ~100-fold increased affinity towards the pHLA.⁴⁶ In this approach Rosetta was first used to identify potential beneficial mutations from the starting crystal structure of the WT-TCR-pHLA. Following experimental testing a scoring function named “ZAFFI” (Zlab AFFInity enhancement) was developed by taking terms from both the ZRANK⁴⁸ and Rosetta⁴⁹ scoring functions and parameterising these terms to improve the fitting to the obtained experimental data. In the second publication, ZAFFI was used to predict single point mutations this time on the melanoma-associated Melan-A/MART-1 peptide bound to HLA-A2.⁴⁷ Combination of two single point mutations resulted in a 400-fold increase in affinity relative to the WT. These results demonstrate the potential of docking and computational chemistry towards rationally designing TCRs.

If the TCR-pHLA complex structure is not known docking will be required to obtain a starting model from which to design mutations. Even with structures of the apo TCR and pHLA, accurate prediction of the complex structure can be challenging. This is in part due to the conformational flexibility available to both the CDR loops and the peptide, as major conformational changes have previously been observed upon going from the unbound to the bound state for both the TCR and peptide.^{24,50} This means rigid body docking approaches are highly likely to be inappropriate for TCR-pHLA complexes.⁵¹ A

docking approach specifically designed for TCR-pHLA docking known as TCRFlexDock⁵² has also been developed. In this approach, iterative rigid-body moves and side chain re-packing is combined with CDR loop and peptide backbone translations, in order to sample a large amount of conformational space to obtain a predicted docking pose(s).⁵²

1.9 Aims and Objectives

Chapters 3-5 of this thesis explore some of the fundamental aspects of TCR-pHLA recognition and pHLA immunology. The research objectives for each chapter are presented below:

1.9.1 Objective 1: The role of Peptide Cargo in Allosterically Regulating the HLA

In **Chapter 3**, we aimed to determine how different peptide cargo can tune the molecular flexibility of the HLA. We explored this using both experiment (pressure and temperature (p/T) perturbation) and computation (MD simulations). Our p/T perturbation experiments demonstrated significant changes in the conformational dynamics of the pHLA, dependant on the peptide cargo. We applied MD simulations to rationalise how different peptide cargo could modulate the molecular flexibility of the rest of the HLA molecule. Our MD simulations identified several regions across the HLA which had their flexibility tuned by the peptide cargo. This includes many regions distal from the peptide binding site. We then analysed how different peptides were able to modulate the regions distal from the peptide binding site, and the pathways used to do so.

1.9.2 Objective 2: Engineering Principles for TCR-pHLA Specificity

In **Chapter 4**, we wished to determine fundamental design principles towards developing high specificity TCR-pHLA therapeutics. This was done through a combination of experimental (biochemical and structural) and computational (MD simulations and free energy calculations) approaches. We demonstrated that our MD simulations and free energy calculations were able to rationalise the observed specificity differences seen for the case studies considered, demonstrating the approaches used could have future applicability in rational design efforts. Our analyses suggested that an increased number of contacts between the TCR and peptide (in particular the side chains of the peptide) combined with binding with a broad energetic footprint (i.e. no hotspots/many favourable contacts) was likely to lead to a more specific TCR.

1.9.3 Objective 3: Engineering Principles for TCR-pHLA Affinity

Following on from the results obtained in **Chapter 4**, in **Chapter 5** we wished to gain insight into what drives affinity enhancement in engineered TCRs. We performed structural analysis, MD simulations and free energy calculations on several case studies of affinity maturation, with the aim to understand at the atomic level how affinity was enhanced. Analysis of the contacts formed between the TCR and pHLA either from the crystal structures or from MD simulations were unable to rationalise the differences in affinity. Analysis of changes in flexibility showed some instances of increased rigidity for the affinity matured apo TCR (which would be entropically favourable for binding). Our free energy calculations were able to reproduce the experimental affinity relationships. The decomposition of these results to the per-residue level allowed for the identification of several approaches used by TCRs to enhance affinity. Finally, we also demonstrated that the energetic footprint between the TCR and pHLA was by and large preserved over the course of the affinity maturation process, an important consideration when attempting to engineer specific TCRs.

Chapter 2:

Theory of Methodology

Physical methods to describe atoms and molecules can be broken down into two major classes, differentiated by the application of either quantum or classical physics to describe them. In quantum mechanical (QM) descriptions, atoms are described as wave-like, with a wave function used to provide information about the probability of finding a given particle at a particular position, as well as said particles momentum. Further, QM methods describe energy in a “quantised” manner, meaning only certain energies (energy levels) are allowed to be adopted by the system. In comparison, classical physics based methods describe energy as continuous and treat atoms as point particles. Generally, QM methods offer a much greater deal of accuracy, yet tend to be significantly slower than classical physics based approaches. As **Chapters 3-5** of this thesis deal with the describing the dynamical properties of large protein complexes, a classical physics based approach (through the use of a molecular mechanics force field) is used as it provides a good balance between accuracy and speed.

The following chapter aims to provide the reader with the theoretical background behind the simulation approaches used throughout this thesis.

2.1. A Molecular Mechanics Force Field

A molecular mechanics force field (MMFF) is an empirically derived model that describes the physical behaviour of atoms and molecules.⁵³ This is achieved by describing atoms (or in some cases groups of atoms) as individual particles and using several empirically derived equations to describe the interactions between particles. Whilst many different MMFFs exist, most take the general form depicted in **Figure 8**. These terms can be broken down into “bonding” (E_{bond} , E_{ang} , E_{dih} and E_{imp}) and “non-bonding” (E_{vdw} and E_{elec}) terms, which sum together to give the total potential energy of the system. Bonding interactions are normally described by making the “harmonic approximation” (i.e. a harmonic potential is used to describe the relationship between the potential energy and the distance between atoms). Whilst the harmonic approximation is accurate for describing near-equilibrium bond length potential energies and can be solved rapidly with a computer, it is important to consider one of its major limitations. Bond breaking is an anharmonic process in which once the bond has “broken” the energetic penalty does not change (significantly). A harmonic model would therefore be highly inappropriate for describing chemical reactions, as bond breaking would be over-penalised. To more appropriately describe chemical reactions, one could modify the reacting atoms MMFF to describe their E_{bond} terms with an anharmonic model, as is done in the case of empirical valence bond simulations.⁵⁴ Alternatively, one could turn to quantum mechanical methods for describing chemical reactivity.⁵⁵

$$E_{total} = E_{bond} + E_{ang} + E_{dih} + E_{imp} + E_{vdw} + E_{elec}$$

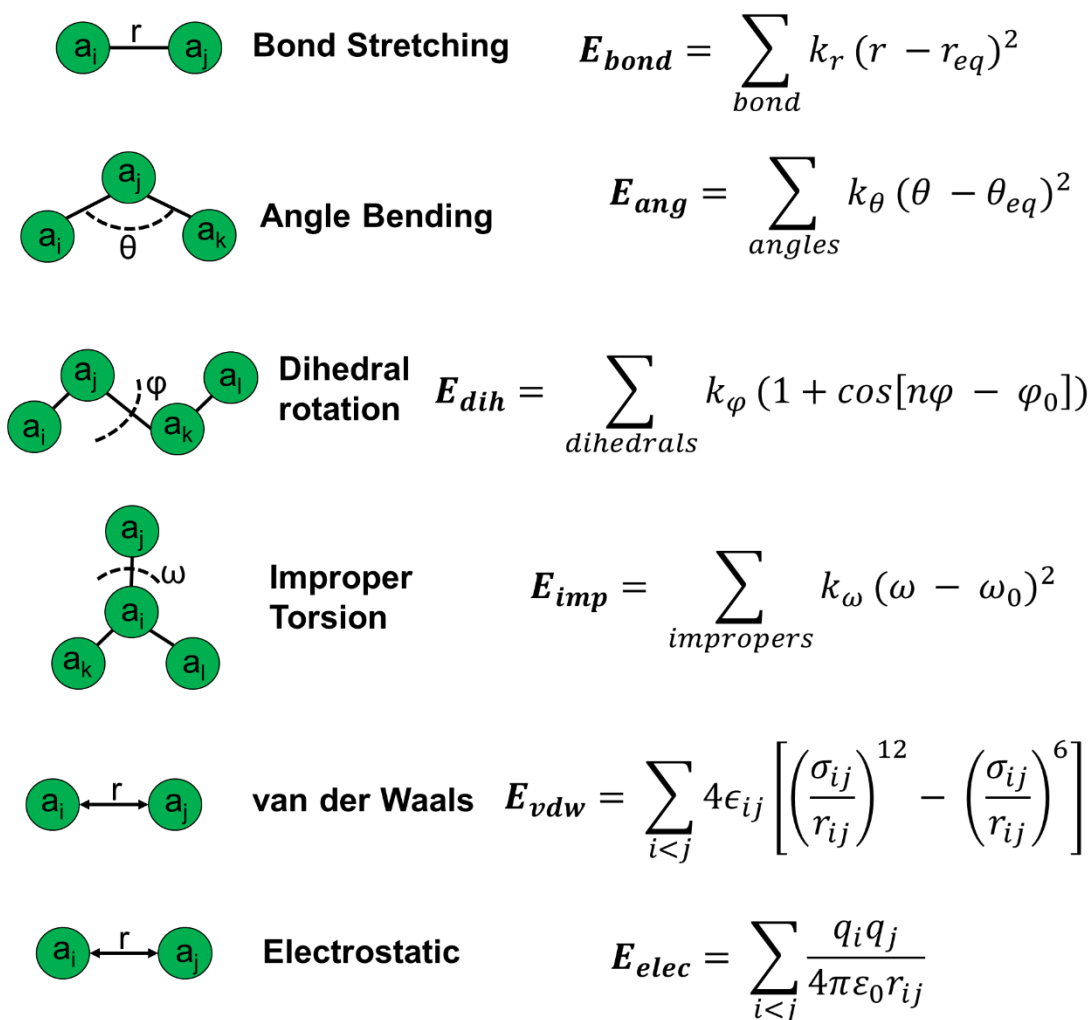


Figure 7: The common terms used in a molecular mechanics force field (MMFF). All terms are summed together to give the total (potential) energy of the system. For the first four terms k_r , k_θ , k_ϕ , and k_ω are force constants dependant on the distance (r), angle (θ), dihedral (ϕ) and torsion angle (ω) respectively. For the van der Waals (vdW) term, ϵ_{ij} is the well depth (minimum potential energy value possible), σ_{ij} is the distance between atoms i and j at which the potential energy is 0, and r_{ij} is the distance between atoms i and j respectively. For the electrostatic term, q_i is the charge of atom i and ϵ_0 is the dielectric constant.

MMFFs are usually parametrised through some combination of experimental and computational calculations, with the quality of the MMFF tested/validated by assessing its ability to reproduce certain experimental observations.⁵³ For example, the TIP3P water model is commonly used in biological simulations to provide a reasonable description of the behaviour of water at physiologically relevant temperatures (298 to 310 K) and

pressures (1 atm).⁵⁶ A further benefit is that several biological MMFFs are co-developed to work with the TIP3P water model (for example several CHARMM MMFFs⁵⁷). TIP3P is however known to be a poor choice for high or low temperature/pressure simulations due to its inability to describe the phase diagram of water (for example, TIP3P has a melting temperature of 146 K⁵⁸). In this case a different water model such as TIP4P/2005 would be better suited to reproduce the phase behaviour of water.⁵⁹ The above example demonstrates the importance of choosing an appropriate MMFF that is able to well describe the properties one wishes to gain insight into. Further, MMFFs are often produced to describe one class of molecules at a time (i.e. amino acids/proteins - Amber FF14SB⁶⁰, CHARMM36M⁶¹; lipids - Lipid14⁶²; sugars - GLYCAM06⁶³; and small organic molecules - GAFF⁶⁴, CGenFF⁶⁵). This is arguably because molecules in the same class are likely to have similar characteristics and trying to make a MMFF too general is likely to lead to an MMFF that is inaccurate.

2.2 Geometry Optimisation

A MMFF allows us to evaluate the potential energy (PE) of a system for a given set of atomic coordinates. A complete evaluation of all possible arrangements for a given system would therefore provide us with a complete description of the potential energy surface (PES), including the location of global and local minima(s) as well as the location of saddle points (i.e. transition states). Evaluating the PE of all possible configurations would be extremely expensive even for relatively small systems and highly inefficient (as many of the configurations evaluated would be of extremely high PE, for instance due to atom overlap). Thankfully, several algorithms exist that allow one to identify local minima or saddle points and even compute the minimum energy pathway between them.⁵³

Energy minimisation is the process by which a given set of coordinates are “optimised” to the nearest local minimum by minimising the potential energy of the system until it cannot be further reduced. Fast geometry optimisation procedures such as the steepest descent (SD) algorithm work by calculating the first partial derivative for an initial set of coordinates.³⁸ The coordinates are then moved a short distance (often referred to as the “step size”) in the direction of the negative gradient. This procedure is iterated until the gradient obtained for the current configuration is smaller than a given threshold, at which point the current geometry is considered to be at the local minimum. Whilst this procedure is rapid it often struggles to converge in systems with a large number of dimensions as it does not “remember” anything about the shape of the PES between steps.

Geometry optimisation is highly similar to energy minimisation (in that both aim to minimise the energy of the system), however geometry optimisation procedures rely on calculating or estimating the second partial derivative of the potential energy (commonly known as the Hessian), increasing the accuracy of the process.³⁸ The Hessian is a matrix of size $3n \times 3n$, where n is the total number of atoms in the system. In the conjugate gradient (CG) approach the Hessian is estimated and used to improve the quality of the prediction of the direction in which to move the coordinates in for a given step.³⁸ The CG method also uses the results of prior steps to help guide the minimisation procedure,

meaning it does not have to move in the direction of the smallest gradient. For systems with a large number of degrees of freedom, this tends to result in a reduction in the number of steps required to optimise the geometry.³⁸

In this thesis, energy minimisation is used to relax the system prior to MD simulations. Energy minimisation is first performed using the SD algorithm followed by the CG algorithm. This takes advantage of the high speed of SD to efficiently remove any “kinks” in the structure and/or relax highly energetically unfavourable conformations. Then the CG algorithm can be used to efficiently converge towards the local minimum.

2.3 Molecular Dynamics Simulations

The first derivative of the potential energy (PE) as a function of the atomic coordinates (r) is equal to the negative of the force (F).⁶⁶ With knowledge of the force “felt” by each atom alongside each atoms mass (m) we can make use of Newton’s second law of motion (Eq 1) to calculate the acceleration (a) of each atom.

$$F(t) = ma = - \frac{\partial PE}{\partial(r)} = m \frac{\partial^2 r}{\partial t^2} \quad (\text{Eq 1})$$

As can be seen above, the second partial derivative of the forces provides the displacement (r) of the atomic coordinates with respect to time (t). In a molecular dynamics (MD) simulation, we assume constant acceleration over a short period of time (referred to as the time-step, Δt). This enables us to calculate the new positions of the atomic coordinates after time Δt has passed. We then recalculate the forces on the atoms of the newly updated coordinates to predict the coordinates after a further time step (i.e. $2\Delta t$ from the starting time). The iteration of this procedure therefore allows us to simulate the dynamics of a system over time. In order to calculate the position of the atomic coordinates after the first time step (i.e. $t + \Delta t$) we solve (Eq 1) numerically to the second order giving (Eq 2).

$$\frac{\partial^2 r}{\partial t^2} \approx \frac{r(t + \Delta t) + r(t - \Delta t) - 2r(t)}{\Delta t^2} = \frac{F(t)}{m} \quad (\text{Eq 2})$$

The above equation can then be rearranged to give the displacement of the coordinates after the first time-step (i.e. $r(t + \Delta t)$, see (Eq 3).

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{\Delta t^2 F(t)}{m} \quad (\text{Eq 3})$$

To deal with the issue that at $t = 0$ we do not know/have the coordinates of $r(t - \Delta t)$, velocities are randomly assigned to each atom, making sure that the total kinetic energy of the system is consistent with simulation temperature (see Section 2.3.2 for further details).⁵³

Whilst the assumption of constant acceleration is incorrect because the forces felt by each atom are positionally dependent, a larger timestep would increase the efficiency of an MD simulation. The major limitation placed on the maximum time-step possible for a MD simulation is the requirement to well describe the fastest motions of the system (bond vibrations and rotations).³⁸ For example, a C-H stretch has a frequency of $\sim 3000 \text{ cm}^{-1}$, which corresponds to a bond oscillation time of $\sim 11 \text{ fs}$. Therefore, time steps close to or higher than the bond oscillation time could result in unrealistic conformations obtained over the course of the simulation. As the oscillation time is inversely proportional to the square root of the reduced mass of the two atoms in the bond, bonds containing hydrogen tend to have the shortest oscillation times. A commonly used procedure in MD simulations is to constrain the bond length for any bond that contains a hydrogen atom, which tends to allow one to increase the timestep used from 1 fs to 2 fs, effectively doubling the simulation speed.⁶⁶ In this thesis, the SHAKE algorithm⁶⁷ is used to achieve this for all MD simulations run. Coarse-grained simulations (in which small groups of atoms are merged into “super atoms”) will simulate particles which have a significantly larger reduced mass, allowing for much larger time steps to be used. For example, for MARTINI, a coarse-grained MMFF for biomolecules, using a timestep of between 20–40 fs is recommended for MD simulations.⁶⁸

The following subsections will now provide detail on some key considerations for running an MD simulation (particularly focussed on simulations of biomolecules in explicit solvent as performed in this thesis).

2.3.1 Periodic Boundary Conditions

Most MD simulations aim to recapitulate experiment, in which a protein would be surrounded by a large amount of bulk water. Simulating a protein surrounded by a droplet of water would give rise to “surface effects”, meaning water molecules at the exterior of the droplet would not be solvated by water (as they should be in bulk solution).⁵³ This could result in water molecules evaporating off into empty space or overcompensating for the limited number of interactions surrounding them to form highly ordered water molecules. To alleviate this issue, MD simulations often use periodic boundary conditions (PBCs). In a PBC simulation, the solute(s) are solvated in a rectangular (other shapes are also possible) box of solvent and this box is replicated in all x, y, and z directions.⁶⁶ These additional boxes are called “images” and are identical to the central box. On a practical level, PBCs mean any atom that passes through the box boundary will re-enter the box from the opposite side, instead of drifting off into empty space (as in the water droplet

example). Furthermore, non-covalent interactions can be “felt” through the box walls, meaning that all solvent molecules are completely solvated, giving a closer representation to the behaviour of bulk water.

2.3.2 Pressure and Temperature Regulation

MD simulations are often run using one of three different ensembles, whereby for each ensemble, the terms provided are kept constant. These ensembles are: NVE, NVT and NPT, where N is the number of particles, V is the volume, E is the energy, T is the temperature and P is the pressure. NVE simulations are the fastest, followed by NVT and then NPT simulations. This is because an NVT simulation requires a thermostat whilst an NPT simulation requires both a thermostat and barostat (pressure regulator), and an NVE simulation requires neither.

The temperature of an MD simulation is usually measured by its relationship to the kinetic energy (KE), see (Eq 4).

$$KE = \frac{3}{2}Nk_bT = \left\langle \sum_{i=1}^N \frac{1}{2}m_i v_i^2 \right\rangle \quad (\text{Eq 4})$$

Where N is the number of atoms, k_b is the Boltzmann constant and m_i and v_i are the mass and velocity of atom i respectively. It is important to note that the average temperature is proportional to the kinetic energy, and that on the microscopic scale the kinetic energy should fluctuate around this average value (following a Gaussian distribution).⁵³ An NVE simulation (in which no thermostat is present) would follow a constant energy scheme, not representative of this microscopic variation. Thermostats enable the representation of this natural variation, through modifying the velocities of the atoms in the system. In this thesis, a Langevin thermostat⁶⁹ is used to regulate the temperature. Langevin thermostats work by implicitly modelling the effects of viscosity and random collisions with the surroundings on the system, resulting in both the addition and removal of kinetic energy around the target temperature.

The pressure (P) of an MD simulation is measured using the Virial theorem⁷⁰, which relates the forces applied to the simulation box to the pressure (see (Eq 5)).

$$P = \frac{1}{3V} \left\langle 3Nk_bT + \sum_{i=1}^N f_i r_i \right\rangle \quad (\text{Eq 5})$$

Where V is the volume and f and r are the forces and positions of atom i . Alike thermostats, barostats regulate the instantaneous pressure (pressure at any given time) to

fluctuate around the average pressure.⁵³ Most barostats work by either modifying the size of the simulation box (whilst scaling the coordinates appropriately so the pressure change is felt throughout the system and not just at boundary) or by modifying the velocities (and therefore forces) to give constant pressure. In this thesis, a Berendsen barostat⁷¹ is used to regulate the pressure. This is achieved by coupling the system to a weakly interacting pressure bath, which scales the volume of the periodic box at set intervals, producing fluctuations in the pressure around the target pressure.

2.3.3 Dealing with Long Range Electrostatics Interactions

As discussed in Section 2.1, non-bonded interactions in MMFFs are described by both van der Waals (vdWs) and electrostatic interactions. Calculation of non-bonded interactions tends to be the slowest part of an MD simulation timestep (as there would be n^2 pairwise distances to evaluate, where n is the number of atoms). The vdWs repulsive and attractive terms scale by r^{-12} to the r^{-6} respectively, meaning after a short distance their strength is negligible (see **Figure 9**). To reduce simulation cost, a distance cut-off for vdWs interactions is normally applied such that only interactions within the cut-off are considered (this is typically between 8–12 Å).⁵³ Electrostatic interactions however scale by r^{-1} , which would mean ignoring their interactions after short distance cut-off would be inappropriate. In implicit solvent simulations, electrostatic interactions normally do not have a cut-off (or at least a very large cut-off is used). In explicit solvent simulations, electrostatic interactions are treated explicitly if within a certain cut-off distance (normally selected to be the same size as the vdW cut-off). Longer range interactions are treated by grid-based Ewald summation approaches like the particle mesh Ewald (PME) method.⁷² In these approaches, explicit charges (from all atoms) are spread onto a grid covering the entire periodic box and used to define the long range electrostatic interactions felt by each atom in all directions.

It is important to consider non-bonding interactions when deciding on an appropriate box size. That is, one does not want to select a periodic box size that is small enough to allow the solute to directly interact with itself. Furthermore, the solute can indirectly interact with itself by perturbing the solvent. It is therefore important to consider that the box size is large enough such that the solvent is in a “bulk-like” state when approaching the box boundaries.⁵³

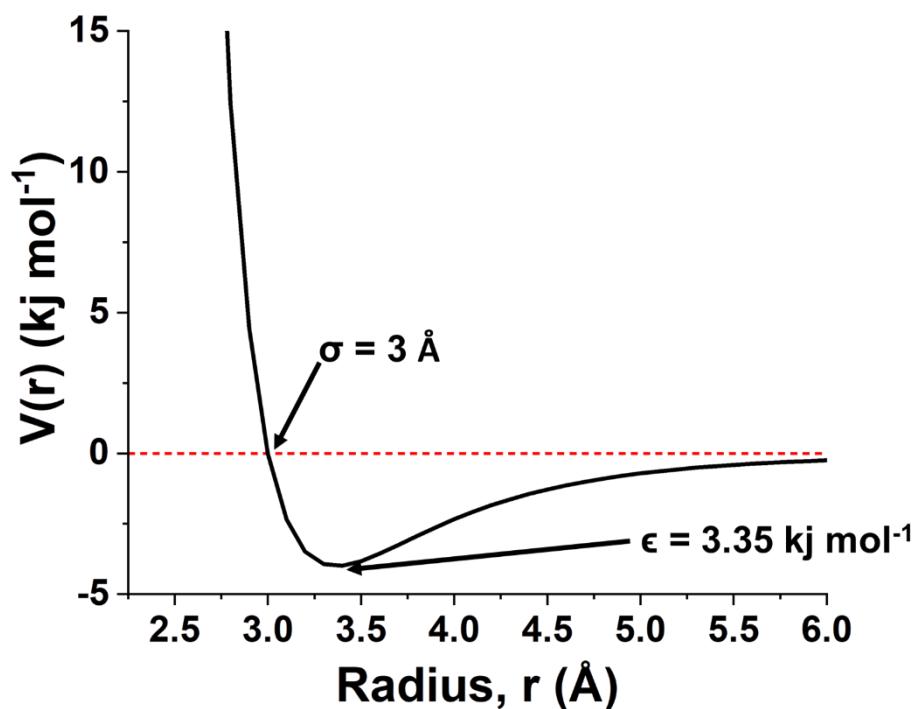


Figure 9: Exemplar Leonard-Jones Potential describing the strength of a vdWs interactions between two interacting particles. The locations of σ (distance at which the intermolecular potential energy is 0, i.e. the vdWs radius) and ϵ (well depth) are indicated on the figure. As can be seen from the graph, after a short distance of $\sim 6 \text{ \AA}$, the potential energy is already near 0. For the equation used to calculate the Leonard-Jones potential, see **Figure 8**.

2.3.4 A General MD Simulation Protocol

The setup and running of an MD simulation can be broken down into five general stages which are: preparation of the system, geometry optimisation, heating, equilibration and production. The following subsection provides details on the key parts of each of these steps.

System preparation involves generating a complete atomic model of the system one wants to study alongside the parameters needed to describe the system. The structure of the solute can be obtained from X-ray crystal, NMR solution or cryo-EM structures. Alternatively, if the structure of the biomolecule is not known, it could be predicted with homology modelling or ab initio structure prediction (or some combination of the two).⁷³ Missing residues (from an experimentally determined structure) should be modelled back in where appropriate. Simulations of protein complexes where the structure of the complex is not known can be performed by manually docking the receptor and ligand or using one of the many different docking algorithms available.⁷⁴ An important

consideration for structures solved by crystallography is that the resolution is not normally $<1 \text{ \AA}$. This means the tautomerisation and/or protonation states of histidine residues and the position of the side chain amine or carbonyl groups of asparagine and glutamine residues cannot be resolved by experiment. Several programs exist to solve this problem by identifying the most likely states based on the criteria of optimising the hydrogen bond network.⁷⁵ It is recommended to validate the obtained results by eye to ensure the outputted results make sense. The protonation states of titratable residues (or other titratable species) should be considered based on the desired simulation pH. Several empirical methods exist to rapidly predict the pK_a of titratable residues which should normally suffice.⁷⁶ Once the solute has been prepared, one needs to solvate the structure in a box of water in order to perform PBC simulations. Finally, the system needs to be parameterised and one should take care not only to select MMFF(s) that well describe the properties one is interested in obtaining, but also that the MMFFs selected are compatible with one another.⁵³

For the energy minimisation step, one should ensure a sufficient number of steps are performed such that the structure is close to the local minimum. If not, instabilities can arise in the subsequent heating step (as the forces projected onto atoms in high energy states can be very large). The heating step is the first point at which velocities (and therefore kinetic energy) is assigned to the system. It is generally recommended to start from a low temperature (to avoid the introduction of large velocities) and relatively slowly heat the system to the desired target temperature.⁵³ In the simulations of TCRs or pHLAs performed here, following an initial energy minimisation of only the solvent molecules or hydrogen atoms, the solvent is heated to the target temperature (using restraints on all solute heavy atoms throughout) to allow it to relax around the solute structure(s). Then, the entire system is energy minimised again and then heated to the target temperature (this time with only gentle restraints on the $C\alpha$ of any protein atoms).

Heating is normally performed in the NVT ensemble. It is important to consider that the solvent box added will be purposefully too large for the size of the system. This is done to allow water molecules adequate room to adjust upon heating, so systems are less likely to become trapped in high energy states. As the box size is kept constant during an NVT simulation, NPT equilibration simulations should be performed so that an appropriate box size can be obtained for the target pressure and temperature. If one plans to later swap back to NVT or NVE simulations after NPT equilibration, one should calculate the average box size (as this property will fluctuate over the course of the NPT simulations) after the box is equilibrated and use this for NVT or NVE simulations. Before moving onto production MD simulations (the point at which trajectory information is used for analysis) one should validate that the biomolecule itself has stabilised. One way to assess this would be to calculate the root-mean-square deviation (RMSD) of the protein over the course of equilibration procedure (only beginning production MD simulations once this has converged to a reasonably fluctuating level).

2.4 Allosteric Analysis of MD Simulations

In **Chapter 3**, some less commonly used simulation analysis techniques are performed in order to probe for allosteric communication between different parts of the protein complex. Herein, a short description is provided on the two approaches used.

2.4.1 Dynamic Cross Correlation Matrices (DCCM)

The identification of correlated motion between atoms over the course of a simulation(s) can help to identify regions distal from one another that appear to be dynamically linked to one another. The amount of cross correlation ($C_{i,j}$) between atoms i and j can be calculated using (Eq 6).⁷⁷

$$C_{i,j} = \langle \Delta r_i \cdot \Delta r_j \rangle / (\langle \Delta r_i^2 \rangle \cdot \langle \Delta r_j^2 \rangle)^{1/2} \quad (\text{Eq 6})$$

Where Δr is the displacement from the mean position for the i th or j th atom. The value obtained will be between 1 and -1 , with 1 indicating perfectly correlated motion, -1 indicating perfectly anti-correlated motion and 0 indicating no correlation between the atoms. In the above terminology, anti-correlated motion between two particles would mean the two particles movements are correlated with one another but they move (on average) in the opposite direction to one another (as opposed to just “correlated”, in which case they move in the same direction as one another). Cross correlation values obtained through MD simulations tend to always be non-zero. A cut-off is therefore often used to ascertain whether the correlation observed between the two atoms should be considered meaningful. Commonly, the correlation between the C α carbon of each residue is used as a proxy to measure the degree of cross correlation between residues. Calculating the degree of cross correlation between all residues would give a matrix of these values (referred to as a dynamic cross correlation matrix, DCCM).

2.4.2 Community Network Analysis (CNA)

Community Network Analysis (CNA, sometimes also referred to as correlation network analysis) is an approach which builds upon a pre-determined DCCM for a given system by clustering groups of highly correlated residues together into communities and calculating the degree of correlation (which acts as a proxy for communication) between different communities.⁷⁸ This data can be represented in a graphical manner, in which a node on a graph corresponds to a community, and the edges between nodes indicate the strength of communication between the two communities (with increasing edge thickness

indicating increased correlation/communication). Clustering of residues into communities is performed using the Girvan–Newman algorithm,⁷⁹ which uses a “non-hierarchical” approach to cluster communities. This means edges are progressively removed from the protein network if they are more likely to be between communities (the smaller the magnitude of correlation, the more likely the edge should be between communities and not in the same community). The quality of the clustering partition can be evaluated by calculating the “modularity”. The modularity provides a value between 0 and 1 and assesses the overall degree of inter-community correlation against the intra-community correlation. A higher modularity value is desired as it is indicative of increased inter-community correlation and decreased intra-community correlation.

2.5 Molecular Mechanics Poisson-Boltzmann/Generalised Born Surface Area Calculations

The following section describes the theoretical basis for the Molecular Mechanics Poisson-Boltzmann/Generalised Born Surface Area (MMPB/GBSA) approach in order to predict binding free energies. Practical considerations for the application of MMPB/GBSA based on prior literature findings are also discussed.

2.5.1 Theoretical basis.

Unlike free energy perturbation methods which slowly transition from one state to another through a series of intermediates, MMPB/GBSA is an end-state free energy calculation method.³⁹ This means the Gibbs free energy of the complex, receptor and ligand are calculated separately and then combined in order to obtain the Gibbs free energy of binding (ΔG_{bind} , as shown in **Figure 10**).

As indicated in **Figure 10**, ΔG_{bind} is obtained by constructing a thermodynamic cycle such that the solvation free energy change is calculated separately from the gas phase free energy change. The general equation used in MMPB/GBSA to calculate all the terms required for ΔG_{bind} is shown in (Eq 7).⁸⁰

$$\Delta G_{bind} = \Delta E_{int} + \Delta E_{el} + \Delta E_{vdw} + \Delta G_{pol} + \Delta G_{npol} - T\Delta S \quad (\text{Eq 7})$$

Where ΔE_{int} , ΔE_{el} and ΔE_{vdw} are the change in internal, electrostatic and vdWs energies upon binding respectively. ΔG_{pol} is the polar contribution to the solvation free energy and is obtained by solving either the Generalised Born (GB) or Poisson–Boltzmann (PB) equations. ΔG_{npol} is the nonpolar contribution to the solvation free energy. Finally, $T\Delta S$ is the change in entropy of the solutes upon binding, which in combination with ΔE_{int} , ΔE_{el} and ΔE_{vdw} form the parts of the equation responsible for gas phase contributions to ΔG_{bind} . The following sub-sections below discuss how each term is obtained.

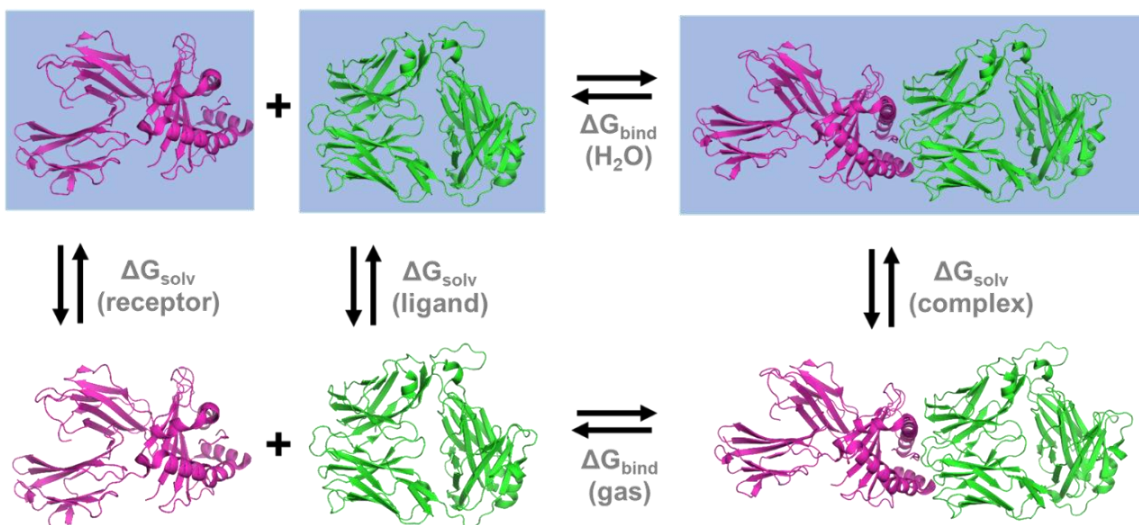


Figure 10: Schematic Representation of the thermodynamic cycle used in MMPB/GBSA calculations. (The top three molecules are depicted inside a blue background to indicate they are solvated, whilst the bottom three are depicted as being in the gas phase). The model system shown is a TCR-pHLA complex, with the pHLA shown as the receptor and the TCR shown as the ligand.

2.5.2 Calculation of Internal Energy, Electrostatic and van der Waals Terms

In the case of ΔE_{int} , ΔE_{el} and ΔE_{vdw} , these terms can be obtained directly from the chosen force field potential energy terms (ΔE_{int} in this case covers all bonding, angle, dihedral and improper torsions terms). These terms can be provided by a single point energy calculation on the given snapshot/structure.

2.5.3 Calculation of the Solvation Free Energy

The polar contributions to the solvation free energy are obtained most commonly by solving either the Poisson–Boltzmann (PB) or Generalised Born (GB) equations (other methods have also been used such as 3DRISM⁸¹). Both the PB and GB models are “implicit” as instead of explicitly treating the solvent molecules as particles, they are described by a continuum model (CM).⁶⁶ In a CM, the solute polarises the continuum (from the partial charges on its atoms) to build up charge density at the border between the solute and continuum. The charge density will be of opposite sign to the local solute charge, with the magnitude of the charge density regulated by both the magnitude of the partial charges in the solute and the dielectric constant (i.e. the polarity) of the solute. If

one can calculate the charge density over the entire surface of the molecule ($\sigma(r)$), it can be used to calculate the polar solvation free energy of the molecule (G_{pol}), see (Eq 8).

$$G_{pol} = \frac{1}{2} \int_0^s \sigma(r) \cdot V(r) dr \quad (\text{Eq 8})$$

Where $V(r)$ is the electrostatic potential at position r on the surface of the molecule. The integration is performed over the entire surface of the calculation, with a reduced integration grid spacing resulting in increased accuracy but increased computational cost. To obtain $\sigma(r)$ one can solve either the PB or GB equation. The solution to the PB equation is shown in (Eq 9).⁸²

$$\sigma(r) = \frac{1}{4\pi} \left(1 - \frac{1}{\varepsilon(r)}\right) \nabla V(r) \cdot n(r) \quad (\text{Eq 9})$$

Where $\varepsilon(r)$ is the position dependant dielectric constant and $n(r)$ is a unit vector (vector with magnitude of one) perpendicular to the solute surface.

The GB equation approximates the PB equation, by treating the point charges exposed to the surface individually before summing the resulting terms together to obtain G_{pol} (see (Eq 10)).⁶⁶

$$G_{pol} = -\frac{1}{2} \left(1 - \frac{1}{\varepsilon(r)}\right) \sum_i^{atoms} \sum_j^{atoms} \frac{q_i q_j}{\sqrt{R_{ij}^2 + r_i \cdot r_j} \cdot e^{(-R_{ij}^2/4r_i \cdot r_j)}} \quad (\text{Eq 10})$$

Where q_i and r_i is the charge and radii of atom i respectively. R_{ij} is distance between the atoms i and j .

Both the PB and GB models calculate the electrostatic contribution to the solvation free energy but do not consider the role of the solute in organising/ordering water molecules (or solvent more generally) to form a hydration layer over the protein. This phenomenon (referred to commonly as the hydrophobic effect) is what is calculated in the term " G_{npol} " and can be estimated by assuming the solvent accessible surface area (SASA) of the protein has a linear relationship with G_{npol} .⁸³

2.5.4 Calculation of Solute Entropy

The final term left to calculate is $T\Delta S$, the change in solute entropy (solvent entropy contribution is calculated as discussed in Section 2.5.2). Several methods now exist to

calculate this term, with the most commonly used approach being normal mode analysis (NMA). This approach is computationally expensive in terms of the time taken per snapshot, as well as being memory intensive to store the Hessian. The time taken is further exacerbated by the need to optimise the snapshots to its local minimum prior to NMA (otherwise the first derivative of the potential energy with respect to the atomic coordinates will not be zero). It has often been observed that the variation between snapshots from NMA can be quite large, meaning that obtaining meaningful results requires a lot of sampling and the use of lots of snapshots. For the above reasons, alongside the fact that MMPB/GBSA is often primarily used for relative (not absolute) binding free energies, explains why many studies skip the calculation of $T\Delta S$ altogether.^{84,85} One approach developed by Ryde and co-workers to deal with some of the limitations associated with NMA is to truncate the protein around the binding site for the optimisation and frequency calculations, whilst retaining a small buffer region that is fixed in place to stabilise the protein over these steps. This approach can significantly reduce the time taken to run NMA as well as the errors obtained.^{86,87}

An alternative approach is to calculate the mass-weighted covariance matrix from a simulation trajectory or trajectories. This approach is known as “quasi-harmonic analysis” (QHA), with the eigenvalues obtained from the covariance matrix corresponding to the vibrational frequencies of the system.⁸⁸ A severe limitation of this approach is the length of time required in order to converge, which is likely to be on the μs or ms time scale for biomolecules.⁸⁹ One way to increase the rate of calculation could be to strip the hydrogen atoms from the calculation (reducing the number of modes and therefore time taken to converge). The vibrational modes of bonds containing hydrogen are relatively high ($\sim 3000\text{ cm}^{-1}$), meaning they will have a small impact on the total entropy value obtained.

A recently developed approach by Duan *et al.* known as the interaction “Interaction Entropy” (IE) method has gained popularity due to the fact that separate calculations are not required to calculate $T\Delta S$.⁹⁰ In this method, the variance (over the course of a simulation(s)) of the gas phase interaction energy (E_{el} and E_{vdw} terms) is used to calculate the entropy of the given state.

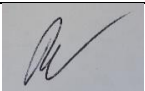
2.5.5 Practical Considerations When Performing MMPB/GBSA Calculations

The approximations associated with the MMPB/GBSA approach mean it is not generally considered reliable for accurately calculating absolute binding free energies.⁸⁵ Instead, MMPB/GBSA can be used to rank or predict relative binding free energies for a series of similar ligands. Each calculation is performed on a single snapshot/conformation at a time making the procedure well suited for parallelisation.³⁹ Whilst MMPB/GBSA can be performed on a single structure, it is generally accepted that sampling (normally via MD simulations) will improve the reliability of the results obtained.⁸⁵ In the case of performing

MD simulations, several studies have shown that running multiple short simulations achieves more (statistically) reliable and accurate results as opposed to one or a few long MD simulations.⁹¹⁻⁹³ The benefit of taking MD simulations snapshots more often than every 10 ps appears to be limited of value, as inclusion of intermediary frames are unlikely to be decorrelated from one another.⁹¹ MMPB/GBSA calculations should theoretically be performed in vacuo as solvent contributions are calculated implicitly. However, as both PB and GB methods represent bulk solvent behaviour (and therefore may not particularly well describe the first few solvation layers) some studies have tested the effect of including a small shell of explicit water molecules around the binding site, with mixed success.⁹⁴⁻⁹⁶ Finally, a common approximation used in MMPB/GBSA is to perform simulations of only the complex and obtain structures for the receptor and ligand by deleting either part as appropriate. This procedure (referred to as the single trajectory approach) makes the ΔE_{int} term cancel out, reduces the calculation time significantly (as less MD simulations needs to be run) and tends to result in a reduction in the magnitude of the error values.^{84,85,93}

Chapter 3: Peptide Dependant Allosteric Regulation of the pHLA Complex

In this chapter, we applied pressure/temperature perturbation experiments, and molecular dynamics (MD) simulations, to explore how different peptide cargo could govern the molecular flexibility of the overall pHLA complex. We found that the motions of the pHLA were dependent on and mediated by (relatively) small alterations in the peptide cargo, suggesting that allosteric mechanisms can mediate HLA flexibility. Differences in protein dynamics for different peptide cargo manifested primarily in the HLA molecule (as opposed to the peptide). Changes in flexibility were found in several regions known to be involved in antigen processing (via tapasin, TAPBPR interactions) and CD8 co-receptor interactions, calling into question the role of the peptide in mediating these interactions. Additionally, these motions might directly influence TCR-mediated antigen discrimination through flexibility motions transmitted to the HLA-binding groove. This work is of significant interest to those studying the molecular mechanisms that govern T-cell mediated antigen recognition, as well as the growing community of researchers and companies developing T-cell based therapies (vaccine design, CARs, bi-specifics, etc). My role in this chapter was in performing all of the MD simulations and subsequent analysis. I also wrote the results and discussion for the MD simulations.

This declaration concerns the article entitled:			
Peptide cargo tunes a network of correlated motions in human leukocyte antigens			
Publication status (tick one)			
draft manuscript	<input type="checkbox"/>	Submitted <input checked="" type="checkbox"/>	In review <input type="checkbox"/>
		Accepted <input type="checkbox"/>	Published <input type="checkbox"/>
Publication details (reference)	<p>Jade R. Hopkins*, <u>Rory M. Crean*</u>, Dragana A. M. Catici, Andrew K Sewell, Vickery L. Arcus, Marc W. Van der Kamp, David K. Cole, and Christopher R. Pudney. (Submitted to FEBS Journal).</p> <p>* = Equal contribution.</p>		
Candidate's contribution to the paper (detailed, and also given as a percentage).	<p>The candidate contributed to/ considerably contributed to/predominantly executed the...</p> <p><u>Formulation of ideas (25 %)</u> Computer simulation experiment ideas was formulated by RMC, MvdK and CRP. Laboratory experiment ideas were formulated by JRH, DAMC and CRP.</p> <p><u>Design of methodology: (40 %)</u> Methodology for simulations was formulated by RMC, and MvdK. Experimental methodology was formulated by JRH, DAMC and CRP.</p> <p><u>Experimental work: (60 %)</u> MD simulations and analysis were carried out by RC. All experiments were carried out by JRH, DAMC.</p> <p><u>Presentation of data in journal format (45 %)</u> Computational figures and results and discussion by RMC. Intro by JRH, CRP and DKC. Experimental figures and results and discussion by JRH and CRP. SI formulation by RMC. All authors edited and reviewed manuscript.</p>		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		Date	31/07/2019

Peptide cargo tunes a network of correlated motions in human leukocyte antigens

Jade R. Hopkins,^{‡,1} Rory M. Crean,^{‡,2,3} Dragana A. M. Catoci², Andrew K Sewell,¹ Vickery L. Arcus⁵, Marc W. Van der Kamp,^{*,6} David K. Cole,^{*,1,7} and Christopher R. Pudney^{*,2,4}

¹Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK.

²Department of Biology and Biochemistry, ³Doctoral Training Centre in Sustainable Chemical Technologies, ⁴Centre for Therapeutic Innovation, University of Bath, Bath, BA2 7AY, UK.

⁵School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand

⁶School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK.

⁷Present address: Immunocore Ltd., Abingdon OX14 4RY, UK

*Correspondence: Dr Marc W. van der Kamp, Email: marc.vanderkamp@bristol.ac.uk, Dr David K. Cole, E-mail: david.cole@immunocore.com, and Dr Christopher R. Pudney, E-mail: c.r.pudney@bath.ac.uk.

[‡]These authors contributed equally.

Keywords. Cancer immunotherapy, peptide-human leukocyte antigen (pHLA), vaccine, T-cell receptor (TCR), fluorescence, pressure, molecular dynamics, allostery

Abstract

Most biomolecular interactions are thought to increase the (local) rigidity of a complex, and this paradigm is applied when designing new drugs. Here, we focussed on the Human Leukocyte Antigen (HLA), which plays a crucial role in the adaptive immune system by presenting peptides for recognition by the $\alpha\beta$ T cell receptor (TCR). The role that the peptide plays in tuning HLA flexibility during TCR recognition is potentially crucial in determining the functional outcome of an immune response, with obvious relevance to the growing list of immunotherapies that target the T-cell compartment. We have applied high-pressure/temperature perturbation experiments, combined with molecular dynamics simulations, to explore the drivers that affect molecular flexibility for a series of different peptide-HLA complexes. We find that different peptide sequences affect peptide-HLA flexibility in different ways, with the peptide cargo tuning a network of correlated motions throughout the pHLA complex, including in areas remote from the peptide binding interface, in a manner that could influence T cell antigen discrimination.

Introduction

The T-cell receptor (TCR), expressed on the surface of T-cells, scans for antigens on the surface of virtually every cell in the body. TCR-antigen recognition can mediate clearance of germs and neoplasms, and plays a major role in autoimmunity and transplantation.^{44,97-100} As such, a better understanding of the molecular determinants that govern TCR-antigen interactions is key to identifying novel therapeutic interventions that can enhance (cancer immunotherapy, vaccines), or inhibit (regulation of autoimmunity) T-cell activation. The natural TCR ligands are the peptide-human leukocyte antigens (pHLA) class I and class II. Classically, pHLA class I is recognized by CD8+ T-cells, and pHLA class II is recognized by CD4+ T-cells. These ligands feature a number of unique characteristics (analogous in both the pHLA class I and pHLA class II systems) that have important implications for both protein dynamics and T-cell mediated immunity. First, the antigen binding site is composed of a composite that includes the HLA-binding groove (formed by the HLA $\alpha 1$ and $\alpha 2$ domains for HLA class I, the focus from hereon in) and a short 9-13 amino acid peptide that can be derived from a completely unrelated protein (the source of these peptides is generally the immune-proteasome that degrades the majority of intracellular proteins, which can derive from foreign or mutated self-proteins).^{101,102} Intriguingly, although the peptide only accounts for ~2% of total amino acids in the pHLA, its position within the binding groove ‘pins’

the entire complex together, i.e. HLA molecules do not generally form a stable structure without a bound peptide.¹⁰³ These peptides are edited by the antigen processing machinery in the endoplasmic reticulum (ER) before being transported to the cell surface for TCR interrogation.¹⁰⁴ Second, during binding, the TCR interacts directly with both the HLA surface and the peptide (composite antigen binding site).^{105,106} How the TCR retains the delicate balance between HLA binding and peptide dependence (peptide-independent recognition of HLA would result in T-cell activation against virtually every nucleated cell in the body) is still not fully understood. Finally, pHLA is unique in biology because it can form a trimeric complex with both the TCR and co-receptor molecules (CD8 for pHLA class I and CD4 for pHLA class II).^{107–110} Although the co-receptors bind to an invariant site distal from the TCR, this interaction is known to play a role in TCR thymic selection¹¹¹, and can tune TCR cross-reactivity by altering T-cell potency.^{112,113}

Many studies have focused on understanding the relationship between the biophysical characteristics of the TCR-pHLA interaction and T-cell potency^{18,23,114–117}, and the role of TCR flexibility during pHLA engagement.^{25,99,126,118–125} These studies have demonstrated that the optimal TCR-pHLA interactions can be mediated by a highly flexible binding mode, probably contributing to the ability of TCRs to recognize multiple different pHLAs.^{16,23,127–130} This flexibility has been observed in the flexible loops that form the binding site of the TCR, contributing towards the notion that TCRs ‘meld’ around the pHLA surface during binding.¹³¹ Although flexibility has also been reported in both the HLA-bound peptide^{50,132–135} and the HLA helices^{136–138}, the role that different peptides play in modulating HLA dynamics globally, and what impact the dynamics might have on T-cell antigen recognition, is only beginning to be explored.^{139,140} On the one hand, a more dynamic pHLA molecule could enable TCR binding of the peptide cargo in an ‘optimal’ conformation for T-cell activation, or to enable recognition by a greater range of different TCRs. On the other hand, a more dynamic pHLA may confer a higher entropic cost during TCR binding that might reduce affinity, or could lead to the unwanted recognition of self-antigens leading to autoimmunity.

In the context of protein-protein and protein-ligand interactions, molecular flexibility is defined by a multi-dimensional free energy landscape (FEL), comprising a large number of energetic minima and maxima that define differently stable conformational sub-states of the same protein (or protein complex). Peptide-dependent effects on HLA dynamics could influence the functional interaction between TCR and pHLA, as well as other molecules known to interact with these

receptors. We hypothesized that nature of the peptide cargo might alter the equilibrium of conformational states that exist and are accessible to the HLA molecule (i.e. its FEL). To address this hypothesis, we use combined pressure/temperature (p/T) dependent fluorescence spectroscopy and molecular dynamics (MD) simulations to expose differences in the thermodynamics of the differing pHLA complexes and to identify the atomistic determinants of pHLA flexibility. These data provide new insights in the role that the peptide plays in tuning the flexibility of HLA, a feature that might contribute to modulation of TCR antigen recognition and T-cell mediated immunity.

Results and Discussion

Pressure-temperature matrices expose differing thermodynamic contributions to pHLA flexibility. We focused on the well characterized 1E6 TCR system, a TCR that naturally recognizes the HLA-A*02:01-restricted ALWGPDPAAA₁₅₋₂₄ peptide from the preproinsulin protein, and plays a biological role in human type 1 diabetes.^{98,141–143} We have previously reported a number of altered peptide ligands (APLs) for the 1E6 TCR using structural, biophysical and cellular analysis. These data demonstrated that, despite a highly conserved, hotspot driven, binding mode (**Figure 11A**), the binding affinity and cellular potency of the 1E6 TCR for the different APLs was substantially affected, independently of pHLA stability (**Figure 11B**).²³ Thus, this well characterized set of APLs provided a biological relevant model system to further examine the contribution of the antigenic peptide on HLA flexibility.

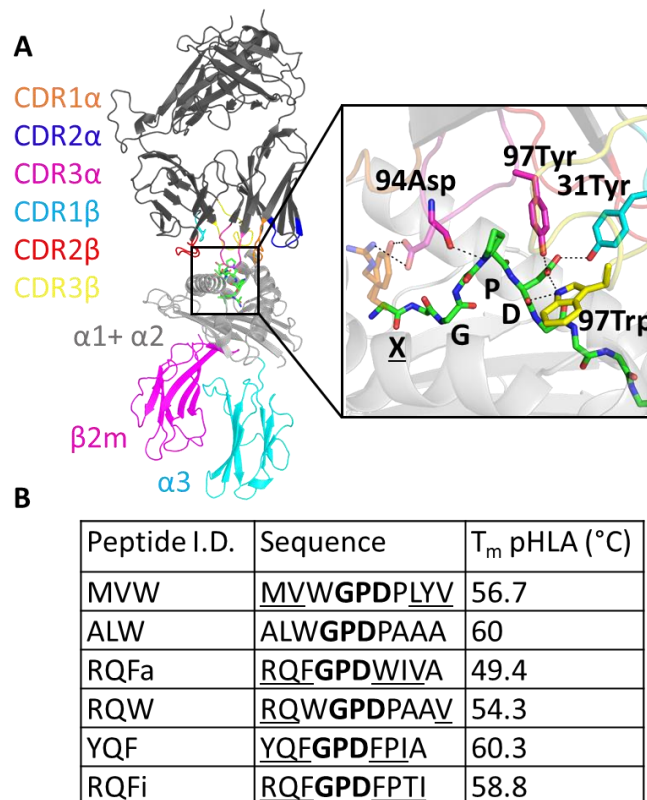


Figure 11: (A) Structural representation of the 1E6 TCR-pHLA interaction with the inset showing a zoom in on the TCR-pHLA binding site. The conserved GPD motif in the peptide and TCR residues that interact with these residues are shown as sticks. For the 4 peptides that also interact with the TCR via their position 1 residue (X), the side-chain of this residue is also indicated with orange sticks. (B) Peptide sequences and their respective pHLA melting temperatures (T_m) as determined by CD spectroscopy (Reported previously, Cole *et al.* 2016).

Molecular flexibility is usefully thought of as the transitions between different conformational states (energetic minima) on the protein FEL. Combined pressure/temperature (p/T) denaturation studies have been used in several cases to extract the complete suite of thermodynamic parameters that define the FEL for protein folding, so called elliptical phase diagrams.¹⁴⁴ In the present study we wish to explore the FEL specifically relating to native protein conformational change. Non-denaturing hydrostatic pressure is an excellent probe of native protein dynamics since it acts by perturbing the pre-existing equilibrium of states, favouring more compact conformations.¹⁴⁵ Non-denaturing pressure therefore gives access to the conformational changes that are natively accessible on the proteins FEL. Intrinsic Trp emission is a ready reporter of the effect of p/T perturbation because Trp emission intensity is sensitive to changes in the immediate molecular environment through a range of mechanisms.¹⁴⁶ Changes in Trp emission are therefore reporters of protein conformational change and as such can be used to calculate an equilibrium constant for

the change across a perturbation series. The change in Trp emission can be converted to an equilibrium constant, see (Eq 11).

$$\frac{F_i}{\sum F} (p, T) = \frac{K(p, T)}{1 + K(p, T)} \quad (\text{Eq 11})$$

Where F_i is the integral of the emission intensity of Trp for a given pressure/temperature. For a simple 2 state transition, e.g. an equilibrium between 2 conformational sub-states, the temperature dependence of the equilibrium constant is given by (Eq 12).

$$\ln K = (-\Delta G)/RT \quad (\text{Eq 12})$$

The combined p/T dependence of ΔG reflects the free energy difference between the 2 notional sub-states and so is a proxy for the degree of conformational ‘flexibility’. $\Delta G_{p,T}$ is then given by (Eq 13).

$$\begin{aligned} \Delta G_{p,T} = & \Delta G_0 + \Delta V_0(P - P_0) + \Delta\alpha'(P - P_0)(T - T_0) + \frac{\Delta\beta'}{2}(P - P_0)^2 \\ & - \Delta S_0(T - T_0) - \Delta C_p \left[T \left(\ln \left(\frac{T}{T_0} \right) - 1 \right) + T_0 \right] \end{aligned} \quad (\text{Eq 13})$$

Where T_0 is a reference temperature. ΔH , ΔS , ΔG_0 , ΔC_p , ΔV_0 , $\Delta\beta$ and $\Delta\alpha$ reflect the changes in enthalpy, entropy, Gibbs free energy, heat capacity, activation volume, compressibility and expansivity between the 2 notional conformational sub-states that define the equilibrium, respectively. Note that this model assumes both ΔC_p and $\Delta\alpha$ are constant with respect to both pressure and temperature. The model assumes a two-state transition because the model for a more complex number of states would be intractable when fitting the experimental data.

These analyses demonstrated that the p/T relationship clearly differed for different pHLA complexes (**Figure 12, Table 1**). The contribution of the different thermodynamic parameters to the magnitude of ΔG were highly specific for each peptide sequence. For example, with RQFa and RQFi, the contribution from ΔS was large compared to other parameters; for ALW the contribution from ΔC_p was large compared to other parameters. MVW, RQW and YQF peptides had significant contributions from ΔV_0 and $\Delta\beta$, which were not observed for RQFa, RQFi and ALW peptides. These data point to a peptide sequence specific effect on the FEL reflecting HLA conformational flexibility. That is, different peptide sequences affect pHLA molecular flexibility in different ways. Most intriguingly, our data suggest the peptide-HLA interaction is governed by

a complex interplay of a range of different thermodynamic contributions, which do not have an obvious relationship to peptide sequence.

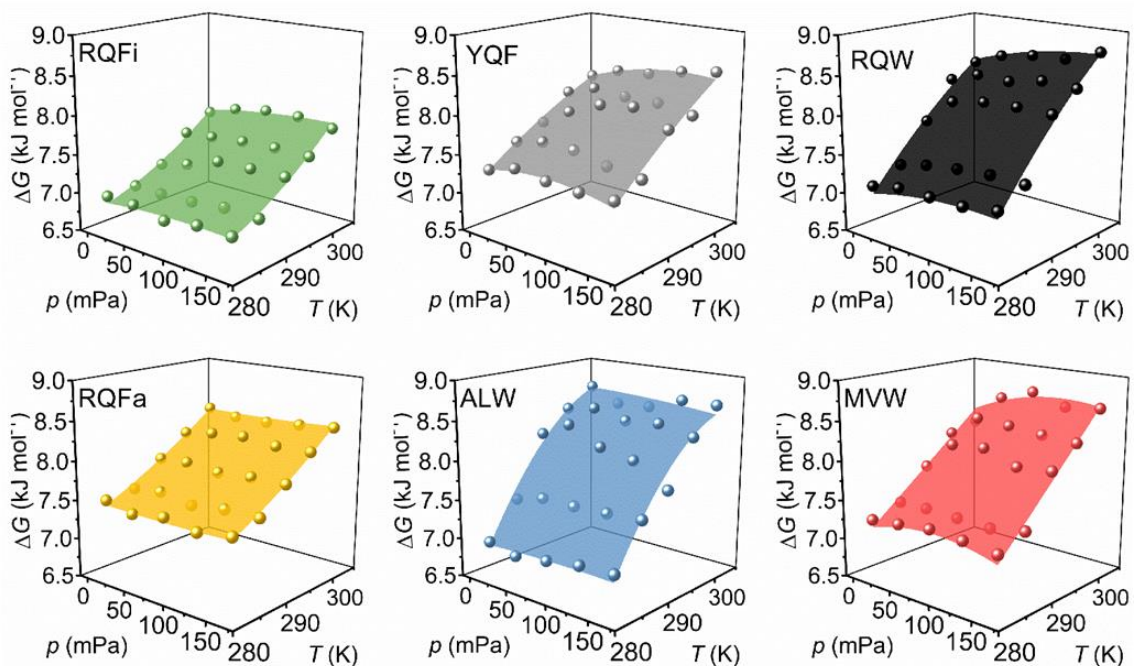


Figure 12: Combined p/T matrices for each pHLA studied. Solid spheres represent the experimental data, transformed to ΔG by (Eq 12). The coloured surfaces are the resulting fit of these values to (Eq 13). Each panel is labelled as the specific pHLA complex.

Table 1: Parameters extracted from surface fits shown in **Figure 12**. T_m values previously published.²³

	T_m (°C)	ΔG_0 (kJ mol ⁻¹)	ΔS (kJ mol ⁻¹)	ΔC_p (kJ mol ⁻¹ K ⁻¹)	ΔV (cm ³ mol ⁻¹) x 10 ⁻³	$\Delta\beta$ (cm ³ mol ⁻¹ mPa ⁻¹) x 10 ⁻⁵	$\Delta\alpha$ (K ⁻¹) x 10 ⁻⁵
RQFi	58.8	8.1 ± 0.1	-0.07 ± 0.6	-0.6 ± 0.2	2 ± 1	-1.3 ± 1.1	3.4 ± 3.5
YQF	60.3	8.3 ± 0.1	-0.02 ± 0.33	0.3 ± 0.3	5 ± 1	-3.6 ± 1.5	8.5 ± 4.9
RQW	54.3	8.7 ± 0.2	-0.04 ± 0.02	0.6 ± 0.4	4 ± 2	2.4 ± 2.1	5.0 ± 6.7
RQFa	49.4	8.8 ± 0.1	-0.06 ± 0.01	-0.3 ± 0.2	0 ± 1	6.3 ± 1.0	1.1 ± 3.1
ALW	60	8.7 ± 0.2	-0.01 ± 0.02	1.7 ± 0.3	0 ± 2	-7.1 ± 2.0	-4.6 ± 6.3
MVW	56.7	8.8 ± 0.2	-0.06 ± 0.02	0.0 ± 0.4	6 ± 2	-4.8 ± 2.4	8.7 ± 7.7

Molecular dynamics simulations identify both local and distal changes in flexibility for different peptide cargo. The p/T analysis demonstrated significant differences in the global molecular flexibility and thermodynamics of the pHLA binary complexes, dependent on the peptide cargo. In an attempt to rationalise the differences observed experimentally, we used molecular dynamics

(MD) simulations, performing 10 replicas of 150 ns each for each pHLA under investigation, giving a total of 9 μ s of simulation time. Using this approach, the (backbone) flexibility of the pHLA can be inferred by calculating the root mean square fluctuations (RMSF) of each residues C α carbon over the course of the simulations.

To investigate the impact of different peptide cargo on the flexibility of the HLA, we first calculated the C α RMSF for the peptide and the α 1 and α 2 domains of the pHLA (**Figure 13, Supp Figure S1**). We then extended this analysis to the α 3 and β 2m domains for each pHLA (**Figure 14, Supp Figure S1**). As we were primarily interested in the differences in pHLA flexibility with different peptides cargos, we calculated the average RMSF value for each residue in all complexes, and subtracted this from each pHLA complex RMSF value, meaning a residue with a positive Δ RMSF value indicates an increased flexibility against the average.

Significant differences ($p < 0.05$) were observed for only 2 of the peptide residues (of 10 total). Interestingly, flexibility differences were not necessarily correlated with regions of the peptide that differed between APLs, demonstrating the interconnected nature of the peptide. For instance, the N-termini had largely very similar flexibility (with the exception of RQW), despite the N-terminal residues differing substantially between peptides. In contrast, Δ RMSF analysis of the central ‘GPD’ motif of peptide (known to be the main binding site for the TCR and conserved in all APLs investigated in terms of sequence and conformation) demonstrated significant changes in flexibility, particularly between RQW and MVW, which showed increased and decreased flexibility relative to the average, respectively. This is of particular interest as the conserved ‘GPD’ motif accounts for 41-50% of all contacts between the 1E6 TCR and the 6 pHLA complexes investigated.²³ These changes in peptide flexibility could, therefore, have a direct impact on the interaction between the TCR-pHLA complex (for instance by modulating the entropic cost of binding).

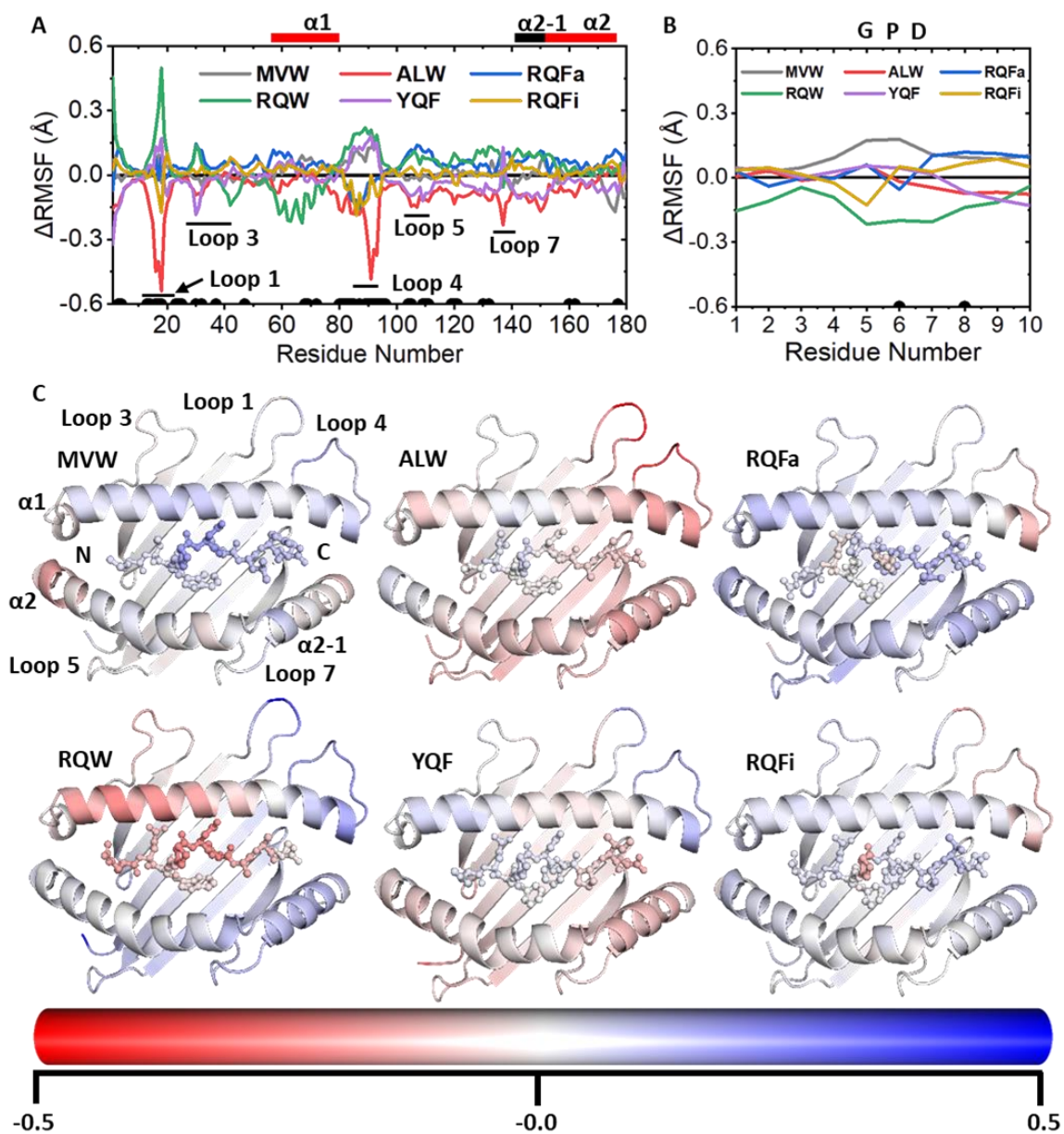


Figure 13: Differences in flexibility at the peptide binding groove for all 6 pHLA complexes investigated. (A-B). Change in C α RMSF (Average – pHLA) for the $\alpha 1$ and $\alpha 2$ domain (A) and peptide (B), meaning a positive Δ RMSF value indicates an increase in rigidity for that pHLA complexes residue relative to the average. C) Δ RMSF values as shown in A-B colour mapped on the pHLA structure (HLA as cartoon, peptide as ball-and-stick), with blue indicating increasing rigidity, and red indicating increasing flexibility (again relative to the average RMSF value for that residue). Heat mapping is scaled from -0.5 – 0.5 Å for all complexes. The black dots towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample t -test ($p < 0.05$). A colour bar is also included for reference.

In contrast to the peptides, which demonstrated relatively small changes in flexibility, significant differences ($p < 0.05$) in flexibility were detected in 45 residues (out of 180 total) for the $\alpha 1$ and $\alpha 2$ helices, 35 residues (out of 97 total) residues in the $\alpha 3$ domain and 54 residues (out of 100 total) in the $\beta 2m$ domain of the HLA. Whilst it is possible that additional HLA residues show differences in flexibility for different peptide cargo, our results demonstrate the importance of performing many replicas and statistical analysis on those replicas to prevent the observation of what may be false positives.¹⁴⁷ The observed significant differences in flexibility were largely confined to the solvent exposed loops in the HLA domains, including loops 3 and 5 in the $\alpha 1$ and $\alpha 2$ domains, respectively. These loops are known to play a role during interactions with tapasin and TAPBPR during peptide editing in the ER.^{148,149} Additionally, loops 8 and 10 on the $\alpha 3$ domain and loop 6 on the $\beta 2m$ domain, which are known to play a role in interactions with the CD8 co-receptor¹⁰⁷, demonstrated high degrees of differences in flexibility. We also note that the statistically significant changes in flexibility we detected in loop 10 in the $\alpha 3$ domain with different peptide cargo are consistent with a previous report demonstrating flexible tuning of this loop during peptide binding.¹⁴⁰ Thus, these changes in flexibility, dependent on peptide cargo, could play a role in tuning the antigen processing pathway, or in modulating the interaction with the CD8 co-receptor, which is known to play a key role in altering T-cell potency and cross-reactivity.^{112,150–152} Overall, differences in flexibility identified by MD analysis were largely observed in the HLA, despite the differences in sequence being confined to the peptide cargo. This unexpected finding of ‘the tail wagging the dog’ may be indicative of allosteric mechanisms in which the sequence of the peptide modulates regions of the HLA known to play a role in different immunological pathways.

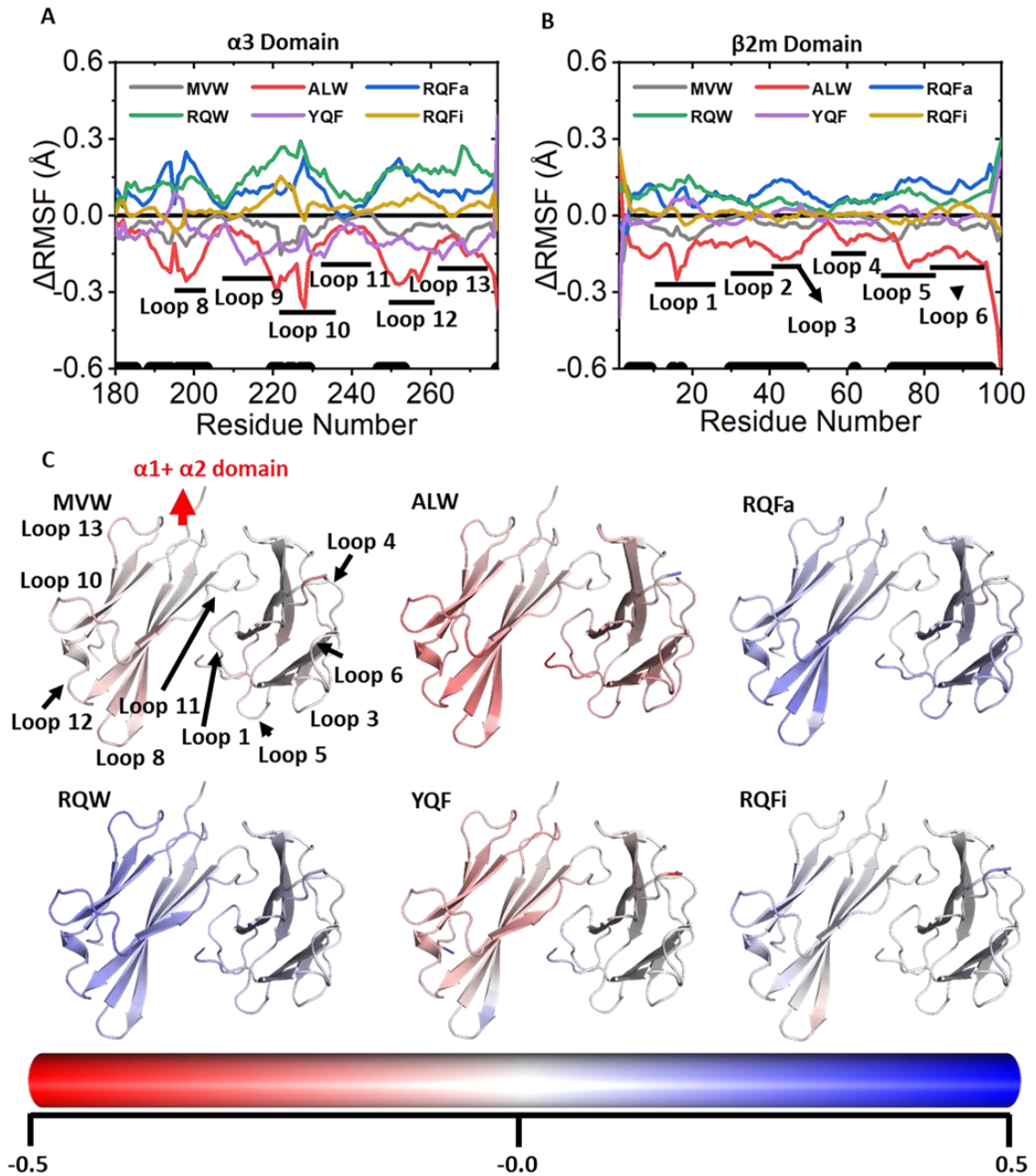


Figure 14 Differences in flexibility at the peptide binding groove for all 6 pHLA complexes investigated. (A-B) Change in C α RMSF (Average – pHLA) for the $\alpha 3$ domain (A) and $\beta 2m$ (B), meaning a positive Δ RMSF value indicates an increase in rigidity for that pHLA complexes residue relative to the average. (C) Δ RMSF values as shown in A-B colour mapped on the pHLA structure, with blue indicating increasing rigidity, and red indicating increasing flexibility (again relative to the average RMSF value for that residue). Heat mapping is scaled from -0.5 – 0.5 Å for all complexes. The locations of the $\alpha 3$ and $\beta 2m$ solvent exposed loops are indicated throughout the figure. The black dots towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample t -test ($p < 0.05$). A colour bar is also included for reference.

Identification of correlated motions between the peptide and HLA. With the observed significant differences in flexibility for regions both local and distal from the peptide binding site in mind, we computed dynamic cross correlation matrices (DCCMs) for all 6 pHLAs under investigation (**Supp Figure S2**). DCCMs measure the degree of correlated motion between each atom (in this case the C α carbon of each residue) over the course of the simulation(s). The measurement assigns a value between +1 (perfectly correlated motion) and -1 (perfectly anti-correlated motion), with 0 indicating no correlation between the residues. This analysis can therefore be used to identify residues distal from one another that are dynamically linked. To focus primarily on the relationship between the peptide and the HLA, we truncated the obtained DCCMs to allow for easier analysis of their relationship (**Figure 15**). Large differences in the overall degree of correlated motion between the peptide and HLA occur, with RQW most strikingly showing a decreased level of correlated motion. Further, a much larger degree of coupling between the C-terminal end of the peptide and the rest of the HLA is observed, as compared to the N-terminus and central portion of the peptide. Thus, these data suggest that the C-terminal residues of the peptide may play a more important role in regulating the global dynamics of the HLA, possibly via the F-pocket of the HLA binding groove. In particular, we observed consistently positively correlated motion between the C-terminal residues of the peptide and the α 1 helix as well as residues 114-134, which make up a large part of the F-pocket. Interestingly, we also observed a consistent change across all pHLA complexes of positively correlated to anti-correlated motion along the α 2 helix (positive starting at the α 2-1 portion of the helix). The degree of correlation between the peptide and domains distal from the peptide binding site (α 3 and β 2m domains) also showed consistent regions of correlated motions for different pHLA complexes. Whilst in the case of RQW, and to a lesser extent RQFa, these correlations were weaker, residues within the range 210-250 on the α 3 domain showed correlated motion to the peptide. These residues include those in loop 10, which we herein, and others¹⁴⁰, observed significant differences in flexibility dependent on the peptide cargo.

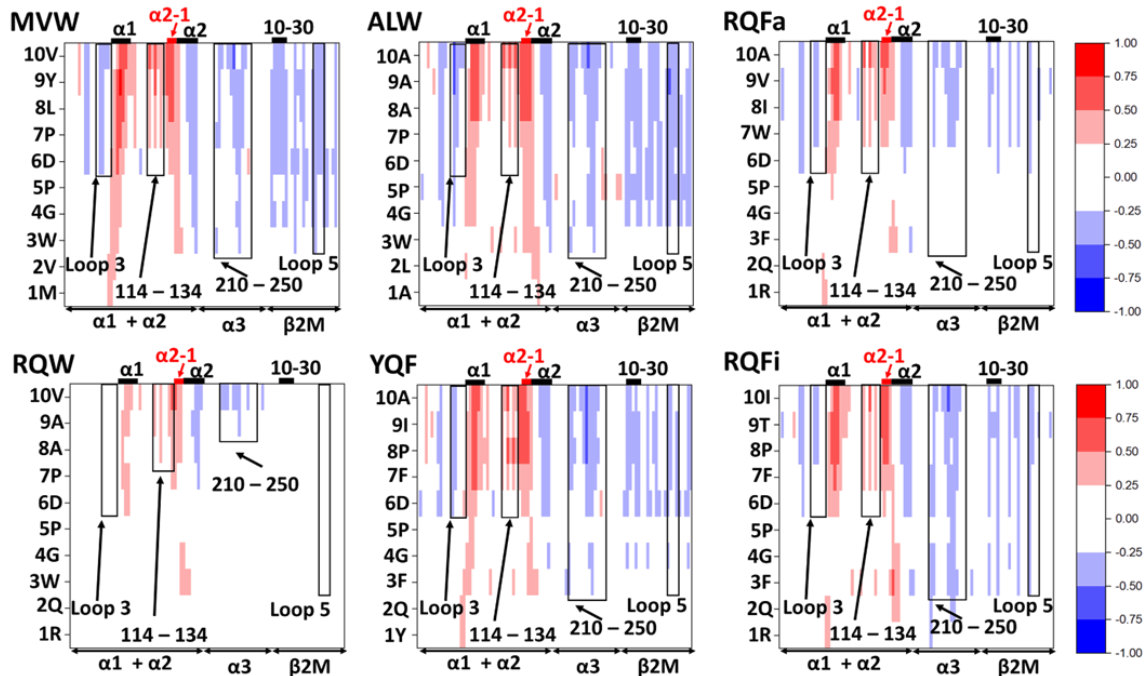


Figure 15: Dynamic cross correlation matrices (DCCMs) for all 6 pHLA complexes simulated. On the Y-axis is each residue of the peptide, which is plotted against all other residues (377 total) on the HLA. The matrices are colour mapped according to the degree of correlated motion between the two residues, with a value of +1 meaning perfectly correlated and -1 meaning perfectly anticorrelated motion. Complete DCCM plots of the pHLA against the pHLA are provided in **Supplementary Figure S3**.

Peptide dependent tuning of the allosteric communication network. The observed differences in flexibility and correlated motions for different peptide cargo in the HLA point to an allosterically linked network across the pHLA complex. With this in mind, we turned to community network analysis (CNA)⁷⁹ to determine the mechanisms by which the peptide communicates dynamical changes to regions distal from the binding site. In CNA, residues are grouped into ‘communities’ of similar dynamics (communities are groups of residues local to one another that share highly correlated motions). The strength of the communication pathway between different communities is determined by the overall amount of correlated motion between members of the 2 given different communities. These data can therefore be represented in graphical form, in which a node corresponds to a community (with the size of the node indicating the number of residues in that community), and edges between nodes indicating the strength of the communication pathway (with an increased thickness indicating increased correlation) (**Figure 16**).

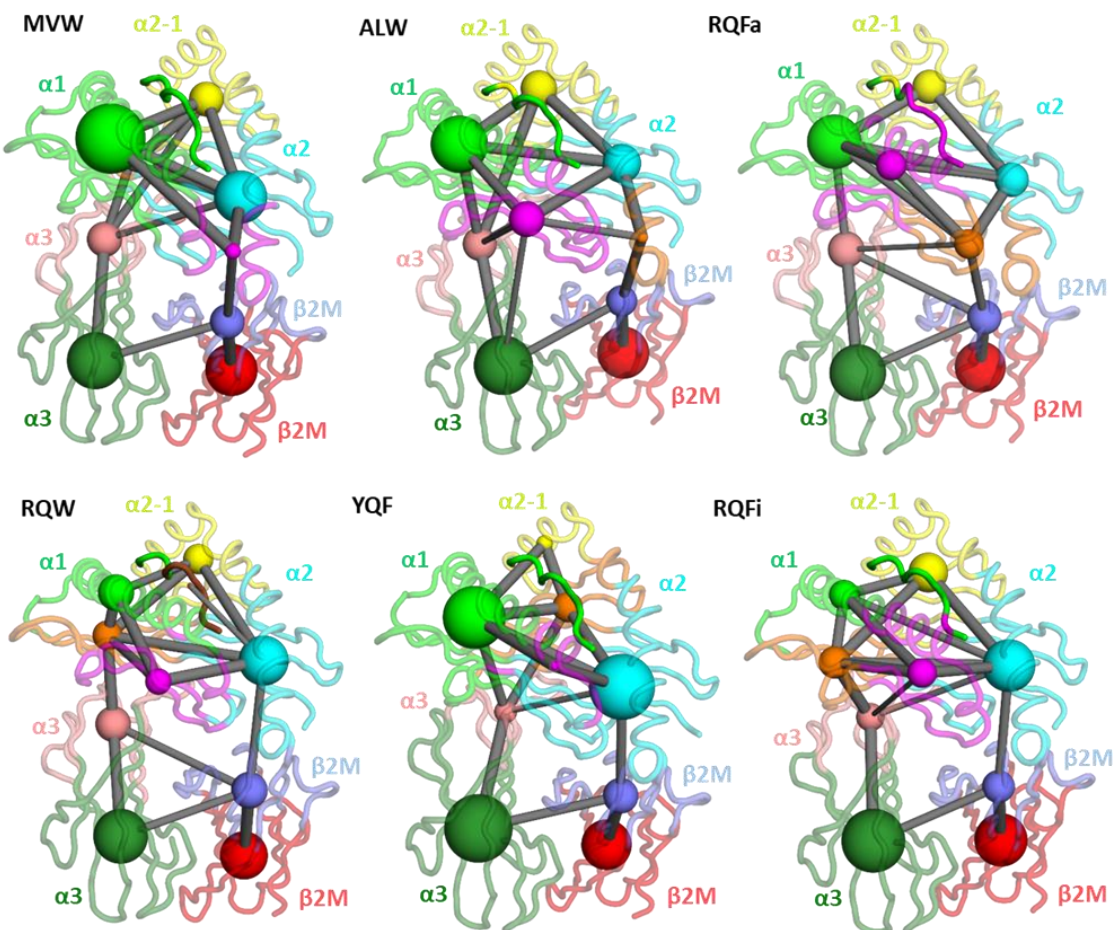


Figure 16: Peptide dependent tuning of the allosteric communication network. Community networks determined for all pHLA complexes studied. Networked communities are shown as coloured spheres, with the radii of the sphere indicating the number of residues within the community. Edges between the nodes/communities represent communication pathways between the nodes, with the thickness of the edge indicating the degree of correlation between the two communities (thicker = greater correlation). All pHLA complexes are shown from the same orientation, such that N-terminus of the peptide is in the foreground. The results here are provided in tabulated form in **Supplementary Table S2**.

The pHLA complexes were partitioned into 9 communities each (chosen based on a consistently high modularity score and an ability to partition the different pHLA complexes as similarly as possible), apart from RQW, which was partitioned into 10 communities because the first 6 residues of the peptide consistently grouped into their own community even at much lower overall community numbers (**Supp Figure S3**). All communities located in the peptide binding groove (peptide and $\alpha 1$ and $\alpha 2$ domains) were highly interconnected to one another. Communication from the binding groove to the $\beta 2m$ domain appeared to occur through a single community, generally located towards the end of the $\alpha 1$ helix and part of the residues that form the F-pocket (peptide C-

terminal binding site). This contrasted with observations made for the top portion of the $\alpha 3$ domain, in which multiple communities in the peptide binding groove showed a significant level of directly correlated communication with residues belonging to the $\alpha 3$ domain. These observations may help to rationalise why the C-terminal peptide residues tended to show a much greater degree of correlated motions with HLA residues. That is, the residues responsible for binding the C-terminal portion of the peptide appear to be dynamically linked to both the $\alpha 3$ and $\beta 2m$ domain, in contrast to the residues responsible for binding the N-terminal portion of the peptide.

Conclusions

Here, we used cutting edge experimental approaches and molecular dynamic simulations to demonstrate that the peptide cargo is able to tune the conformational dynamics of HLA. More specifically, the precise amino acid composition of the peptide cargo differentially engages a network of correlated protein dynamics that spans the HLA. For instance, the C-terminus of the peptide appears to be able to regulate the conformational dynamics of the entire pHLA complex as well as the main TCR-peptide contact zone, potentially modulating TCR binding. Our data point to the peptide cargo having the ability to tune a network of allosteric dynamics in the pHLA complex and may play a role in tuning a number of pathways involved in T-cell mediated immunity. These include peptide editing during antigen processing, interactions with the CD8 co-receptor, and direct TCR interactions with the peptide and HLA. These findings may be pertinent for peptide vaccine design and may help explain why even minor alterations in peptide sequence can completely alter the direction of the immune response.^{134,153,154} Our study also has broader implications for the understanding of protein interaction networks, particularly allosteric mechanisms, in which changes in a relatively small component of the protein complex (in this case a few mutations in a 10 amino acid peptide) can modulate flexibility distal to the changes and throughout the complex (in this case HLA, a 4 domain protein complex made up of nearly 400 amino acids).

Materials and Methods

Protein expression and refolding. HLA-A*02:01 and β 2m were expressed and refolded using competent BL21 DE3 E. coli cells transfected with pGMT7 expression plasmids as previously described.^{155,156} Refolded protein was purified by anion exchange using a Poros50HQ column, followed by size exclusion into phosphate buffered saline using gel filtration column - Superdex™ 200 Increase 10/300 GL. Purification followed previously described protocol.^{155,156} Dynamic light scattering (Zetasizer) was used to confirm sample homogeneity. For pressure/temperature varying fluorescence measurements samples were exchanged into a HEPES buffer (50mM HEPES, 150mM NaCl, pH 7.4) using a PD-10 desalting column containing Sephadex™ G-25 medium, following manufacturer instructions.

Pressure/temperature dependent fluorimetry. Pressure/temperature measurements were performed using an ISS high-pressure cell (ISS, Champaign, IL, USA) fitted with a custom fibre optic mounting connecting to the fluorimeter and the water bath. Peptide-HLA complexes were excited at 295nm, tryptophan emission was measured between 325-500 nm. Emission and excitation slits were set to 15nm to minimise the signal to noise ratio (due to optimal set up of the pressure cell). Initial measurements were made at 10 °C and increased in 5 °C increments up to 30 °C. The pressure dependence at each temperature was measured at 50, 400, 800, 1200, and 1600 bar. Measurements were taken in triplicate. Following each full pressure/temperature range, repeat scans were taken at lower pressure/temperature conditions to ensure extreme the pressure/temperature conditions had not denatured the protein. For all measurements the appropriate buffer controls were subtracted prior to data processing.

MD simulations. Previously solved X-ray crystal structures of the 6 pHLA complexes^{23,98} were used as the starting point for all MD simulations (see **Supp Table S1** for a list of structures used). Any missing residues were added using Modeller v9.¹⁵⁷ PropKa 3.0¹⁵⁸ was used to predict the protonation states of all proteins investigated for a pH 7 (resulting in all residues being simulated in their standard protonation states). MolProbity⁷⁵ was used to determine the optimum tautomerisation states for every His residue (tautomerisation states used for all simulations are provided in **Supp Table S1**) and make any required Asn/Gln side chain flips (under the criteria of optimising the hydrogen bonding network). The results were visually inspected and care was taken to ensure consistency between all pHLAs investigated. All systems were then solvated in an octahedral water box (retaining any crystal waters) such that no protein atom was within 10 Å

of the box boundary. Simulations were performed at an effective [NaCl] of 150 mM (to match experiments), with excess Na⁺ ions added as required to ensure neutrality. MD simulations were performed using Amber16, describing the protein and water molecules with the ff14SB force field⁶⁰ and TIP3P water model¹⁵⁹ respectively. Following a protocol of minimisation, heating and equilibration (see section “Structure Equilibration Procedure” below), all pHLA complexes were subjected to 10 x 150 ns of production MD simulations in the NPT ensemble (at 300 K and 1 atm), with snapshots collected every 10 ps. Production MD simulations were performed using a 2 fs time step and with the SHAKE algorithm applied. An 8 Å direct space non-bonded cut-off was applied with long range electrostatics evaluated using the particle mesh Ewald algorithm.⁷² Temperature was regulated using Langevin temperature control (collision frequency of 1 ps⁻¹), whilst pressure was controlled with a Berendsen barostat (setting the pressure relaxation time to 1 ps).

Structure Equilibration Procedure Upon preparation of all six pHLAs complexes investigated, the following procedure was used to equilibrate structures for production MD simulations performed at 300 K and 1 atm. Minimisation of all hydrogen atoms and solvent molecules (including Na⁺ and Cl⁻), using 500 steps of steepest descent followed by 500 steps of conjugate gradient. To keep all other atoms (i.e. the protein heavy atoms) in place during the minimisation, 10 kcal mol⁻¹ Å⁻¹ positional restraints were applied. Retaining the positional restraints on all protein heavy atoms, the system was then heated rapidly from 50 K to 300 K in the NVT ensemble over the course of 200 ps. This system was again minimised for a further 500 steps of steepest descent followed by 500 steps of conjugate gradient, this time only applying positional restraints (of size 5 kcal mol⁻¹ Å⁻¹) to the C α carbon atoms. These C α restraints were retained as the system was again heated from 25 K to 300 K over the course of 50 ps in the NVT ensemble. Simulations were then performed in the NPT ensemble (1 atm, 300 K), first gradually reducing the 5 kcal mol⁻¹ Å⁻¹ C α carbon restraints over the course of 50 ps of simulation time. This was done in 5 steps (5, 4, 3, 2, 1 kcal mol⁻¹ Å⁻¹) of 10 ps each. A final 1 ns long simulation was then performed in which no restraints were used. The end structure from this run was then used as the starting structure for a production MD simulation. All dynamics steps used the SHAKE algorithm. 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 again used Langevin temperature control (collision frequency of 1 ps⁻¹) and a Berendsen barostat (1 ps pressure relaxation time), with a simulation timestep of 2 fs. Simulations of replicas were

performed by taking the structures after the second minimisation step (and before the second heating step). Replicas were therefore assigned different random velocity vectors on the subsequent heating step.

MD Trajectory Analysis. Routine trajectory analysis was performed with CPPTRAJ.¹⁶⁰ $^{160}\text{C}\alpha$ RMSF calculations were performed for all complexes after discarding the first 10 ns of simulation time (for equilibration) and averaged over each run. RMS fitting was performed to the $\text{C}\alpha$ of stable (over the course of our MD simulations) secondary structure residues of the HLA. We used the following residues for RMS fitting: 4-13, 22-38, 51-54, 58-86, 95-104, 111-127, 134-181, 187-196, 199-209, 215-220 and 242-263 of Chain A (i.e. the $\alpha 1$, $\alpha 2$ and $\alpha 3$ domains), and residues 6-11, 21-30, 36-41, 60-70, 78-83 and 91-94 of Chain B (i.e. the $\beta 2\text{m}$ domain). For both RMSF DCCM and CNA calculations, RMS fitting was first performed to the crystal structure in order to create an average structure. Following this, all snapshots were then re-fitted to the average structure for the subsequent calculations. Dynamic cross correlation matrices (DCCMs) and community network analysis (CNA) were calculated using a combination of the Bio3D¹⁶¹ and igraph¹⁶² libraries within the package *R*. Briefly, all ten independent simulations were combined into a single trajectory, RMS fitting each frame to an average structure of all ten simulations combined. DCCMs were calculated for all 387 x 387 residues in each pHLA before truncating the matrix to show the degree of correlated motion between the peptide and all HLA residues. CNA was performed on the aforementioned complete DCCM results, using a Girvan-Newman clustering protocol⁷⁹ to cluster communities of similar dynamics together. Edges with a correlation score of $< |0.4|$ were discarded prior to clustering. The resulting communities were further filtered using a maximum distance cut-off between pairs of atoms of 8 Å (throughout 100% of the simulation time). Whilst the standard procedure in CNA is to plot the community number that gives the highest modularity, it is also acceptable when comparing multiple similar complexes, to choose a high scoring modularity value that better groups the resulting structures.¹⁶³ The modularity is an overall measure of the level of correlation between community members and non-community members, with a higher score indicating increased intercommunity correlation and decreased intra-community correlation, and therefore a better division of the data. We choose a community number of 10 for RQW and 9 for all other pHLA complexes based on the above criteria and in all cases, the difference between the maximal possible modularity score and the selected community score was no greater than 0.02 (**Supp Figure S3**). A value of 10 was selected for RQW as the N-terminal portion of the peptide consistently grouped to itself even at much lower community numbers (lowest evaluated community number was 4).

Conflict of interest

DKC is an employee of Immunocore LTD. The authors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

JRH, DAMC, DKC and CRP performed and/or directed experiments. RMC and MWvdK performed and/or directed simulations. JRH, RMC, DAMC, AKS, VLA, MWvdK, DKC and CRP analysed data, and critiqued the manuscript. MWvdK, DKC and CRP conceived and directed the project. JRH, RMC, MWvdK, DKC and CRP wrote the manuscript.

Acknowledgements

JRH's PhD studentship is funded by the MRC (MR/N013794/1). RMC's PhD studentship is funded by the EPSRC. AKS is a Wellcome Senior Investigator (WT086716MA). MWvdK is a BBSRC David Phillips Fellow (BB/M026280/1). DKC was a Wellcome Career Development Fellow (WT095767). 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.

Supporting Information for Peptide cargo tunes a network of correlated motions in human leukocyte antigens

Jade R. Hopkins,^{‡,1} Rory M. Crean,^{‡,2,3} Dragana A. M. Catoci², Andrew K Sewell,¹ Vickery L. Arcus⁵, Marc W. Van der Kamp,^{*,6} David K. Cole,^{*,1,7} and Christopher R. Pudney^{*,2,4}

¹Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK.

²Department of Biology and Biochemistry, ³Doctoral Training Centre in Sustainable Chemical Technologies, ⁴Centre for Therapeutic Innovation, University of Bath, Bath, BA2 7AY, UK.

⁵School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand

⁶School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK.

⁷Present address: Immunocore Ltd., Abingdon OX14 4RY, UK

*Correspondence: Dr Marc W. van der Kamp, Email: marc.vanderkamp@bristol.ac.uk, Dr David K. Cole, E-mail: david.cole@immunocore.com, and Dr Christopher R. Pudney, E-mail: c.r.pudney@bath.ac.uk.

‡ These authors contributed equally.

Keywords. Cancer immunotherapy, peptide-human leukocyte antigen (pHLA), vaccine, T-cell receptor (TCR), fluorescence, pressure, molecular dynamics, allostery

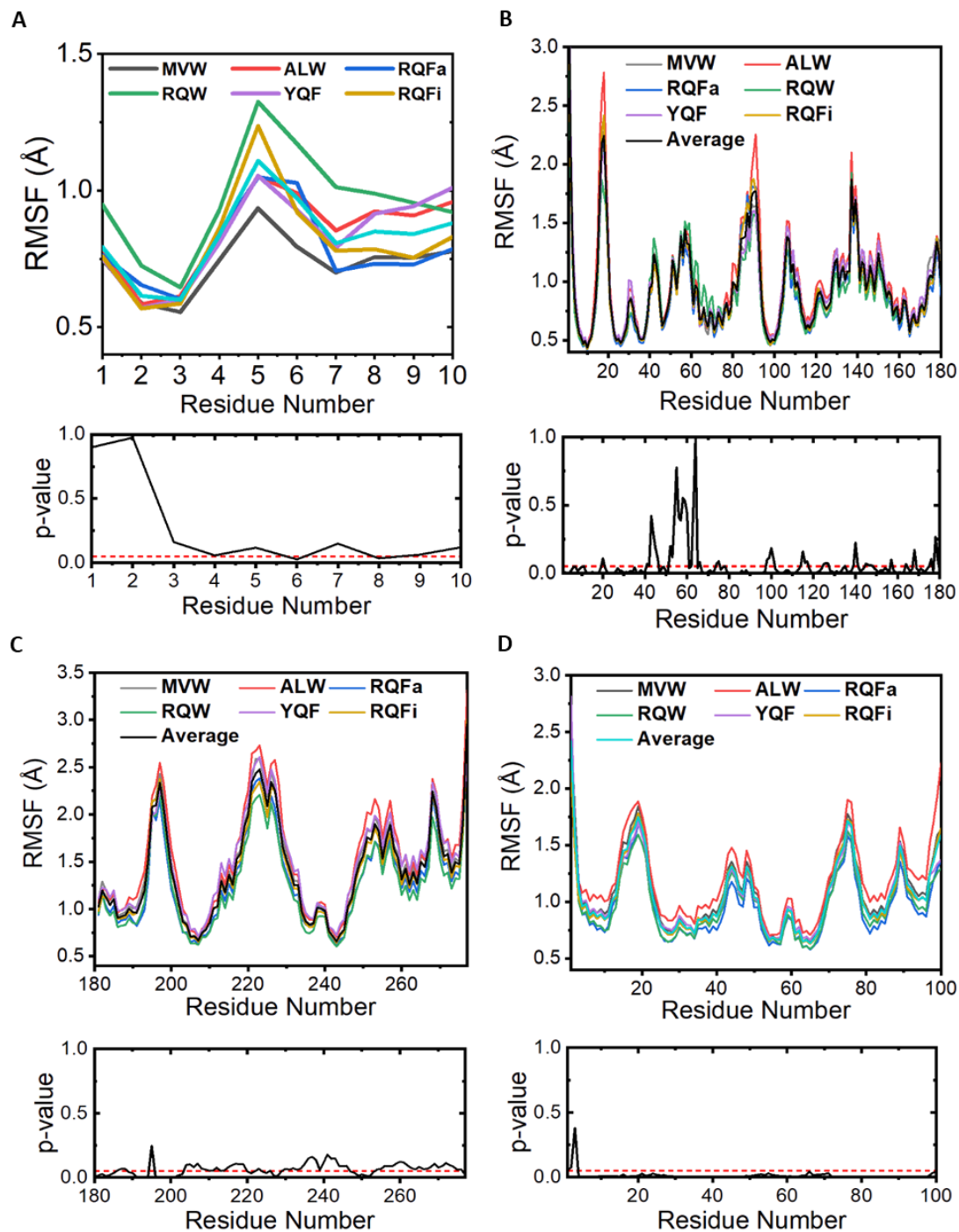


Figure S1: Per residue Ca RMSF values of (A) the peptides, (B) the $\alpha 1$ and $\alpha 2$ domains, (C) the $\alpha 3$ domains and (D) the $\beta 2m$ domains for all 6 pHLA complexes investigated, with the average per residue value also indicated. Below the plot is the p-value obtained from a two-sample t-test between the most and least flexible RMSF value of each residue. A red dotted line is plotted at a p-value of 0.05, which is the cut-off used to determine significance.

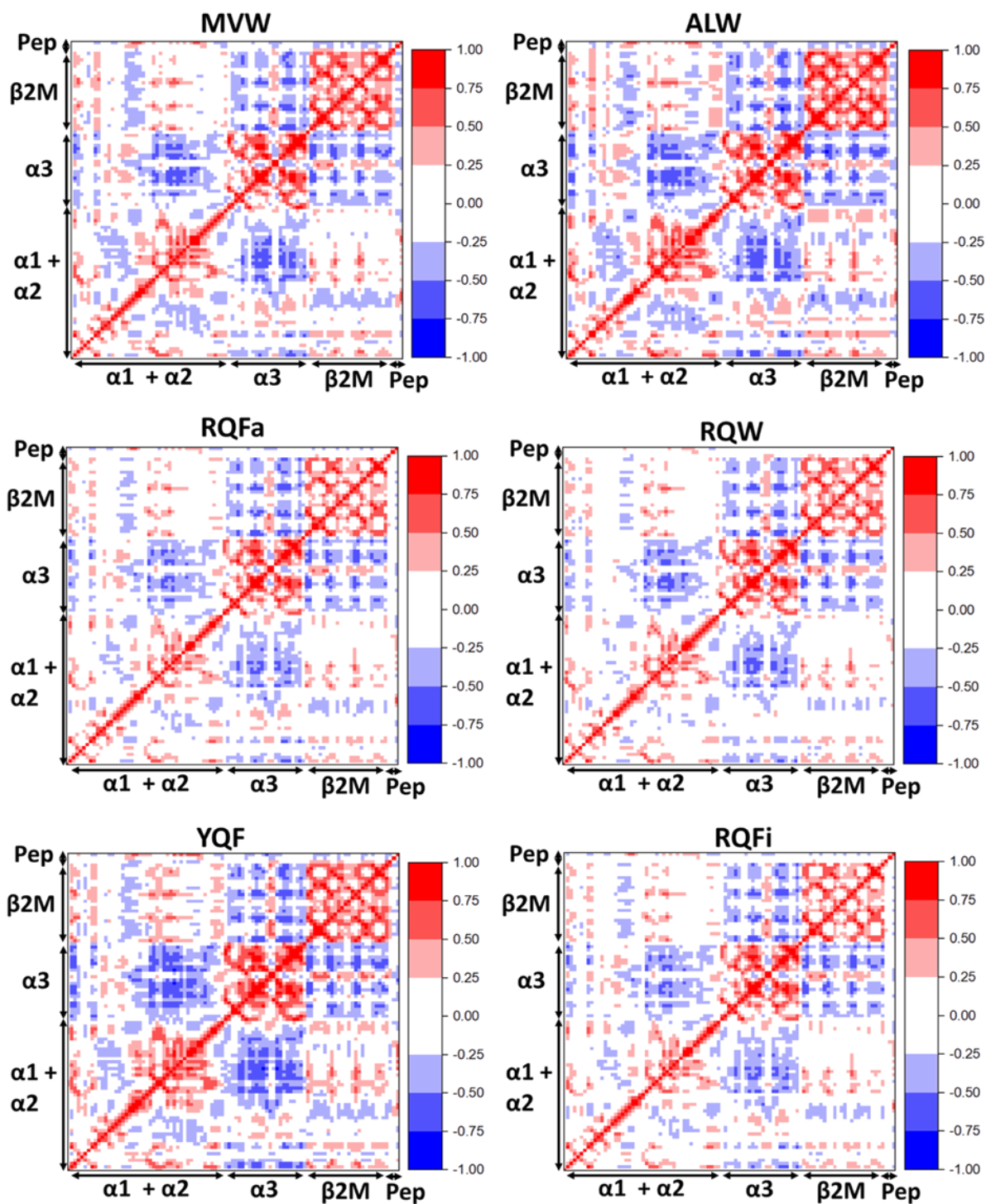


Figure S2: Complete dynamic cross correlation matrices (DCCMs) for all six pHLA complexes investigated. Both x and y-axis consist of 387 residues total (287 HLA, 100 $\beta2m$ and 10 peptide), with regions indicated on the figure. The $C\alpha$ carbon of each residue is used for the measurement. All Matrices are colour mapped according to their degree of correlated motion calculated from the aggregate 1.5 μs of MD simulations per pHLA complex. A value of +1 (red) indicates perfectly correlated motion, whilst -1 (blue) means perfectly anticorrelated motion.

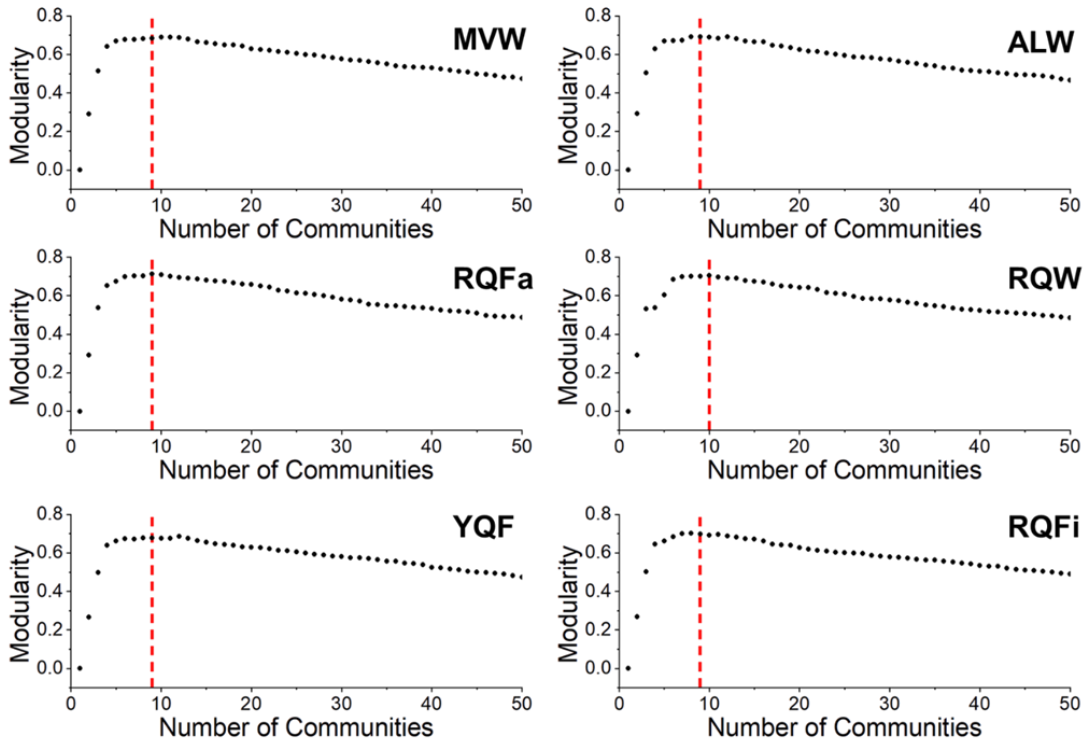


Figure S3: Modularity vs number of communities for each pHLA. The red line on each plot indicates the number of communities chosen for community network analysis as seen in Figure 6 in the main text. In all cases except RQW, this is 9 communities, with RQW set to 10 communities.

Supplementary Table S1. PDB IDs used and histidine tautomerisation state assignments for all MD simulations.

pHLA Complex and peptide sequence	PDB ID	HID^a Tautomerisation State
MVW: MVWGPDPLOYV	5COH	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.
ALW: ALWGPDPAAA	3UTQ	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.
RQFa: RQFGPDWIVA	5COJ	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.
RQW: RQWGPDPAAV	5COF	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.
YQF: YQFGPDFPIA	5COE	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.
RQFi: RQFGPDFPTI	5COI	Chain A: 4, 71, 115, 189, 193, 261. Chain B: 51.

^a, HID corresponds to a histidine residue which is singly protonated on its N δ 1 nitrogen, with all other histidine residues simulated as singly protonated on their N ϵ 2 nitrogen.

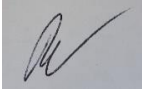
Supplementary Table S2. Community Assignments and total correlation values between communities for all pHLA complexes studied. (This data is represented graphically in **Figure 16** of the main text of the publication). Residues are broken down into their PDB chain IDs (e.g. “A:” corresponds to Chain A).

pHLA Complex	Community Number and Residues Assignments	Correlation Between Communities
MVW	1:(A:1-10+25-35+49-50+98-115+129+158-170) 2:(A:11+95-97+119-122) 3:(A:12-24+36-48+57-94)+(C:1-9) 4:(A:51-56+171-180) 5:(A:116-118+123-157)+(C:10) 6:(A:181-187+205-214+232-244+263-269) 7:(A:188-204+215-231+245-262+270-277) 8:(B:1-7+28-37+54-65+83-92) 9:(B:8-27+38-53+66-82+93-100)	1---2, 0.881, 1---2, 0.881, 1---3, 0.820, 1---4, 0.882, 1---5, 0.898, 1---8, 0.570, 2---3, 0.872, 2---5, 0.908, 2---8, 0.736, 3---4, 0.924, 3---5, 0.900, 4---6, 0.702, 5---8, 0.715, 6---7, 0.965, 6---9, 0.662, 8---9, 0.927.
ALW	1:(A:1-9+29-30+99-115+129+159-170) 2:(A:10-24+63-98+119-120)+(C:1-7+9) 3:(A:25-62) 4:(A:116-118+123-128+130-157+158)+(C:8+10) 5:(A:121-122)+(B:1-7+29-37+55-64+83-92) 6:(A:171-182) 7:(A:183-187+206-215+233-244+263-269) 8:(A:188-205+216-232+245-262+270-277) 9:(B:8-28+38-54+65-82+93-100)	1---2, 0.862, 1---3, 0.904, 1---4, 0.929, 1---5, 0.623, 1---6, 0.879, 2---3, 0.814, 2---4, 0.899, 2---5, 0.819, 3---5, 0.490, 3---6, 0.569, 3---9, 0.500, 4---5, 0.908, 5---9, 0.935, 6---7, 0.790, 7---8, 0.969, 7---9, 0.671.
RQFa	1:(A:1-8+100-114+129+160-171) 2:(A:9-25+69+71-99+115-123)+(C:7+9) 3:(A:26-36+48-56+172-181) 4:(A:37-47+57-68+70)+(C:1-6) 5:(A:124-128+130-159)+(C:8+10) 6:(A:182-187+206-215+232-244+263-270) 7:(A:188-205+216-231+245-262+271-277) 8:(B:1-9+27-37+54-65+84-91) 9:(B:10-26+38-53+66-83+92-100)	1---2, 0.885, 1---3, 0.894, 1---4, 0.464, 1---5, 0.884, 2---3, 0.688, 2---4, 0.861, 2---5, 0.863, 2---8, 0.685, 3---4, 0.826, 3---6, 0.664, 3---8, 0.424, 6---7, 0.947, 6---8, 0.663, 6---9, 0.635, 8---9, 0.905.
RQW	1:(A:1-8+29-32+52+99-115+159-181) 2:(A:9-25+93-98+116-124+338) 3:(A:26-28+33-56+61) 4:(A:57-60+62-92)+(C:7-10) 5:(A:125-158) 6:(A:182-188+205-+214+232-244+262-271) 7:(A:189-204+215-231+245-261+272-277) 8:(B:1-9+27-37+55-60+62-65+83-92) 9:(B:10-26+38-54+66-82+93-100) 10:(C:1-6)	1---2, 0.835, 1---3, 0.901, 1---5, 0.835, 1---6, 0.685, 1---10, 0.453, 2---3, 0.720, 2---4, 0.858, 2---5, 0.902, 2---8, 0.918, 3---4, 0.910, 4---5, 0.689, 6---7, 0.961, 6---8, 0.666, 6---9, 0.652, 8---9, 0.906.
YQF	1:(A:1-8+10+25-36+46-54+101-113+158-181) 2:(A:9+98-100+114-136+154-157)+(B:1-2) 3:(A:11-24+37-44+64-97+1-9) 4:(A:45+55-63) 5:(A:137-153)+(C:10) 6:(A:182-187+206-215+233-244+262-271) 7:(A:188-205+216-232+245-261+272-277) 8:(B:3-7+28-35+54-65) 9:(B:8-27+36-53+66-100)	1---4, 0.858, 1---6, 0.733, 1---8, 0.478, 2---3, 0.907, 2---5, 0.882, 2---8, 0.806, 3---4, 0.863, 3---5, 0.850, 3---8, 0.753, 6---7, 0.974, 6---9, 0.735, 8---9, 0.923.
RQFi	1:(A:1-8+26-34+101-113+158-181) 2:(A:9-25+38-44+93-97+119-122)+(B:338) 3:(A:35-37+45-69)+(C:2) 4:(A:70-92)+(C:1+3-10) 5:(A:98-100+114-118+123-157) 6:(A:182-188+205-213+232-244+264-269) 7:(A:189-204+214-231+245-263+270-277) 8:(B:1-7+28-37+54-60+62-65+85-87) 9:(B:8-27+38-53+66-84+88-100)	1---2, 0.838, 1---3, 0.811, 1---4, 0.636, 1---5, 0.886, 1---6, 0.707, 1---8, 0.495, 2---3, 0.833, 2---4, 0.804, 2---5, 0.901, 2---8, 0.935, 3---4, 0.856, 3---8, 0.408, 4---5, 0.776, 6---7, 0.968, 6---9, 0.702, 8---9, 0.935.

Chapter 4:

Design Principles for Engineering Highly Specific TCRs

The use of soluble TCRs as therapeutics requires high affinities towards the pHLA target and a large therapeutic window to avoid unwanted side-effects. Experimental testing of engineered high affinity TCRs for specificity is both time-consuming and expensive, requiring a large number of cell-based assays. Therefore, a clear understanding at the atomistic level of how to rationally engineer TCRs specific for their targets has clear benefits in the field of rational drug design. In this chapter we applied a combination of structural, biochemical and computational approaches to investigate the molecular rules that define pHLA specificity. In this manuscript we compare three different examples of a TCR and TCR-mimic (antibodies engineered to bind pHLA) targeting the same pHLA. We determine thoroughly characterise their specificity experimentally before using computation to rationalise these observations and suggest design principles for producing highly specific TCRs or TCR mimics. My role in this project was in performing all of the MD simulations and free energy calculations. I also wrote the results and discussion sections pertaining to my obtained results.

This declaration concerns the article entitled:									
Specificity of bispecific TCRs and antibodies targeting peptide-HLA									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input checked="" type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input type="checkbox"/>
Publication details (reference)	<p>Christopher J. Holland*, <u>Rory M. Crean</u>*, Johanne Pentier*, Ben de Wet*, Angharad Lloyd*, Nikolai Lissin, Katy Lloyd, Thomas Blicher,, Chris Pudney, Marc W. van der Kamp, Pierre J. Rizkallah, Bent K. Jakobsen and David K. Cole. (Under review at The Journal of Clinical Investigation).</p> <p>* = Equal contribution.</p>								
Candidate's contribution to the paper (detailed, and also given as a percentage).	<p>The candidate contributed to/ considerably contributed to/predominantly executed the...</p> <p><u>Formulation of ideas (25 %)</u> Laboratory experiment ideas were formulated by CJH, JP, BdW, AL, NL and PJR, BKJ, DKC. Computer simulation experiment ideas was formulated by RMC, MvdK.</p> <p><u>Design of methodology: (25 %)</u> Methodology for simulations was formulated by RMC, and MvdK. Experimental methodology was formulated by CJH, JP, BdW, AL, NL and PJR, BKJ, DKC.</p> <p><u>Experimental work: (35 %)</u> MD simulations and free energy calculations were carried out by RC. All experiments were carried out by CJH, JP, BdW, AL, NL.</p> <p><u>Presentation of data in journal format (35 %)</u> Computational figures and results and discussion by RMC. Experimental figures and results and discussion CJH, JP, BdW, AL and DKC. Formulation of manuscript with the above starting point by DKC. All authors edited and reviewed manuscript.</p>								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						Date	31/07/2019		

Specificity of bispecific TCRs and antibodies targeting peptide-HLA

Christopher J. Holland^{*1}, Rory Crean^{*2,3}, Johanne Pentier^{*1}, Ben de Wet^{*1}, Angharad Lloyd^{1*}, Nikolai Lissin¹, Katy Lloyd¹, Thomas Blicher¹, Stephen Hearty¹, Paul Conroy¹, Miriam Hock¹, Robert J. Pengelly¹, Thomas E. Spinner¹, Brian Cameron¹, Anitha Jeyanthan¹, Peter Molloy¹, Annelise Vuidepot¹, Malkit Sami¹, Milos Aleksic¹, Nathaniel Liddy¹, Ross Robinson¹, Stephen Harper¹, Velupillai Srikanthasan¹, Marco Lepore¹, Chris Pudney², Marc W. van der Kamp⁴, Pierre J. Rizkallah⁵, Bent K. Jakobsen¹ and David K. Cole¹

¹Immunocore Ltd., Milton Park, Abingdon, UK.

²Department of Biology and Biochemistry, University of Bath, Bath, UK.

³Doctoral Training Centre in Sustainable Chemical Technologies, University of Bath, Bath, UK.

⁴School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, UK.

⁵Cardiff University School of Medicine, Heath Park, Cardiff, CF14 4XN, UK.

*These authors contributed equally to this study.

Correspondence: Dr David Cole, E-mail: david.cole@immunocore.com, and Dr Bent Jakobsen, E-mail: bent.jakobsen@immunocore.com.

Keywords: Cancer immunotherapy, TCR-mimic antibody, bi-specific, crystal structure, peptide-human leukocyte antigen (pHLA), surface plasmon resonance (SPR), T cell cross-reactivity, T cell receptor (TCR), high affinity TCR.

Non-standard abbreviations: pHLA, peptide-human leukocyte antigen; RU, response unit; SPR, surface plasmon resonance.

Abstract

Tumour-associated peptide-human leukocyte antigen complexes (pHLA) represent the largest pool of cell surface expressed cancer-specific epitopes, making them attractive targets for cancer therapies. Soluble bispecific molecules, that incorporate an anti-CD3 effector function, are being developed to redirect T cells against these targets using two different approaches. The first achieves pHLA recognition *via* affinity-enhanced versions of natural TCRs (e.g. ImmTAC molecules), whereas the second harnesses an antibody-based format (TCR-mimic antibodies). For both classes of reagent, target-specificity is vital considering the vast universe of potential pHLA molecules that can be presented on healthy cells. Here, we made use of structural, biochemical and computational approaches to investigate the molecular rules underpinning the reactivity patterns of pHLA-targeting bispecifics. We demonstrate that affinity-enhanced TCRs engage pHLA using a comparatively broad and balanced energetic footprint, with interactions distributed over several HLA and peptide side-chains. As ImmTAC molecules, these TCRs also retained a greater degree of pHLA-selectivity, with less off-target activity in cellular assays. Conversely, TCR-mimic antibodies tended to exhibit binding modes focussed more towards hotspots on the HLA surface and exhibited a greater degree of cross-reactivity. Our findings extend our understanding of the basic principles that underpin pHLA selectivity and exemplify a number of molecular approaches that can be used to probe the specificity of pHLA-targeting molecules, aiding the development of future reagents.

Introduction

The ability to selectively target tumour-specific antigens holds great promise for the development of specific cancer treatments, but their identification remains a key challenge. Peptide fragments presented on the cell surface by human leukocyte antigens (pHLAs) represent the intracellular proteome, and because this also includes dysregulated and cancer-specific proteins^{7,164}, pHLAs constitute an important source of tumour-specific antigens. However, targeting these molecules is difficult for two reasons. First, their natural presentation levels can be very low (often below 10 copies of each specific peptide epitope per cell)¹⁹; and second, peptides are co-recognised in the context of HLA, a molecule expressed by most cells (i.e. peptide-selectivity could be lost if HLA interactions dominate the binding interface).^{165,166}

The immune system naturally overcomes these hurdles *via* selective T cell receptor (TCR) recognition of pHLA, enabling T cell triggering towards low-level antigens.^{167–169} Although the mechanisms that determine peptide-selectivity by natural TCRs are not fully understood, the binding mode employed by the TCR is likely to be fundamentally important, evidenced by the conserved binding mode observed for virtually all TCR-pHLA structures solved to date.⁶ This canonical interaction places the TCR diagonally across the HLA binding groove, positioning the somatically rearranged TCR CDR3 loops centrally over the antigenic determinant (peptide) with the germline-encoded CDR1/2 loops positioned primarily over the HLA helices, enabling natural TCRs to detect pHLA in a peptide-dependent manner. Despite the need for precise peptide-selectivity, a limited number of TCRs must still maintain the ability to recognise millions of potential target antigens.^{127,128} Consequently, TCRs have been shown to cross-react with a vast array of different peptides^{16,127,129,130}, but are selected in the thymus to avoid having specificities overlapping with abundant self-epitopes to maintain self-tolerance. Although the mechanisms that underpin these characteristics have yet to be determined, the relatively weak binding affinity of thymically selected TCRs (K_{Ds} in the micromolar affinity range^{18,170}) has been shown to be important for T cell sensitivity¹¹⁶, and is likely also important for maintaining self-tolerance.

The weak affinity of naturally selected TCRs, combined with difficulties manufacturing a membrane-bound protein as a soluble reagent, imposes certain challenges on their use for therapeutic applications. Consequently, the most widely used T cell-based therapies involve the adoptive transfer of either expanded antigen-specific T cells, or T cells genetically modified to express an artificial antigen-specific TCR (specific peptide affinity-enhanced receptor; SPEAR)¹⁷¹ or antibody (chimeric antigen receptor; CAR).¹⁷² Although promising, these therapies are complicated by the need to prepare therapeutic T cells on a patient-by-patient basis and an inability to control dosing in response to potential toxicities.²⁸

Soluble bispecific T cell redirectors, consisting of antigen recognition and T cell engaging domains, bypass many of the limitations of the adoptive transfer approach.¹⁷³ The antigen recognition of pHLA-targeting reagents can be *via* a TCR or antibody domain. ImmTAC molecules (Immune-mobilising monoclonal T cell receptors Against Cancer) are bispecific molecules with an engineered soluble TCR fused to an anti-CD3 effector function³³; thus, it that redirects T cells specifically towards cells presenting a target pHLA.³³ The TCR component of ImmTAC molecules are stabilised with an inter-chain disulphide bond¹⁷⁴ and affinity-enhanced using phage display to generate highly stable, soluble TCR reagents that can bind to pHLA with

low-picomolar affinities, and with binding half-lives of several hours (in comparison to half-lives of seconds for natural soluble TCRs).^{33,175} These attributes enable ImmTAC molecules to elicit anti-tumour responses at picomolar concentrations against cells expressing very low levels of pHLA on the cell surface. In comparison, BiTEs can utilise antibodies to target pHLA (TCR-mimic antibodies) as soluble T cell engaging bispecific molecules.^{34,35,183–186,36,176–182} Antibodies, unlike TCRs that are anchored in the cell membrane, can exist naturally as soluble effector molecules (and as such are easier to engineer as soluble reagents) and typically have a natural strong affinity for their antigen (nanomolar range), making them attractive for development as soluble therapeutics¹⁷³. The main challenge for a targeted pHLA therapeutic is achieving sufficient specificity in the context of a vast landscape of potential self-antigens. For instance, even on individual cell types, data from our in-house mass spectrometry database and published direct evidence demonstrates that the number of unique peptides can be in the range of tens of thousands.^{187–190} Considering the full human protein coding genome, the number of peptides presented has been estimated to be over 11 million.¹⁹¹

In this study, we used a combination of structural, molecular and computational approaches to understand the molecular mechanisms underpinning the pHLA-selectivity and, consequently, the potential cross-reactivity of soluble bispecific T cell redirectors. We demonstrate that utilisation of a native TCR-like binding mode was not predictive of peptide-selectivity. In fact, peptide-selectivity, as defined by lower levels of pHLA cross-reactivity and less off-target activity in cellular testing, was associated with an energetic signature characterised by broad interactions with several peptide side-chains as well as the peptide backbone. These findings have important implications for the underlying rules that determine pHLA discrimination and identify key considerations in the design of immunotherapeutics that target these cell surface proteins.

Results

Structural analysis of pHLA-targeting reagents

We selected TCR-mimic antibodies (TCR-mimics) according to *in vitro* and *in vivo* testing and based on the availability of crystal complex structures to enable molecular analysis (**Supplementary Table S1**). Several additional TCR-mimics have been reported; however, most lacked sufficient published information for inclusion (sequence, structure, and specificity data). As the peptide antigen can have a major influence on specificity (i.e. some peptides may have

close homology to self-peptides), we chose affinity-enhanced TCRs based on their recognition of identical, or closely related, pHLA determinants compared to the TCR-mimics. Here, we assessed reagents designed to recognise the NY-ESO-1₁₅₇₋₁₆₅ cancer testis antigen-derived peptide, SLLMWITQC, presented by HLA-A*02:01 (A2-SLL), the MAGE-A1₁₆₁₋₁₆₉/MAGE3-A3₁₆₈₋₁₇₆ melanoma-associated antigen-derived peptides, EADPTGHSY (A1-EAD) and EVDPIGHLV (A1-EVD), presented by HLA-A*01:01, and the WT1₂₆₋₁₃₄ Wilms tumour antigen derived peptide, RMFPNAPYL, presented by HLA-A*02:01 (A2-RMF). Although not a direct comparison, the A1-EAD and A1-EVD still represented a useful system to include due to the similar tumour expression patterns of both proteins, the same HLA restriction, and similar peptide sequences – a consequence of both peptides representing the same region of the highly related MAGE proteins.

In addition to the previously published crystal structures for TCR-mimic antibodies and affinity-enhanced TCRs in complex with A2-SLL^{34,43,45}, A2-RMF³⁶, and A1-EVD/A1-EAD^{192,193} (**Figure 17A**), we solved the structure of the affinity-enhanced WT1_α7β2 TCR in complex with A2-RMF at 2.8 Å to complete the set (**Table 2**). Together, these data were analysed to identify any structural features that might influence the peptide-selectivity of each reagent. We compared the normal range of binding (crossing and engagement zone) of natural TCRs⁶ with both affinity-enhanced TCRs and TCR-mimic antibodies. The affinity-enhanced TCRs (1G4_α58β61 TCR, MAG-IC3 TCR and WT1_α7β2 TCR) bound within the normal range of natural TCR topologies with the CDR3 loops of both chains focussed over the central peptide bulge (amino acids 4-6) (**Figure 17B**). This binding mode enabled the affinity-enhanced TCRs to form contacts with five of the nine amino acids in the peptide, and multiple interactions with the α1 and α2 domains of the HLA (**Figure 17C**).

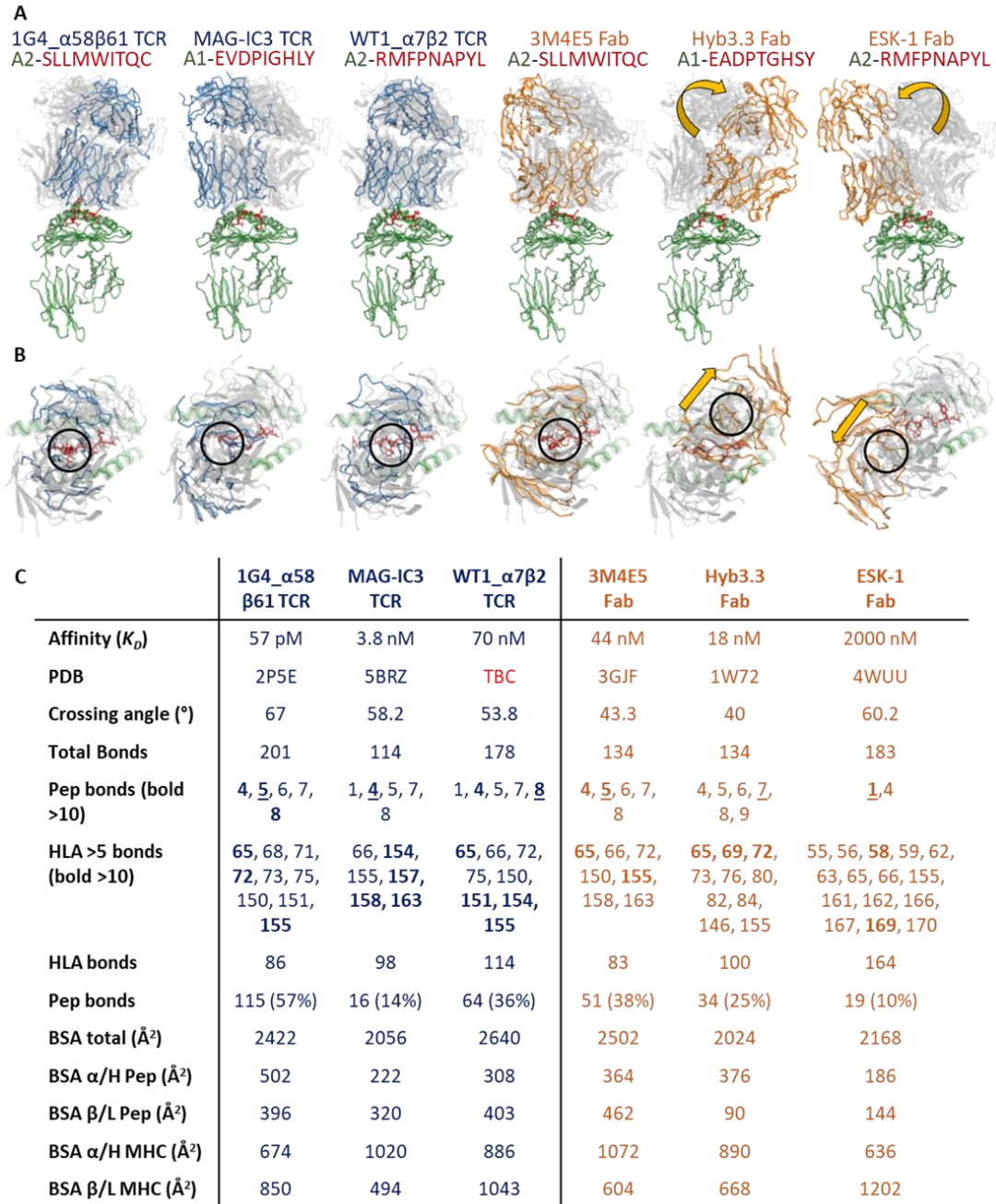


Figure 17: Structural analysis of pHLA-targeting reagents. Structures of the TCRs and TCR-mimics, in complex with pHLA, were analysed to determine the structural mechanism underpinning their binding characteristics. **(A)** TCRs (blue ribbon) or TCR-mimics (orange ribbon) binding to peptide (red sticks) and HLA (green ribbon) compared to the binding range employed by all published wild-type TCR structures (grey cartoon). Yellow arrows indicate unconventional binding modes. **(B)** Coloured as in **A**, top down view of TCR, or TCR-mimic binding to pHLA. Black circles represent the center of binding. Yellow arrows indicate unconventional binding modes. **(C)** Structural analyses of TCRs versus TCR-mimics binding to pHLA. Bonds were assessed using the program contact (CCP4), implementing a 3.4 Å cut-off for H-bonds and a 4 Å cut-off for Van der Waals interactions. Any peptide residue with at least one interaction with the TCR is documented, with peptide residues with >10 contacts shown in bold, and the residue making the most contacts underlined. Any HLA residue with >5 interactions with the TCR is documented, with HLA residues with >10 contacts shown in bold. Buried surface area (BSA) Å² was determined using ePISA. 91

Table 2: Data collection and refinement statistics for WT1_α7β2-A2-RMF and NYBR1-A2-SLS structures. One crystal was used for solving each structure. Figures in brackets refer to the highest resolution bin.

	WT1_α7β2-A2-RMF	NYBR1-A2-SLS
PDB code		6R2L
Data collection		
Space group	P 21 21 21	P 1 21 1
Cell dimensions		
a, b, c (Å)	94.6, 114.8, 185.4	52.2, 99.3, 111.4
a, b, γ (°)	90, 90, 90	90, 90.5, 90
Resolution (Å)	114.75 – 3 (2.95 -2.95)	52.2 – 2.36 (2.36-2.3)
Beam Line		I04-1
Beam time code	In17077-18	14843-1
R _{merge} (%)	32.6 (220.4)	20.7 (108.9)
I / σI	8.8 (1.2)	9.7 (1.3)
Completeness (%)	99.9 (99.9)	99.7 (99.7)
Redundancy	14.1 (14.3)	3.7 (3.6)
No. reflections	700,229 (54,486)	188,042 (13,553)
Refinement		
No R _{free} reflections	2,370	2,435
R _{work} / R _{free}	25.1/29.1	21.4/26.5
R.m.s. deviations		
Bond lengths (Å)	0.006	0.011
Bond Angles (°)	1.253	1.853
Mean B value (Å ²)	75.5	40.0
Wilson B factor (Å ²)	75.8	25.9
Estimated coordinate error based on maximum likelihood (Å)	0.423	0.221

The 1G4_α58β61 and WT1_α7β2 TCRs were more peptide focussed than the MAG-IC3 TCR in terms of the percentage (57%, 36% and 14%, respectively) and number (115, 64 and 16, respectively) of peptide bonds. The 1G4_α58β61 and WT1_α7β2 TCRs also exhibited a larger buried surface areas (BSAs) than MAG-IC3, but all were within, or near to, the normal range (1240 – 2400 Å²)⁶. The 3M4E5 TCR-mimic antibody bound in a very similar fashion to natural TCRs, making contacts with five peptide residues, 38% peptide contacts, and a BSA of 2502 Å². However, the structural analysis revealed focussing of interactions at peptide residue W5, where half of the peptide contacts were concentrated (24/51). Although the 1G4_α58β61 TCR also made many contacts with the large exposed side chain of W5 (54/115), binding was less focussed on this residue and additional important contacts were made with other peptide residues, particularly M4 (31/115) and Q8 (15/115). Peptide binding hotspots have been detected for natural TCRs^{23,130,194}, however, this structural feature has been associated with auto-reactive TCRs and can correlate with a high level of TCR cross-reactivity.^{16,23} In contrast, the Hyb3.3 and ESK-1 mimic antibodies both bound to their respective pHLA using unconventional topology. Hyb3.3 binding was C-terminally shifted with an engagement angle substantially outside of the natural TCR-pHLA range. Despite this unusual topology, Hyb3.3 retained broad peptide contacts across 6 of the 9 amino acids, making 25% peptide contacts and a BSA of 2024 Å² (**Figure 17C**). ESK-1 binding was N-terminally shifted with an unconventional engagement angle. This binding mode positioned the TCR-mimic antibody so that the CDR3 loops were focussed over the α2 helix of the HLA, resulting in a very limited interaction with the peptide. This binding mode resulted in ESK-1 making only 10% peptide contacts and the majority of these were formed with peptide residue 1. This HLA-centric binding mode, where only one peptide residue was contacted, raised questions about the ability of the ESK-1 TCR-mimic to retain specificity. However, for all the other HLA-targeting reagents considered here, their structures appear to retain many of the features observed for natural TCR-pHLA interactions.

Alanine scan analysis reveals distinct molecular recognition patterns

Alanine scan mutagenesis was used to investigate the molecular recognition pattern of the affinity-enhanced TCRs and TCR-mimic antibodies using surface plasmon resonance (**Figure 18; Suppl Figure S1**). For each reagent, we generated a panel of soluble pHLAs in which each peptide residue was replaced with an alanine (or a serine if the native residue was already an alanine), except for the canonical anchor residues at positions 2 and 9. The 1G4_α58β61 TCR bound to A2-SLL with a K_D of 57 pM. However, binding was not detected when residues 4, 5 and 6 were

mutated to alanine, and the affinity was substantially reduced when residues 7 and 8 were mutated (**Figure 18A**). These findings are consistent with the 1G4_α58β61-A2-SLL co-complex crystal structure showing that the central MW motif forms a central peptide bulge making multiple contacts with the TCR CDR3 loops, and peptide residue Q8 points up away from the HLA surface enabling contacts with the TCR CDR1β loop (**Figure 17**). A similar pattern was observed for the 1G4_α5β51 A2-SLL restricted TCR (K_D of 1.4nM), with small reductions in affinity also observed at peptide positions 1 and 3, whereas the 1G4_α5β100 A2-SLL restricted TCR, that bound with a weaker affinity (K_D of 5 nM) was highly sensitive to alanine mutations at every position along the peptide backbone (**Figure 18A**). We repeated the alanine scan analysis on the A2-SLL-reactive 3M4E5 TCR-mimic and included two published higher affinity versions of 3M4E5 (3M4E5_T2 and 3M4E5_T3) because they were closer in affinity to the 1G4_α5β100 and 1G4_α5β51 affinity-enhanced TCRs, allowing a more direct comparison. The 3M4E5 ($K_D = 44$ nM in scFv format) and 3M4E5_T2 TCR-mimic antibodies ($K_D = 2.8$ nM in scFv format) were both sensitive to alanine mutation at peptide residues 4, 5 and 6 (**Figure 18B**), whereas mutations at all other positions of the peptide did not reduce binding affinity. 3M4E5_T3 ($K_D = 5.5$ nM in scFv format) demonstrated a similar trend, being sensitive to alanine substitution at peptide residues 4 and 5 (**Figure 18B**). Alanine substitutions at peptide residues 1, 3, 7, and 8 had no impact on binding affinity for any of the A2-SLL TCR-mimics, demonstrating a more focussed binding mode around peptide residues 4, 5, and 6, compared to the affinity-enhanced TCRs. These findings were also consistent with the crystal structure of 3M4E5-A2-SLL that demonstrated binding was focussed towards these central residues of the peptide.

The high level of sensitivity to alanine substitutions across the peptide backbone was also observed for the A1-EVD specific MAG-IC3 ($K_D = 3.8$ nM) and MAG-IC5 ($K_D = 17$ nM) TCRs (**Figure 17C**). The stronger affinity MAG-IC3 TCR demonstrated reduced or abrogated affinity towards every alanine mutant tested, whilst the MAG-IC5 TCR was sensitive to mutations at all positions apart from peptide residues 6 and 7. The MAG-IC3-A1-EVD co-complex crystal structure was consistent with this finding, demonstrating a complex network of contacts across the peptide backbone (**Figure 17**). The Hyb3.3 TCR-mimic antibody recognises the same peptide region as MAG-IC3 and MAG-IC5 but derived from a different MAGE protein (MAGE-A1) and binds with an affinity of $K_D = 18$ nM. The MAGE-A3 and MAGE-A1 peptides are conserved at all positions except for the N-terminal anchor (position 2), position 5 and position 8. Although the Hyb3.3-A1-EAD complex structure demonstrated non-canonical topology, the interactions with the peptide apparent in the structure were comparable to most TCR-pHLA complexes (**Figure 17**).

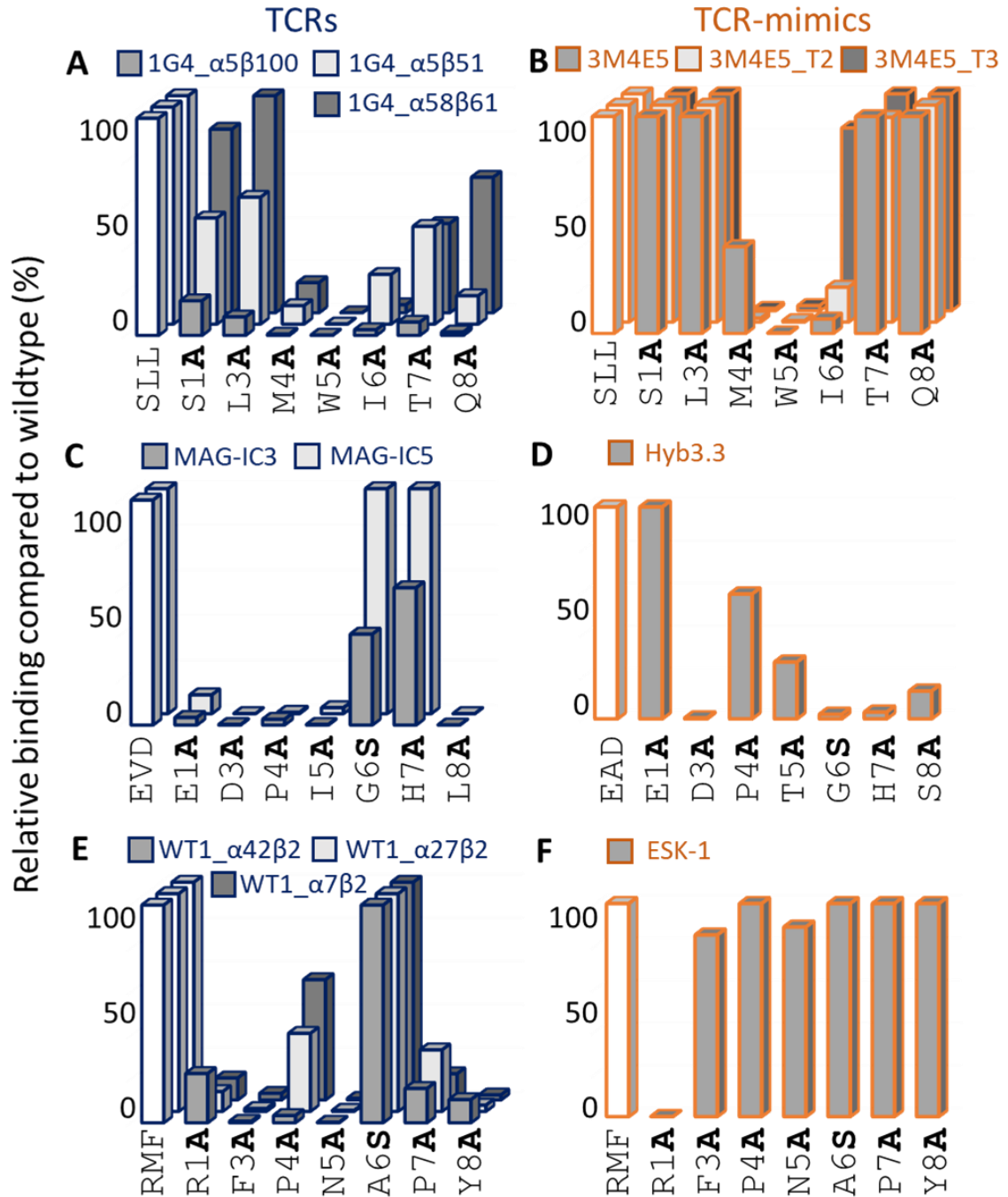


Figure 18: Alanine scan analysis reveals distinct molecular recognition patterns. The contribution of peptide side chains to binding specificity was analysed using alanine scan mutagenesis (by surface plasmon resonance). Binding affinities of the TCRs and TCR-mimic antibodies were determined using single cycle kinetic analysis. Bar graphs show binding affinity as a % relative to the binding affinity to the index peptide. (A) A2-SLL affinity-enhanced TCRs, (B) A2-SLL TCR-mimics, (C) A1-EVD affinity-enhanced TCRs, (D) Hyb3, (E) A2-RMF affinity-enhanced TCRs, and (F) ESK-1.

This observation was reflected in the alanine scan analysis that demonstrated some degree of sensitivity to alanine substitutions at all positions tested apart from peptide residue 1 (**Figure 18D**). The WT1_α7β2 TCR, which bound A2-RMF with a K_D of 70 nM, exhibited abrogated or highly reduced binding for all residues except at peptide position 6. The WT1_α27β2 ($K_D = 13$ nM) and WT1_α42β2 ($K_D = 0.76$ nM) TCRs showed a similar trend, with the strongest binding WT1_α42β2 TCR exhibiting the greatest level of sensitivity to alanine substitutions across the peptide backbone (**Figure 18E**). The ESK1 TCR-mimic antibody, which has a relatively weak affinity for A2-RMF of $K_D = 2$ μM in scFv format (**Supp Figure S1**), demonstrated broad degeneracy in peptide binding, being tolerant to alanine substitutions at all positions of the peptide except for peptide residue 1 (**Figure 18F**). Here, availability of the ESK-1-A2-WT-1 structure provided insight into this observation, confirming that virtually all contacts between ESK-1 and the WT-1 peptide were focussed on peptide residue 1 (**Figure 17**).

TCR-mimic antibodies bind to several commonly expressed self-peptides

Although alanine scan analysis is useful to understand positional sensitivity of pHLA-targeting receptors, it is unclear how these data relate to the broader cross-reactivity of these reagents, particularly their ability to discriminate against common self-peptides. To gain further insight into self-discrimination, we designed a new experimental approach for screening multiple pHLA complexes in a high throughput format by modifying the MAGPIX platform. We designed multiplex experiments using MagPlex beads coated with HLAs in complex with a range of commonly expressed self-peptides. Self-peptide-HLAs recognised by affinity-enhanced TCRs or TCR-mimics were detected using MAGPIX analysis (**Table 3+Table 4**). In all cases, the affinity-enhanced TCR reagents only generated a signal against their respective index peptides, whereas the TCR-mimic antibodies (in scFv format to avoid avidity-mediated binding) were more promiscuous. The 3M4E5_T2 scFv was reactive against four broadly expressed HLA-A*02:01 restricted self-peptides and the ESK1 scFv demonstrated reactivity against six broadly expressed HLA-A*02:01 restricted self-peptides, in addition to their target antigens A2-SLL and A2-RMF, respectively (**Table 3**). The Hyb3.3 scFv was reactive against almost all the HLA-A*01:01 restricted self-peptides tested (9/12), as well as A1-EAD (**Table 4**). In all cases, the self-peptides recognised by the TCR-mimic antibodies shared very little sequence similarity revealing a high level of potential cross-reactivity compared to the affinity-enhanced TCRs developed to target identical, or very similar, peptide antigens.

Table 3: MAGPIX analysis to investigate a panel of commonly expressed HLA-A*0201-restricted self-peptides. Peptides generating a signal above background (3 times median intensity of all bead regions bound to native helper-phage) were classified as positive binders, and binding expressed as a percentage of signal obtained from binding to index peptide in each case. Percentage binding for each interaction is reported as the average of several experimental repeats.

HLA Antigen	Peptide	1G4_α5β100	1G4_α58β61	3M4E5_T2	WT1_α7β2	ESK-1
		TCR	TCR	scFv	TCR	scFv
A2	SLLMWITQC	100.0	100.0	100.0	0.0	0.0
A2	NY-ESO(157-165)					
A2	RMFPNAPYL	0.0	0.0	0.6	100.0	100.0
A2	WT1(126-134)					
A2	AIVDKVPSV	0.0	0.0	0.0	0.0	0.0
A2	COPG1(147-155)					
A2	ALVVQVAEA	0.0	0.0	0.0	0.0	0.1
A2	HEXB(34-42)					
A2	SLDQPTQTV	0.0	0.0	0.0	0.0	0.0
A2	EIF3C(834-842)					
A2	GLATDVQTV	0.0	0.0	0.0	0.0	0.0
A2	PSMB3(55-63)					
A2	ILTDITKGV	0.0	0.0	0.0	0.0	0.0
A2	EEF2(661-669)					
A2	SNRPGP15(68-76)	0.0	0.0	0.0	0.0	0.0
A2	IMLEALERV					
A2	VMDSKIVQV	0.0	0.0	0.1	0.0	0.2
A2	KPNA1(434-442)					
A2	RLQEDPPAGV	0.0	0.0	0.0	0.0	0.3
A2	UBE2A(15-24)					
A2	KIYEGQVEV	0.0	0.0	0.0	0.0	0.2
A2	RPL5(117-125)					
A2	NLAENISRV	0.0	0.0	0.0	0.0	0.0
A2	PYGM(271-279)					
A2	LLDVPTAAV	0.0	0.0	0.0	0.0	0.0
A2	IFI30(16-24)					
A2	SLSEKTVLL	0.0	0.0	0.0	0.0	0.0
A2	CD59(106-114)					
A2	ALNEKLVNL	0.0	0.0	0.0	0.0	0.0
A2	EIF3F(349-357)					
A2	ILDKKVEKV	0.0	0.0	0.1	0.0	0.3
A2	HSP90AB1(570-578)					
A2	ILDQKINEV	0.0	0.0	0.0	0.0	0.0
A2	ODC1(23-31)					
A2	VLIDYQRNV	0.0	0.0	0.0	0.0	0.0
A2	XPO1(784-792)					
A2	GLIEKNIEL	0.0	0.0	0.0	0.0	0.0
A2	DNMT1(425-433)					
A2	FLDPNNIPKA	0.0	0.0	0.0	0.0	0.0
A2	ALG8(305-314)					
A2	SLQSTILGV	0.0	0.0	0.2	0.0	0.0
A2	LONP2(51-59)					

Table 4: MAGPIX analysis to investigate a panel of commonly expressed HLA-A*0101-restricted self-peptides. Peptides generating a signal above background (3 times median intensity of all bead regions bound to native helper-phage) were classified as positive binders, and binding expressed as a percentage of signal obtained from binding to index peptide in each case. Percentage binding for each interaction is reported as the average of several experimental repeats.

HLA Antigen	Peptide	MAG-IC3 TCR	Hyb3.3 scFv
A1	EVDPIGHLY MAGE-A3(168-176)	100.0	70.0
A1	EADPTGSHY MAGE-A1(161-169)	0.0	100.0
A1	YSDKYGLGY PLK1(417-425)	0.0	90.7
A1	DTDHYFLRY PIGT(165-173)	0.0	16.8
A1	STDHIPILY GFPT1(218-226)	0.0	0.4
A1	HSDPSILGY GIGYF1(1012-1020)	0.0	15.2
A1	KSDVHLNFY HLTF(499-507)	0.0	0.0
A1	HTDILKEY DTWD1(262-270)	0.0	0.0
A1	IADMGHLKY PCNA(241-249)	0.0	1.0
A1	LTELPDWSY MRPL52(43-51)	0.0	0.0
A1	ASDPFFRHY GPN2(210-218)	0.0	77.3
A1	ETEKDFSRY AQR(1757-895)	0.0	4.3
A1	GTVYEDLRY UBE2C(71-79)	0.0	0.1

Deep sequencing of peptides from randomised pHLA phage-libraries demonstrates the binding degeneracy of pHLA-targeting reagents

Despite the ability of 3M4E5 to bind pHLA in a TCR-like conformation, the alanine scan profile and MAGPIX analysis revealed lower levels of peptide-selectivity compared to the affinity-enhanced TCRs. To probe this discrepancy further, we developed a novel approach for the characterisation of TCR-peptide degeneracy using randomised pHLA libraries displayed on phage (Coles C *et al.* Submitted). We used this system to identify peptide motif preferences for the affinity-enhanced TCRs and TCR-mimics that recognised A2-SLL. In broad agreement with the structural and alanine scan analyses, the affinity-enhanced TCRs, 1G4_α5β100, 1G4_α5β51 and 1G4_α58β61 showed strong preferences for the native SLL peptide sequence at residues W5, T7, Q8 and V9 (**Figure 19A-C**). In contrast, the TCR mimic 3M4E5 demonstrated a preference for the native SLL peptide sequence only at W5, with all other positions showing very little amino acid preference (**Figure 19D-F**). Although W5 was selected by 3M4E5_T2 and 3M4E5_T3, it was not the dominant amino acid preference at this position, with phenylalanine (F) being preferred by both reagents. 3M4E5_T2 and 3M4E5_T3 also displayed very little selectivity in

terms of amino acid preference at any other position. These findings are consistent with the alanine scan results demonstrating that the A2-SLL reactive TCR-mimics could tolerate alanine substitutions at any residue outside of the central MW-peg in the SLL peptide, while the affinity-enhanced TCRs were selective across the peptide backbone.

Using this information, we assessed the number of unique peptides selected by each reagent to gain insight into their comparative cross-reactivity. Analysis of next generation sequencing data identified an average of 7,068 unique peptides (687,241 total reads) for 1G4_α5β51, 4,455 unique (689,928 total reads) for 1G4_α5β100, and 9,012 unique (696,992 total reads) for 1G4_α58β61. The TCR-mimics selected between three and fifteen times more unique peptides compared to the affinity-enhanced TCRs with 50,765 unique peptides (740,196 total reads) for 3M4E5, 60,699 unique (692,455 total reads) for 3M4E5_T2 and 32,934 unique (727,824 total reads) for 3M4E5_T3. Overall, these data suggest that the A2-SLL affinity-enhanced TCRs are less cross-reactive (in terms of total number of peptides recognised) and less promiscuous in terms of their ability to tolerate amino acid variation across the peptide backbone, compared to the A2-SLL TCR-mimics.

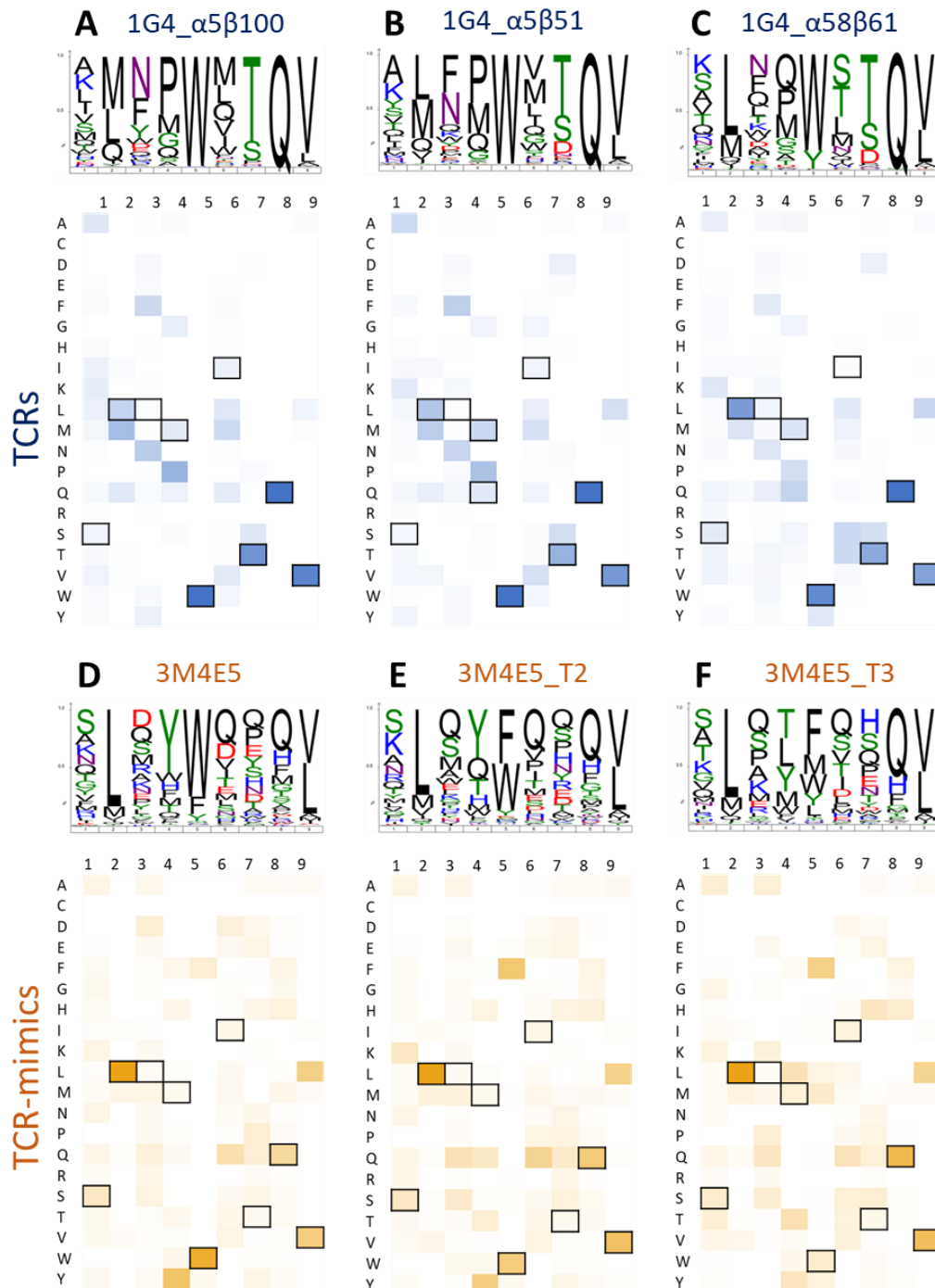


Figure 19: Deep sequencing of peptides from randomised pHLA phage-libraries demonstrates the binding degeneracy of pHLA-targeting reagents. Sequence logos (Icelogo software) and heat maps were generated from NGS sequencing of pan 3 data identifying the distribution of amino acid identities per position of the peptide selected by A2-SLL-reactive affinity-enhanced TCRs and TCR mimic antibodies. The abundance of an amino acid is shown by intensity of colour. Outlined boxes identify the amino acids of the cognate antigen SLL. Data shown is the average of two experimental repeats. (A) 1G4_α5β100 (B) 1G4_α5β51, (C) 1G4_α58β61, (D) 3M4E5, (E) 3M4E5_T2, and (F) 3M4E5_T3.

Molecular dynamics simulations reveal peptide-selectivity is associated with distinctive energetic modes of binding

Although the structural and alanine scan analyses provided useful insights into the recognition mode employed by the pHLA-targeting molecules described here, they were not fully predictive of the recognition patterns observed in the MAGPIX and randomised pHLA library analysis. For example, despite the A2-SLL and A1-EAD reactive TCR-mimics forming seemingly broad peptide contacts, according to the structural analysis, and promising alanine scan profiles, these reagents bound substantially more self-peptides in the MAGPIX analysis than the affinity-enhanced TCRs. Furthermore, the A2-SLL reactive TCR-mimics were characterised by more degenerate peptide binding in the randomised pHLA library analysis. Consequently, we performed molecular dynamics (MD) simulations to extend the ‘snapshot’ view available from crystal structures. We subjected all six structures described in **Figure 17** to two, 500 nanosecond long MD simulations to investigate the biochemical nature and lifetime of contacts formed between the peptide and affinity-enhanced TCRs or TCR-mimics. Interactions were dissected into contacts formed between the peptide side chain (amino acid specific) and main chain (conformation specific) *versus* time and separated into hydrogen bonding and Van der Waals (VdW) type interactions. In all cases, the affinity-enhanced TCRs made a higher number of long-lived contacts with side chain atoms across the peptide compared to main chain interactions (**Figure 20A-C**). In contrast, 3M4E5 made a lower number of peptide side chain contacts (**Figure 20D**), reflected by a lower overall ratio of peptide side chain contacts (Hbond ratio: 0.5, vdW ratio: 2.23) compared to the 1G4_α58β61 TCR (Hbond ratio: 1.49, vdW ratio: 5.41) (**Supp Figure S2**). Hyb3.3 made virtually no contacts with peptide side chain residues (Hbond ratio: 0.16, vdW ratio: 0.24), focussing primarily or exclusively on interactions with the peptide backbone (**Figure 20E, Supp Figure S2**). ESK-1 did make peptide side-chain interactions, but only with the exposed R1, in line with the alanine scan analysis (**Figure 20F, Supp Figure S2**). These data suggest that more limited side chain-mediated recognition pattern, as observed for the TCR-mimics, might contribute to greater levels of peptide degeneracy.

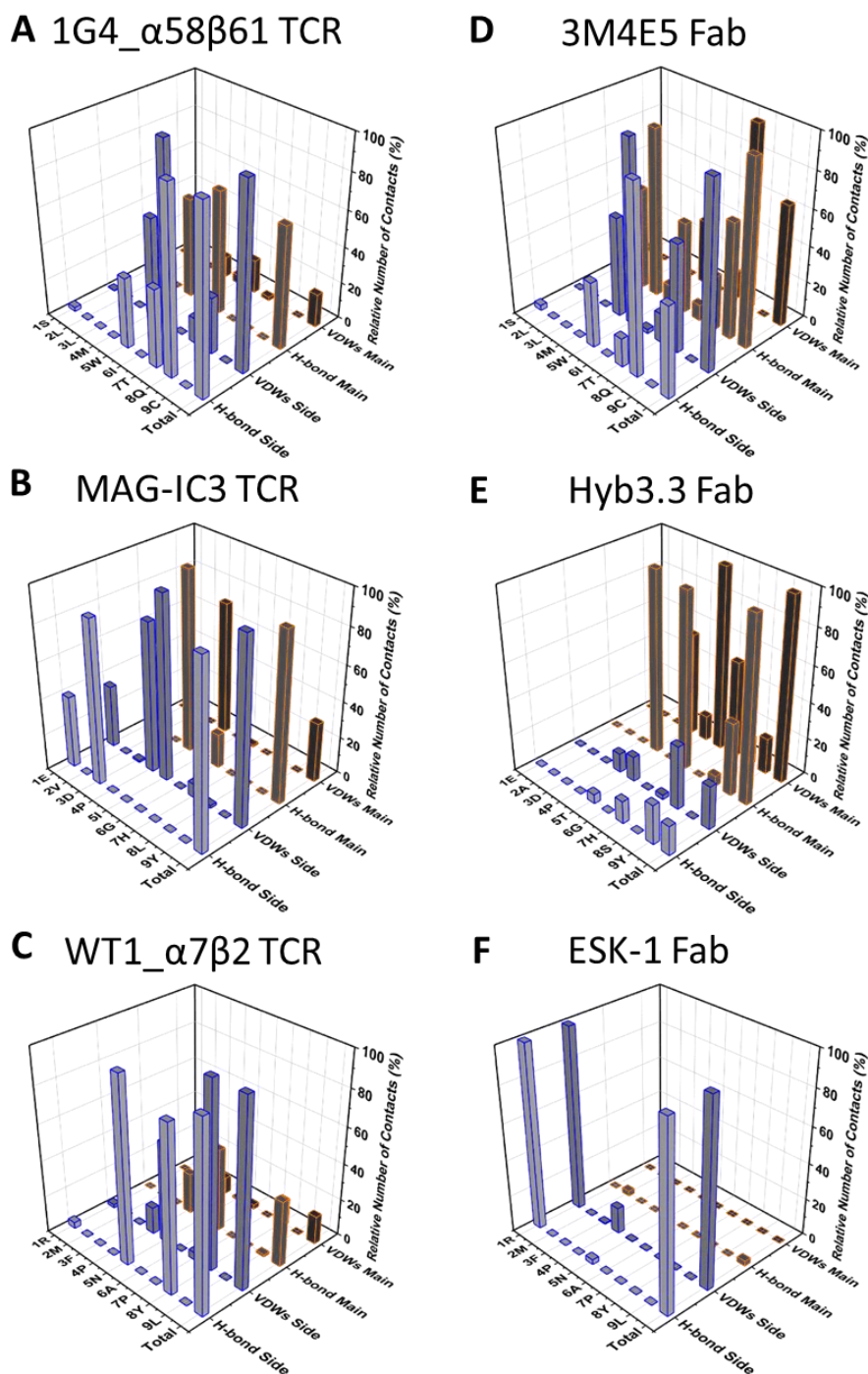


Figure 20: Molecular dynamics simulations reveal broad side chain contacts with the peptide drive specificity. Relative number of Hydrogen bonds (H-bonds) and Van der Waals (vdWs) interactions formed between either the main or side chain of each peptide residue to the TCR/TCR-mimic over the course of our MD simulations. Total side versus main chain ratios for H-bonds and vdWs interactions are shown, with the larger value (side or main for each category) scaled to 100 % (absolute values for all contacts are provided in **Supp Figure S2**). (A) 1G4_α58β61-A2-SLL, (B) MAG-IC3-A1-EVD, (C) WT1_α7β2-A2-RMF, (D) 3M4E5-A2-SLL, (E) Hyb3.3-A1-EAD, (F) ESK-1-A2-RMF.

The energetic landscape of each affinity-enhanced TCR/TCR-mimic-pHLA complex was characterised by calculating their binding free energies (using the Molecular Mechanics Poisson–Boltzmann Surface Area (MMPBSA) approach).³⁹ The MMPBSA approach has been used extensively to predict relative protein-ligand and protein-protein binding free energies.^{85,195} More specifically, it has been used to rationalise the effect of mutations on antibody-antigen complexes¹⁹⁶, the role of a water bridged interaction in TCR-pHLA affinity¹⁹⁷, and to predict reliable relative binding free energies of pHLA complexes.⁹² Here, we performed 25 x 4 nanosecond MD simulations per complex for MMPBSA analysis, to ensure reliable and converged results.^{85,91,92} MMPBSA calculations have the advantage that they can be decomposed into per-residue contributions to the binding free energy, allowing one to predict each residue's preference towards binding. Analysis of the decomposition results demonstrated that the affinity-enhanced TCRs were characterised by broad energetic signatures, whereby the binding energy was distributed over at least three peptide residues, with multiple disparate regions driving affinity across the HLA surface (**Figure 21A-C; Supp Figure S3A-C**). On the other hand, the TCR-mimics were more HLA-focussed with one, or two, energetic hotspots focussed on single HLA residues (**Figure 21D-F; Supp Figure S3D-F**). For example, although the 1G4_α58β61 TCR made substantial energetic interactions with HLA residue R65 (-12 kcal mol⁻¹), this was balanced by interactions with multiple peptide residues (W5: -11 kcal mol⁻¹, M4: -8 kcal mol⁻¹, Q8: -7 kcal mol⁻¹) and other HLA residues (Q155: -6 kcal mol⁻¹). No energetic hotspots were detected for the WT1_α7β2 TCR, with most of the binding energy being equally balanced over HLA residues R65, R75 and Q155 and peptide residues P4, N5 and Y8 (all around -6 kcal mol⁻¹) and the MAG-IC3 TCR had a balanced energetic footprint with important binding contributions from HLA residues E63, N66 and V158, as well as peptide residues E1, D3, P4 and H7 (all around -5 kcal mol⁻¹). In contrast, all of the TCR-mimics utilised more focussed energetic binding to engage their cognate pHLAs. The chief energetic contribution for 3M4E5 was made by HLA residue R65 (-16 kcal mol⁻¹) with an additional peptide hotspot at W5 (-7 kcal mol⁻¹). For ESK-1, energetic hotspots were detected at HLA residue K66 (-15 kcal mol⁻¹) and peptide residue R1 (-13 kcal mol⁻¹), with very little contribution from any other peptide residues. Finally, Hyb3.3 binding was characterised by strong energetically favourable interactions only with HLA residues K146 (-11 kcal mol⁻¹) and R65 (-11 kcal mol⁻¹), with the EAD peptide playing a minor role.

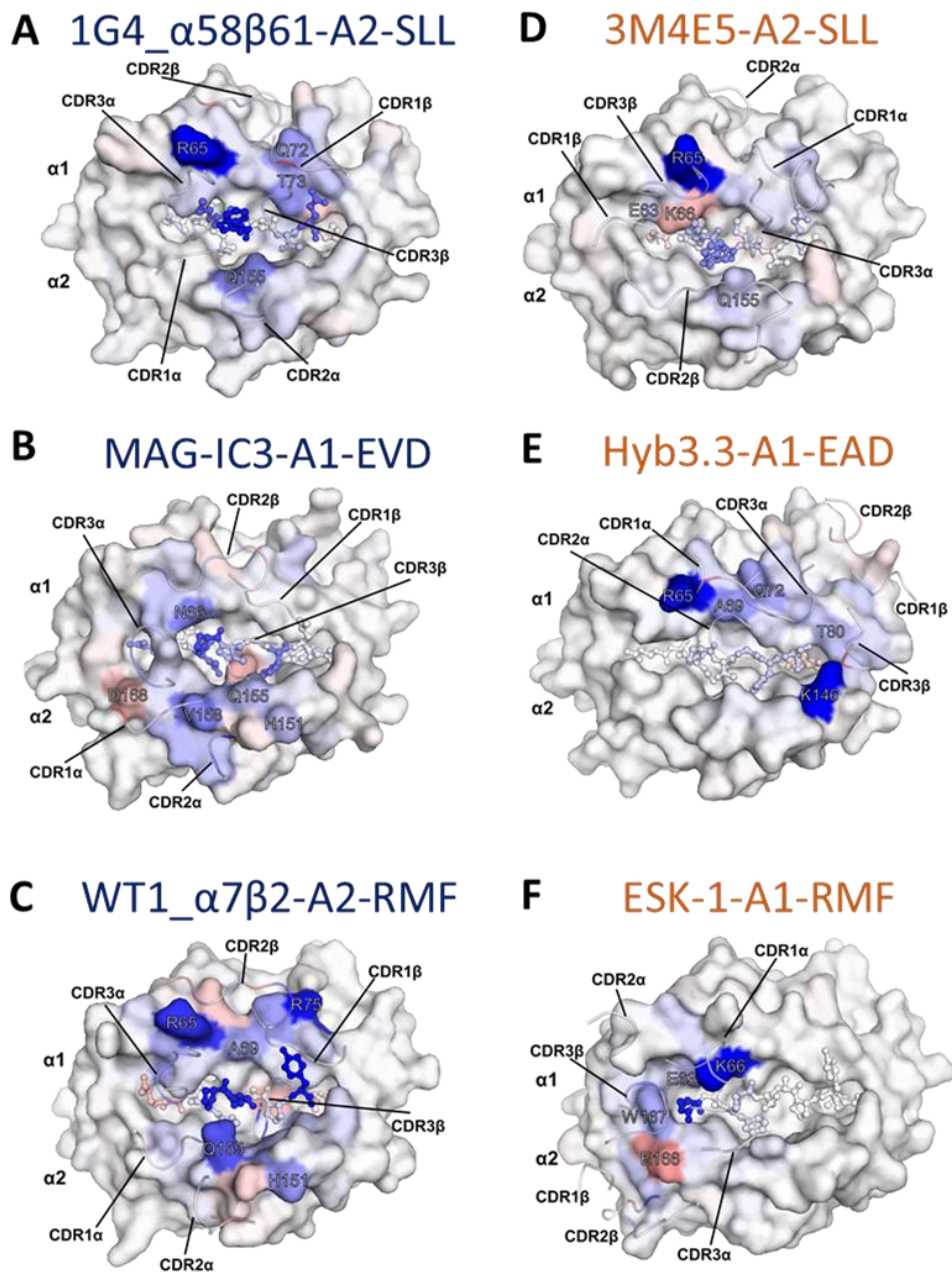


Figure 21: Binding free energy decomposition analysis of TCR and TCR-mimic-pHLA interactions. Per-residue decomposition of the binding free energy obtained from our MMPBSA calculations, to identify energetic hotspots for each TCR or TCR-mimic interaction with cognate pHLA. A top down view of each pHLA is shown, with the peptide depicted as sticks and the HLA as a surface. Colour mapping of the decomposition results for each residue was performed across the entire binding interface and used to indicate which residues across this interface favour (blue) or disfavour (red) binding (with white indicating no preference). (Bar graphs for all decomposition results are provided in **Supp Figure S3**). (A) 1G4_α58β61-A2-SLL, (B) MAG-IC3-A1-EVD, (C) WT1_α7β2-A2-RMF, (D) 3M4E5-A2-SLL, (E) Hyb3.3-A1-EAD, (F) ESK-1-A2-RMF.

Together with the previous analyses, the molecular dynamic simulations provide further evidence that, rather than being driven by the recognition of a dominant amino acid at a single position on the HLA or peptide, broad interactions across the peptide (particularly with peptide side chains) was associated with greater peptide-selectivity.

Re-directed T cell killing of antigen-positive and negative cell lines using pHLA-targeting bispecifics

On-target *versus* off-target reactivity of affinity-enhanced TCRs and TCR-mimics were assessed in functional T cell redirection assays against antigen-positive and antigen-negative cell lines. Soluble bispecific molecules were generated by fusing an anti-CD3 scFv to the β -chain of the affinity-enhanced 1G4_ α 58 β 61 (IMC-1G4_ α 58 β 61), and MAG-IC3 TCRs (IMC-MAG-IC3), or the heavy chain of the TCR-mimic scFv 3M4E5_T2 (3M4E5_T2-anti-CD3), 3M4E5_T3 (3M4E5_T3-anti-CD3), and Hyb3.3 (Hyb3.3-anti-CD3). Reagents recognising A2-RMF were not included because, consistent with evidence from other studies¹⁹⁸, we were unable to detect the peptide on WT1⁺ tumour cells by mass spectrometry analysis (data not shown). IMC-1G4_ α 58 β 61 redirected T cell killing of A2⁺/NYESO⁺ NCIH-1755 cells was approximately twenty-times more sensitive when compared to 3M4E5_T2-anti-CD3 and 3M4E5_T3-anti-CD3, in line with the difference in affinity between these reagents (**Figure 22A, Supp Figure S4**). No redirected T cell killing of A2⁺/NYESO⁻ targets was detected for IMC-1G4_ α 58 β 61, whereas 3M4E5_T2-anti-CD3 and 3M4E5_T3-anti-CD3 both induced redirected T cell killing of the A2⁺/NYESO⁻ antigen-negative cell lines HEP-G2, Ren8 and HISMC at EC₅₀s similar to those for the A2⁺/NY-ESO-1⁺ cell line NCIH-1755. Similarly, Hyb3.3-anti-CD3 demonstrated T cell redirected killing of multiple A1⁺/MAGE-A1⁻ antigen-negative cell lines, as well as A1⁺/MAGE-A1⁺ antigen positive cell lines, demonstrating no, or a very small specificity window. In contrast, IMC-MAG-IC3 mediated redirected T cell killing of A1⁺/MAGE-A3⁻ antigen negative cells was absent (HISMC cells) or only occurred at very high concentrations (NCI-H1703 and COLO205), demonstrating a clear specificity window compared to redirected T cell killing against HCC1428 A1⁺/MAGE-A3⁺ antigen positive cells (**Figure 22B, Supp Figure S5**).

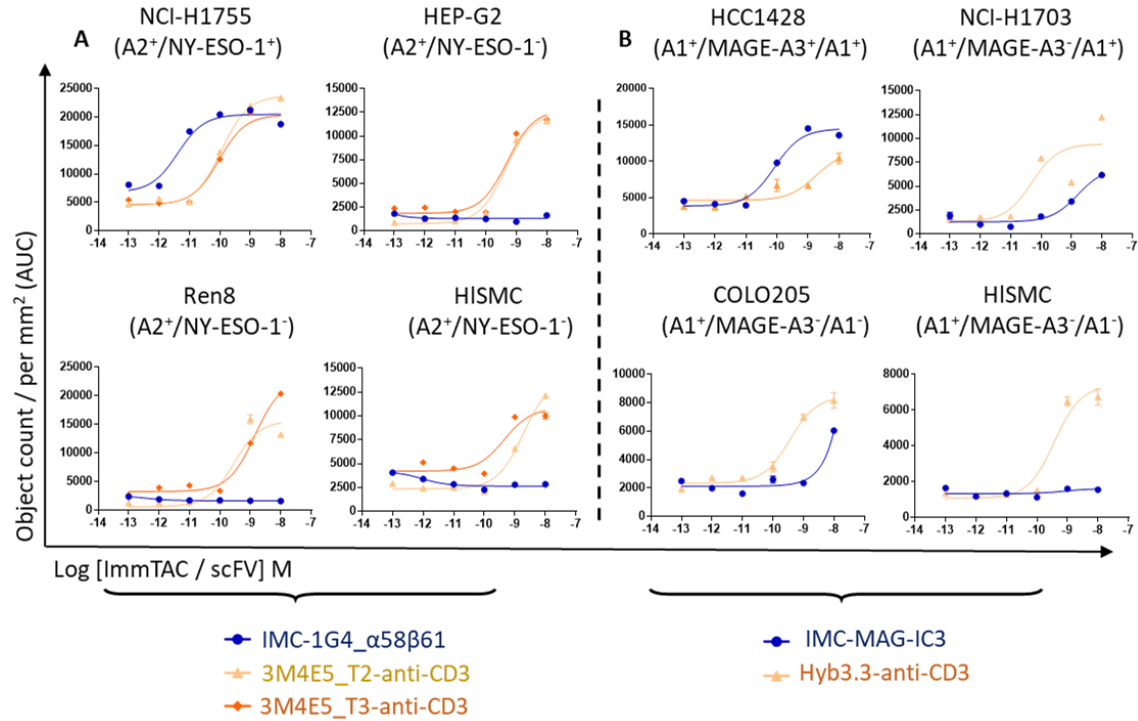


Figure 22: Redirected T cell killing of antigen positive and negative cell lines using pHLA-targeting bispecifics. The activity of the ImmTAC molecules and the TCR-mimic-anti-CD3 fusions was tested against a range of antigen positive and antigen negative cell lines (tumour and healthy cells) using IncuCyte killing assays. Data are plotted using area under the curve analysis. **(A)** IMC-1G4_α58β61, 3M4E5_T2-anti-CD3 and 3M4E5_T3-anti-CD3 T cell re-direction against HLA-A*02:01+/NY-ESO-1+ (NCI-H1755) and HLA-A*02:01+/NY-ESO-1- (HEP-G2, Ren8 and HISMIC) cells lines. **(B)** IMC-MAG-IC3 and Hyb3.3-anti-CD3 T cell re-direction against HLA-A*01:01+/MAGE-A3+ (HCC1428), HLA-A*01:01+/MAGEA1+ (HCC1428 and NCI-H1703) and HLA-A*01:01+/MAGE- (COLO205 and HISMIC) cells lines.

Overall, the HLA-targeting bispecifics based on the natural TCR scaffold retained much higher levels of specificity in cellular testing, consistent with the MAGPIX, and randomised pHLA library analyses. These findings support the hypothesis that dispersed peptide contacts with the comparatively broad peptide side-chain focussed energetic signature is more predictive of the ability to discriminate between different peptides.

The NYBR1 TCR exhibits no cellular off-target reactivity and utilises a broad, peptide side-chain centric binding mode

In order to test the notion that peptide-selectivity is associated with broad contacts with peptide side chains and a dispersed energetic profile, we extended our analysis to include an affinity-enhanced TCR that was known to be highly selective for its target pHLA. The NYBR1 TCR was affinity-matured against a novel cancer specific HLA-A*02:01 restricted peptide (SLSKILDTV; referred to as SLS peptide from hereon) derived from the NY-BR-1 lineage antigen and used to

generate an ImmTAC molecule (IMC-NYBR1). In cellular testing, as assessed by both IFN γ release and target cell killing, IMC-NYBR1 demonstrated exquisite specificity, as evidenced by absence of T cell redirection against 10 A2⁺/NY-BR-1⁻ antigen-negative cell lines, even at high concentrations (up to 2 nM of IMC-NYBR1 for a 47 pM affinity reagent) (**Figure 23A**). The structure of the NYBR1-A2-SLS complex (solved at 2.3 Å resolution; **Table 2**) demonstrated that NYBR1 bound canonically, with a normal crossing angle (63.6°) and a BSA slightly above the reported range (2835 Å²), engaging with 7 of the 9 peptide residues (32% peptide contacts) (**Figure 23B**). This binding mode enabled fine specificity across the peptide, evidenced by sensitivity to alanine substitutions at every peptide position apart from peptide position 1 (**Figure 23C**). Analysis of contacts and energetics from MD simulations demonstrated a highly peptide side-chain mediated interaction (**Figure 23D**) with important contributions from 6 of the 9 residues in the SLS peptide (**Figure 23E, Supp Figure S6**). Although there was a slight energetic focus towards HLA residue Q155 (-6 kcal mol⁻¹) and peptide residue K4 (-9 kcal mol⁻¹), a number of energetic contributions were made across the entire HLA surface and peptide (**Figure 23E**). Thus, consistent with our other observations in this study, the fine specificity of the NYBR1 TCR was associated with a broad energetic binding mode characterised by interactions with multiple peptide side-chains.

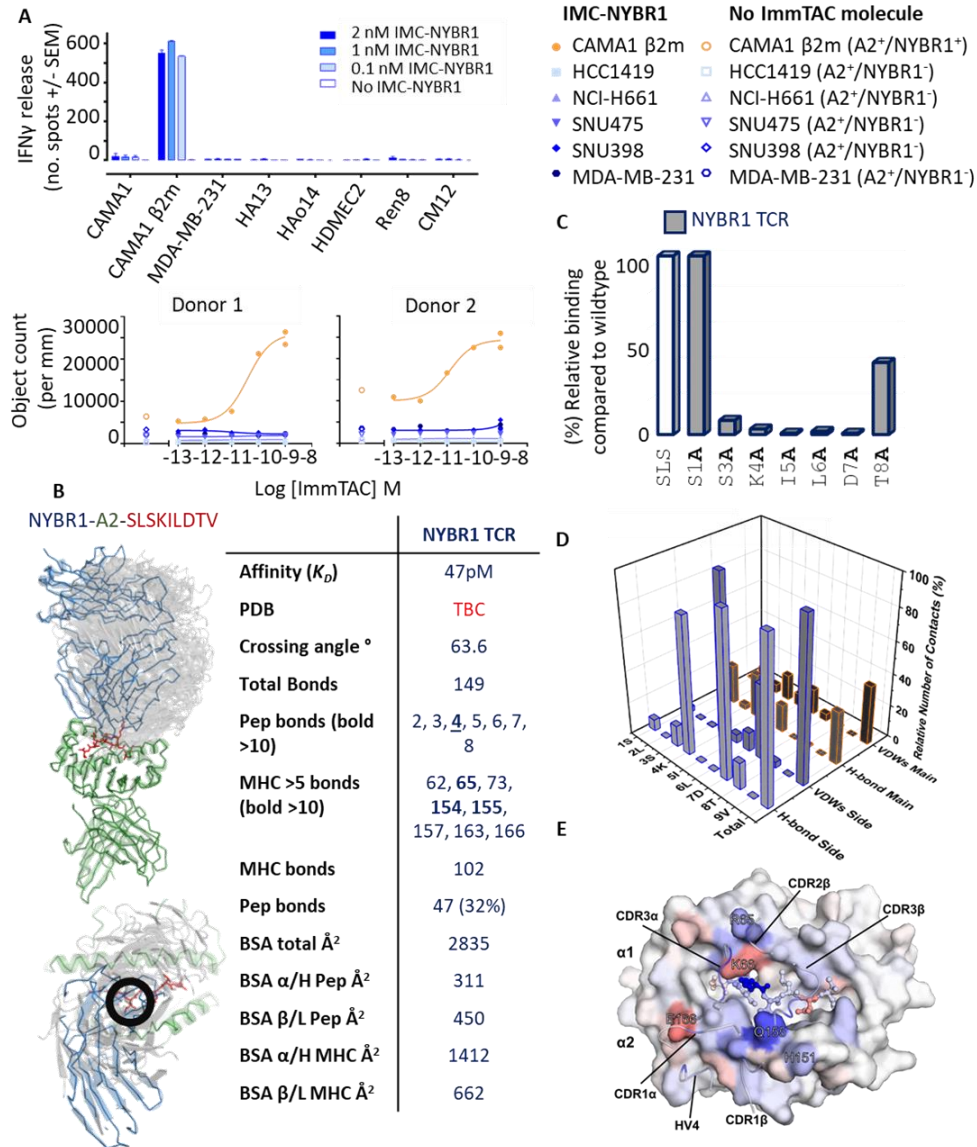


Figure 23: The NYBR1 TCR exhibits no cellular off-target reactivity and utilises a broad, peptide side-chain centric binding mode. **(A)** The activity of IMC-NYBR1 was tested against a range of HLA-A*02:01+/NYBR1+ (CAMA1 and CAMA1 β 2m) and HLA-A*02:01+/NYBR1- cell lines (MDA-MB-231, HA13, HAo14, HDMEC2, Ren8, CM12 HCC1419, NCI-H661, SNU475 and SNU398) using IFN γ ELISpot (bar graphs) and InCuCyte killing assays (area under the curve analysis) in two donors. **(B)** Above: Side on view of the structure of the NYBR1 TCR (blue ribbon) in complex with A2 (green ribbon) -SLS (red sticks). Below: Top down view of the NYBR1 TCR-A2-SLS interaction. Black circles represent the center of binding. The table shows a structural analysis of the NYBR1-A2-SLS complex. Bonds were assessed using the program contact (CCP4), implementing a 3.4 \AA cut-off for H-bonds and a 4 \AA cut-off for vdWs interactions. Buried surface area (BSA) \AA^2 was determined using ePISA. **(C)** The contribution of peptide side chains to binding specificity was analysed using alanine scan mutagenesis (by surface plasmon resonance). Binding affinities of the NYBR1-A2-SLS interaction was determined using single cycle kinetic analysis. Bar graphs show binding affinity as a % relative to the binding affinity to the index peptide. **(Figure Legend Continued on the next page).**

(Figure Legend Continued)

(D) Relative number of Hydrogen bonds (H-bonds) and vdWs interactions formed between either the main or side chain of each peptide residue to the NYBR1 TCR over the course of our MD simulations. Total side versus main chain ratios for H-bonds and vdWs interactions are shown, with the larger value (side or main for each category) scaled to 100 % (absolute values for all contacts are provided in **Supp Figure S2**). (E) Per-residue decomposition of the binding free energy obtained from our MMPBSA calculations. A top down view of the pHLA is shown, with the peptide depicted as sticks and the HLA as a surface. Colour mapping of the decomposition results for each residue was performed across the entire binding interface and used to indicate which residues across this interface favour (blue) or disfavour (red) binding, (with white indicating no preference) (bar graphs for all decomposition results are provided in **Supp Figure S3**).

Discussion

The identification of cancer-specific targets for solid tumours is challenging because cell surface antigens are often expressed on a wide range of tissues. Molecules such as CD19 and others have proved effective targets for liquid tumours because on-target toxicity is limited to mainly haematopoietic cells that can repopulate from the bone marrow following treatment.¹⁹⁹ Another key source of antigens is dysregulated or mutated intracellular proteins. However, many of these proteins are only presented on the cell surface in the context of pHLA. It is interesting to note that, even in response to very common human pathogens (i.e. influenza), natural antibodies recognising pHLA have not been detected, suggesting that humoral responses to pHLA are either ineffective, or dangerous to the host. Unlike antibodies, TCRs are selected to recognise pHLA in the thymus, which deletes T cells with TCRs that bind strongly to self-pHLA (and presumably removes TCRs that bind in a peptide-independent manner). This process is controlled in part by HLA co-engagement by the T cell co-receptors, CD8 and CD4, removal of which in CD8/CD4 co-receptor knockout mice has been shown to select T cells with TCRs that can bind to non-major histocompatibility complex proteins, such as CD155.^{111,200} A number of recent reports have also suggested that, probably as a consequence of thymic selection, TCRs have co-evolved to recognise pHLA.^{201–203} This ‘co-evolution’ model is consistent with findings from a recent study revealing that TCRs have a unique variable domain orientation compared to antibodies, the absence of which might restrict antibodies from reproducing the natural ability of TCRs to discriminate between different HLA-presented peptides.²¹ Thus, TCRs have multiple unique selection mechanism and structural features that guide pHLA recognition. We have previously demonstrated that affinity-enhanced TCRs maintain many of the binding characteristics of their

thymically-selected wild-type progenitors, with stronger binding generally being driven through the formation of new interactions with both the peptide and HLA surface.^{43,45,129,204–206} This feature likely provides an advantage for generating affinity-enhanced TCRs with the ability to retain a similar specificity profile compared to natural, thymically-selected TCRs.

The literature indicates that T cell potency is tuned by a TCR affinity threshold that is optimal in the low micromolar to high nanomolar range.^{116,180,207–211} For soluble bispecific reagents, this affinity threshold is controlled at the effector end of the molecule, while the affinity of the pHLA targeting end must be optimised according to antigen presentation. For most tumour associated pHLAs, their natural presentation levels can be very low (often below 10 copies of each specific peptide epitope per cell).^{19,182} Consequently, binding in the femto- to picomolar affinity range is needed to achieve a therapeutically relevant receptor occupancy level; a feat that has been achieved for monovalent TCR-based bispecifics.^{33,175} The literature, together with the data presented here, suggests that, with current technologies, engineering a TCR-mimic antibody to achieve this affinity range whilst maintaining peptide-selectivity is more challenging.¹⁷⁷ Some soluble TCR-mimics have been designed as full antibodies, thus achieving much stronger binding through avidity effects.¹⁷⁹ However, this essentially halves the number of effector molecules per target cell; a major issue if antigen expression is already limiting.

Here, we interrogated the molecular basis of pHLA-recognition of a panel of T cell redirecting bispecifics using a combination of structural, biochemical and computational approaches. All the affinity-enhanced TCRs utilised a canonical native TCR-like binding mode, maintaining broad contacts across the peptide backbone. This finding likely represents the advantage of using reagents that have been developed from a thymically-selected TCR progenitor. Our data also reveal that a native TCR-like binding mode is necessary, but not sufficient for enabling peptide-selectivity. For instance, despite the 3M4E5 TCR-mimic binding in an almost identical fashion to a native TCR, with peptide contacts across the peptide backbone, it still performed poorly in cellular cross-reactivity assays. Our MD simulation analysis demonstrated that 3M4E5 bound *via* two main hotspots: one on peptide residue W5 and one on HLA residue R65. These findings were mirrored in the randomised pHLA library analysis that demonstrated the A2-SLL-reactive TCR-mimics were all more degenerate (in terms of amino acid preferences across the peptide backbone) as compared to the A2-SLL-reactive affinity-enhanced TCRs and were predicted to select a far larger number of unique peptide sequences. Interestingly, this hotspot mediated binding mode has also been observed for some natural TCRs, but has been shown to correlate with high levels of

TCR cross-reactivity and has been implicated in autoimmunity.^{23,212} The ESK-1 and Hyb3.3 TCR-mimics also employed hotspot-binding modes, especially focussed towards residues on the HLA surface, and demonstrated loss of peptide-selectivity in biochemical (ESK-1 and Hyb3.3) and cellular (Hyb3.3) testing. In contrast, the TCR-based reagents tested all exhibited superior peptide-selectivity in biochemical and cellular tests and were characterised by binding modes that included a greater combination of balanced energetic interactions across the peptide and HLA surface. This was exemplified by the NYBR1 TCR (by far the cleanest molecule tested in both the molecular and cellular analysis), which also demonstrated a broad energetic binding mode with the majority of peptide contacts through side-chain interactions. Although the TCR-mimics selected in this study were generated using a different approach to the affinity-enhanced TCRs, our data suggest that the reagents based on the natural TCR-scaffold were better able to discriminate between different peptides by utilising peptide-specific binding interactions.

Even though we focussed on three published TCR-mimic antibodies because of available structures³⁴⁻³⁶, our data also have implications for other published studies of TCR-mimics. For instance, the TCR-mimic ESK-1 has been shown to target the Wilms tumour antigen in mouse models^{178,179}, despite evidence demonstrating that the A2-RMF antigen is not expressed.¹⁹⁸ This finding, combined with the data shown here, suggest that the activities reported for ESK-1 may be mediated by recognition of another peptide, or in a peptide-independent fashion. Although it is clear that engineering pHLA selectivity is still one of the major challenges for TCR-mimics²¹³, there are emerging reagents that exhibit more promising specificity profiles, including reagents targeting an HLA-A*02:01 restricted EBV LMP2A₄₂₆₋₄₃₄ (CLGGLTMV) epitope²¹⁴, and an alpha-fetoprotein₁₅₈₋₁₆₆ (FMNKFIYEI) epitope²¹⁵, the second of which has entered clinical trials as a CAR for the treatment of liver cancer. Our data demonstrate the importance of robust specificity testing of pHLA-targeting molecules, in line with our own preclinical safety testing package²¹⁶, that should be considered for the development of soluble pHLA-targeting bispecifics.

In summary, we demonstrate that by combining structural and biochemical data with atomistic MD simulations, the interactions underpinning pHLA recognition can be dissected in detail and can be used to better understand the specificity of pHLA-targeting reagents. Overall, these findings extend our understanding of the molecular rules that determine selective recognition of pHLA and shed additional light on how TCRs engage pHLA in a peptide-dependent fashion.

Finally, these observations also highlight the challenges associated with engineering pHLA-targeting molecules that can truly mimic TCR-like specificity.

Methods

TCR engineering

To obtain TCRs with affinity-enhanced for HLA-A*02:01 SLLMWITQC (NY-ESO-1₁₅₇₋₁₆₅), HLA-A*01:01 EVDPIGHL Y (MAGE-A3₁₆₁₋₁₃₄) and HLA-A*02:01 RMFPNAPYL (WT-1₁₂₆₋₁₃₄), the wild-type 1G4, MAGE-A3 and WT1 TCRs were subjected to phage display as previously described¹⁷⁵. A panel of high affinity TCRs were obtained with mutations in the α and/or the β chain (data not shown). TCRs from these panels were selected for this study according to their similarity in affinity to available TCR-mimic antibodies. Additionally, some stronger affinity-engineered TCRs were selected according to the availability of corresponding TCR-pHLA complex structures, to enable direct structural comparisons with the TCR-mimic antibodies.

Construct design, protein expression and purification

The modified TCRs, the TCR-mimics, β 2m and the HLA-A*01:01, and HLA-A*02:01 heavy chains were cloned into the pGMT7 vector and expressed in the BL21 (DE3) Rosetta pLysS *E.coli* strain as described previously.^{110,170} TCR constructs for biophysical analysis were designed to include the variable and constant domains of both chains (α and β) with an engineered inter-chain disulphide bond as previously described.¹⁷⁴ Antibody reagents for biophysical analysis were generated as single chain fusions (scFv) with a linker between the variable heavy and light chains (the constant domains were not included in the construct). Hyb3.3 scFv expression cassette was cloned into pCEP4, and protein expressed in mammalian cells using the ExpiCHO Expression System (ThermoFisher Scientific), as it did not express in *E. coli*. TCR constructs for T cell re-direction experiments were generated with an anti-CD3 scFv fused to the TCR β -chain as previously described (ImmTAC molecules). TCR-mimic antibody constructs for T cell re-direction experiments were designed as scFv with an anti-CD3 scFv fused to the heavy-chain. The HLA-A*01:01 and HLA-A*02:01 heavy chains were expressed with a biotinylation tag (for SPR experiments), or without it (for crystallisation screens) and refolded in the presence of β 2m and a specific peptide, as previously described.⁹⁸ TCRs and scFv were both refolded and purified using a previously described TCR refolding protocol.¹⁷⁴ For a 250 ml ImmTAC molecule refold, 6.5 mg α -chain was mixed with 16 mg β -chain. The refolds were extensively dialysed against 20 mM Tris

pH8 at 8°C and purified by Poros50HQ™ 10/100, Poros50HS™ 10/100 (Life Technologies) and Superdex S200HR™ 10/300 (GEH) columns.³³

Analysis of on-target and off-target T cell reactivity via re-direction using anti-pHLA/anti-CD3 bispecific reagents

The activity of the ImmTAC molecules (TCR-CD3 scFv fusions) and the TCR-mimic scFv-CD3 scFv fusions was tested through their ability to redirect T cells against a range of antigen-positive and antigen-negative cell lines (tumour and healthy cells). Incubation killing assays were performed according to the manufacturer's instructions (Sartorius, UK). Briefly, 100,000 PBMCs per well were added to 10,000 target cells per well. Target cells included HLA-A*02:01 and HLA-A*01:01 positive tumour cell lines, which were positive or negative for expression of target protein, as well as a panel of healthy cell lines. Target cells were incubated on plates overnight at 37°C in 5% CO₂ before addition of ImmTAC molecules, or TCR-mimic bispecific reagents at 10⁻⁷ – 10⁻¹² M followed by PBMCs and CellPlayer kinetic caspase apoptosis assay kit™. Cells were scanned every 3 hours for 70 hours, and data were quantified using IncuCyte ZOOM software™ (Sartorius, UK). Results were analysed using GraphPad Prism. For IFN γ ELISPOT assay, experiments were performed according to the manufacturer's instructions (BD BioSciences, UK). Briefly, 80,000 PBMCs per well were added to 50,000 target cells per well. Target cells included HLA-A*02:01 positive tumour cell lines, positive or negative for expression of target protein, as well as a panel of healthy cell lines. ImmTAC molecules, or TCR-mimic bi-specific reagents were added at 10⁻⁷ – 10⁻¹² M and incubated overnight at 37°C in 5% CO₂. Data were quantified after development using an automated ELISPOT reader (ImmunoSpot Series 5 analyser, Cellular Technology Ltd.).

MAGPIX peptide screening

Affinity-enhanced TCRs and TCR-mimic scFvs were subjected to peptide cross-reactivity analysis using a MagPLEX bead kit (Invitrogen, UK). Several common peptides expressed by healthy cells were refolded with either biotin-tagged HLA-A*02:01, or biotin-tagged HLA-A*01:01 (detailed in **Table 3**). Phagemid-encoded TCRs/TCR-mimics were expressed on the surface of bacteriophage M13, fused to capsid protein pIII²¹⁷, and binding to biotinylated self-peptide-HLA complexes attached to neutravidin-conjugated MagPLEX™ magnetic beads assayed. Positively bound beads were identified by MAGPIX analysis using a phage-specific PE-conjugated antibody. Peptides generating a signal above background (3 times median intensity of all bead regions bound to native helper-phage) were classified as positive binders, and binding expressed as a percentage of signal obtained from binding to index peptide in each case. Averages

of triplicate measurements for each interaction were determined within each experiment, and percentage binding for each interaction is reported as the average of several experimental repeats.

SPR Single cycle kinetic analysis

Purified TCRs and TCR-mimic scFv were subjected to SPR analysis using a BIAcore3000™. Briefly, pHLAs were biotinylated as described previously¹⁵⁰ and were immobilised onto a streptavidin-coupled CM5 sensor chip. For alanine scan analysis, 500 RUs of each alanine scan mutant were loaded onto individual flow cells. Flow cell one was loaded with free biotin to act as a control surface. All measurements were performed at 25°C in PBS buffer (Sigma, UK) at a flow rate of 30 µl/min. Binding profiles of the TCRs and TCR-mimic antibodies were determined using single cycle kinetic analysis as previously reported.^{205,206} TCRs and TCR-mimic antibodies were injected at a top concentration of around 20 µM, followed by four injections using serial 1/3 dilutions. K_D values were calculated assuming Langmuir binding ($AB = B \cdot AB_{max} / (KD + B)$) and data were analysed using the kinetic titration algorithm (BIAevaluation™ 3.1).²¹⁸

Generation of scHLA libraries and panning

scHLA libraries were generated as previously described (Coles *et al.* Submitted). Briefly, scHLAs were displayed on the surface of phage with the peptide component disulphide trapped in a single chain trimer (dsSCT). Diversity was encoded at the peptide level by introducing a flat distribution of 19 amino acids (excluding cysteine, to avoid cyclic peptide formation). All 19 amino acids were represented at the HLA primary anchors, Pos2 and Pos9, however, to maximise the functionality of the library with peptide correctly bound in the antigen binding groove, the amino acid distribution was biased towards to known anchor residue preferences for HLA-A*02:01. This phagemid library was introduced by electroporation into *E. coli* TG1 cells and grown in 2xYT amp 2% glucose media to OD600 = 0.5 and HelperPhage added at an infection ratio of ~ 20:1 phage to *E. coli*. Phage particles were isolated by PEG precipitation and 0.45 µM filtration. Panning was performed using 200nM for pan 1, and then decreasing concentrations (0.048 nM – 94nM) in subsequent pans to increase selection pressure. Biotinylated affinity-enhanced TCRs (1G4_α5β100, 1G4_α5β51 and 1G4_α58β61) and TCR-mimic antibodies (3M4E5, 3M4E5_T2 and 3M4E5_T3) were captured on streptavidin-coated paramagnetic beads and incubated with the library of purified phage particles preblocked in 3% MPBS buffer. Phage particles were eluted in trypsin and used to infect early log phase TGI *E. coli* cells and plated onto YTEag plates at 30°C for 16 h. Three rounds of selection were performed.

Deep sequencing of pHLA libraries

DNA was isolated from each glycerol stock by miniprep (Zymoprep II kit, Zymo Research). Sequencing libraries were prepared with molecular indexing based on a method described in.²¹⁹ Briefly, a primer containing molecular index was annealed to a region upstream of the peptide sequence on the scHLA-pIM627 phagemid DNA and single primer extension reaction was carried out with Kapa HiFi DNA polymerase (Roche Diagnostics). Following a reaction cleanup with ExoProStar (GE Healthcare) and column purification (Macherey-Nagel), second PCR reaction was carried out with primers specific to the primer containing molecular index and a reverse primer designed to the β 2M gene. Sequencing libraries were prepared from purified PCR products (Ampure XP beads, Beckman coulter) using NebNext Ultra II DNA library prep kit (NEB) according to manufacturer's instructions. Library QC was performed with Agilent bioanalyser HS kit (Agilent biosystems) and library DNA concentrations were measured with Qubit HS dsDNA kit (Life technologies). Libraries were sequenced using Illumina V3 SBS chemistry on the MiSeq sequencer. Basecalling and sample demultiplexing was performed using MiSeq reporter to generate fastq files and were processed with custom analysis pipeline. Peptide repertoire analysis was performed using excel and sequence Logos were generated using IceLogo standalone tool.²²⁰ Sequence clustering analysis was performed with GibbsCluster-2.0 web server using default settings.²²¹

X-ray crystallography

Crystals were grown at 18°C by vapour diffusion via the sitting drop technique. All crystallization screening and optimisation experiments were completed with an Art-Robbins Phoenix dispensing robot (Alpha Biotech Ltd, U.K.). 200 nL of 10-15 mg/ml TCR-pHLA complex mixed at a 1:1 molar ratio was added to 200 nL of reservoir solution. Intelli-plates were then sealed and incubated in a crystallization incubator at 18°C (RuMed, Rubarth Apperate GmbH, Germany) and analyzed for crystal formation using a Rock Imager 2 (Formulatrix, Bedford, MA USA). Crystals selected for further analysis were cryoprotected with ethylene glycol to 25% and then flash-cooled in liquid nitrogen in Litho loops (Molecular Dimensions, UK). For WT1_ α 7 β 2-A2-RMF, optimal crystals were obtained in 0.1 M HEPES pH 7, 0.1M ammonium, 20% v/v Sok-CP7. For NYBR1-A2-SLS, optimal crystals were obtained in Pact premier (Molecular Dimensions) condition B07 (0.2 M sodium chloride, pH 6.0, 0.1 M MES and 20% PEG 6000. Diffraction data were collected at several different beamlines at the Diamond Light Source, Oxford, using a Dectris Pilatus 6M detector. Using the rotation method, 1000 frames were recorded each covering 0.2° of rotation. Reflection intensities were estimated with the XIA2 package and the data were scaled, reduced

and analysed with AIMLESS and the CCP4 package. TCR/pHLA complex structures were solved with molecular replacement using PHASER, using PDB 4I4W as a starting model for pHLA, and PDB 3O4L as a starting model for NYBR1 TCR. Accession code WT1_α7β2-A2-RMF: TBD and NYBR1-A2-SLS: 6R2L.

MD Simulations and MMPBSA calculations

Periodic boundary simulations and Molecular Mechanics Poisson–Boltzmann Surface Area (MMPBSA) calculations were performed with the Amber16 suite of programs.²²² X-ray crystal structures of the seven TCR/Fab-pHLA complexes investigated were used as the starting points for molecular dynamics (MD) simulations. Missing residues were added using Modeller v9.⁷⁵ MolProbity was used to modify histidine tautomerisation states (tautomerisation states used can be found in **Supp Table S1**) and Asn/Gln side chain orientations under the criteria of optimising the internal hydrogen bonding network. PropKa 3.0¹⁵⁸ was used to check the likely protonation states of all titratable residues for pH 7 (all residues were modelled in their standard protonation states). Each system was solvated in a rectangular box of water (with all crystallographic waters retained), extending at least 10 Å away from any protein atom. Sodium or chloride ions were added as necessary to neutralise the total system charge. The ff14SB⁶⁰ and TIP3P¹⁵⁹ force-fields were used to describe protein and water molecules, respectively. Following minimisation, heating and equilibration (see section titled “Structure Equilibration Procedure” below for further details), each system was subjected to two, 500 ns long production MD simulations (random velocity vectors assigned upon heating) in the NPT ensemble (1 atm, 298 K). Production MD simulations were run with the SHAKE algorithm applied, a 2 fs time step and a collision frequency of 1 ps⁻¹. An 8 Å direct space non-bonded cut-off was applied with long range electrostatics evaluated using the particle mesh Ewald algorithm.⁷² Hydrogen bonding (including water bridged hydrogen bonds) and vdW interactions were evaluated from snapshots saved every 10 ps, using the last 450 ns of each trajectory (900 ns per TCR/Fab-pHLA). A hydrogen bond was defined as ‘on’ if the donor acceptor distance was within 3.0 Å and the donor hydrogen acceptor angle was within 45° to 180°. A vdW interaction was defined as ‘on’ if two heavy atoms were within 4 Å of one another.

Molecular Mechanics Poisson–Boltzmann Surface Area (MMPBSA) calculations were performed using MMPBSA.py.MPI³⁹, using 25 independent (random velocity vectors assigned upon heating) 4 ns long MD simulations (see Supporting Information for further details) for each structure. From each run, 300 equally spaced snapshots were taken from the last 3 ns of each MD simulation for MMPBSA calculations, giving a total of 7500 frames per complex. MMPBSA calculations were

performed with an implicit salt concentration of 150 mM, and with 30 explicit water molecules (which were all defined as part of the receptor) retained in each snapshot. The 30 closest water molecules to any binding site residue heavy atom were retained in each snapshot by using the ‘closest’ command in CPPTRAJ¹⁶⁰, (see section titled “Selection of explicit waters for MMPBSA calculations” below for further details).

Structure Equilibration Procedure.

The following procedure was used to prepare for production MD simulations for both the long time-scale (2 x 500 ns) and short time scale (25 x 4 ns) simulations used in this study. First, hydrogen atoms and solvent molecules were relaxed with 500 steps of steepest descent followed by 500 steps of conjugate gradient (using 10 kcal mol⁻¹ Å⁻¹ positional restraints on all protein heavy atoms). The system was then heated linearly from 50 K to 298 K (NVT ensemble) over the course of 200 ps (retaining the 10 kcal mol⁻¹ Å⁻¹ positional restraints on all protein heavy atoms). The whole system was then minimised for a further 500 steps of steepest descent followed by 500 steps of conjugate gradient with 5 kcal mol⁻¹ Å⁻¹ positional restraints on all C α carbon atoms. Retaining the C α carbon restraints, each system was again heated from 25 K to 298 K over the course of 50 ps in the NVT ensemble. The C α carbon restraints were then gently removed in linear steps of (5, 4, 3, 2, 1, 0 kcal mol⁻¹ Å⁻¹) of 10 ps each in the NPT ensemble. Following this, production MD simulations were run. For NVT simulations, the timestep was set to 1 fs (with the SHAKE algorithm applied) and a collision frequency of 1 ps⁻¹ was used with Langevin temperature control. Simulations in the NPT ensemble were performed with a timestep of 2 fs (with the SHAKE algorithm applied), using a Berendsen barostat for pressure control (1 ps pressure relaxation time) and Langevin temperature control (collision frequency of 1 ps⁻¹).

Selection of explicit waters for MMPBSA calculations

The InterfaceResidues.py script (available at <https://pymolwiki.org/index.php/InterfaceResidues>) run through PyMOL on each crystal structure was used to identify in an unbiased manner the binding site residues for each TCR/Fab-pHLA. Criteria for selecting interfacial residues were set based on the results of Maffucci *et al.*²²³ (Cut-off for change in solvent accessible surface area was set to 0.5 Å² for all residues). Following this, CPPTRAJ¹⁶⁰ (part of the AmberTools suite of programmes), was used to select the closest 30 water molecules to the selected interfacial residues using the ‘closest’ command. Stripping solvent using large residue selections such as the ones generated in these calculations can be time consuming if performed on the entire water box. To greatly increase the speed of this calculation, we performed two ‘closest’ calculations (one to

remove most of the waters in the periodic box, and the second to select the 30 closest waters molecules for MMPBSA calculations). An example script of how to do this within CPPTRAJ is available upon request.

Conflict of interest

CJH, JP, BW, AL, NL, KL, TB, KL, SH, PC, MH, RJP, TES, BC, AJ, PM, AV, MS, MA, NL, RR, SH, VS, ML, BKJ and DKC are employees of Immunocore LTD. The authors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

CJH, RC, JP, BW, AL, NL, KL, TB, KLowe, SH, PC, MH, RJP, TES, BC, AJ, PM, AV, MS, MA, NL, RR, SH, VS, ML, CP, MWK, PJR, BKJ and DKC performed and/or directed experiments, analysed data, and critiqued the manuscript. DKS and BKJ conceived, and directed the project. DKC wrote the manuscript.

Acknowledgements

The authors would like to thank Diamond Light Source for beam time, and the staff for assistance with crystal testing and data collection. The authors would also like to thank Kate Lowe, Ita O'Kelly and Michelle McCully for critically reading the manuscript.

Supporting Information for: Specificity of bispecific TCRs and antibodies targeting peptide-HLA

Christopher J. Holland^{*1}, Rory Crean^{*2,3}, Johanne Pentier^{*1}, Ben de Wet^{*1}, Angharad Lloyd^{*}, Nikolai Lissin¹, Katy Lloyd¹, Thomas Blicher¹, Stephen Hearty¹, Paul Conroy¹, Miriam Hock¹, Robert J. Pengelly¹, Thomas E. Spinner¹, Brian Cameron¹, Anitha Jeyanthan¹, Peter Molloy¹, Annelise Vuidepot¹, Malkit Sami¹, Milos Aleksic¹, Nathaniel Liddy¹, Ross Robinson¹, Stephen Harper¹, Velupillai Srikannathasan¹, Marco Lepore¹, Chris Pudney², Marc W. van der Kamp⁴, Pierre J. Rizkallah⁵, Bent K. Jakobsen¹ and David K. Cole¹

¹Immunocore Ltd., Milton Park, Abingdon, UK.

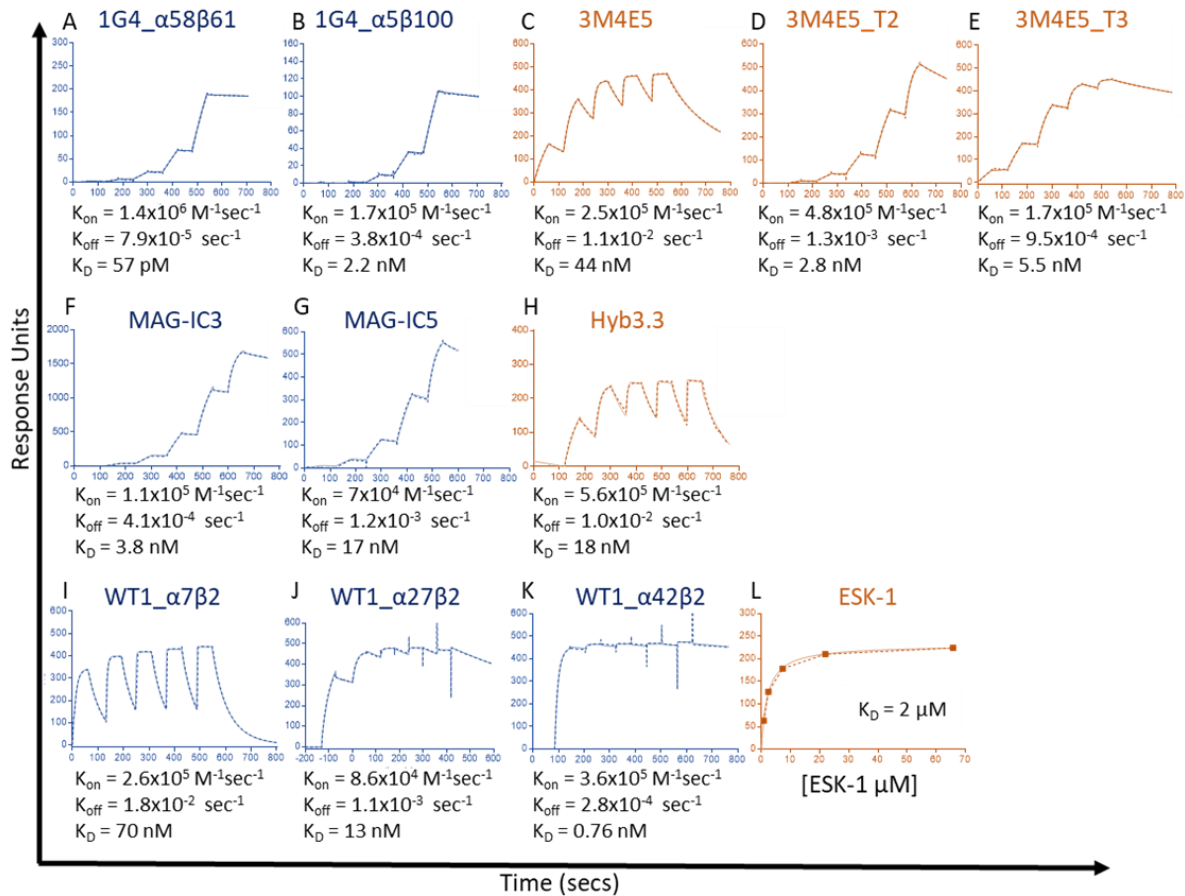
²Department of Biology and Biochemistry, University of Bath, Bath, UK.

³Doctoral Training Centre in Sustainable Chemical Technologies, University of Bath, Bath, UK.

⁴School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, UK.

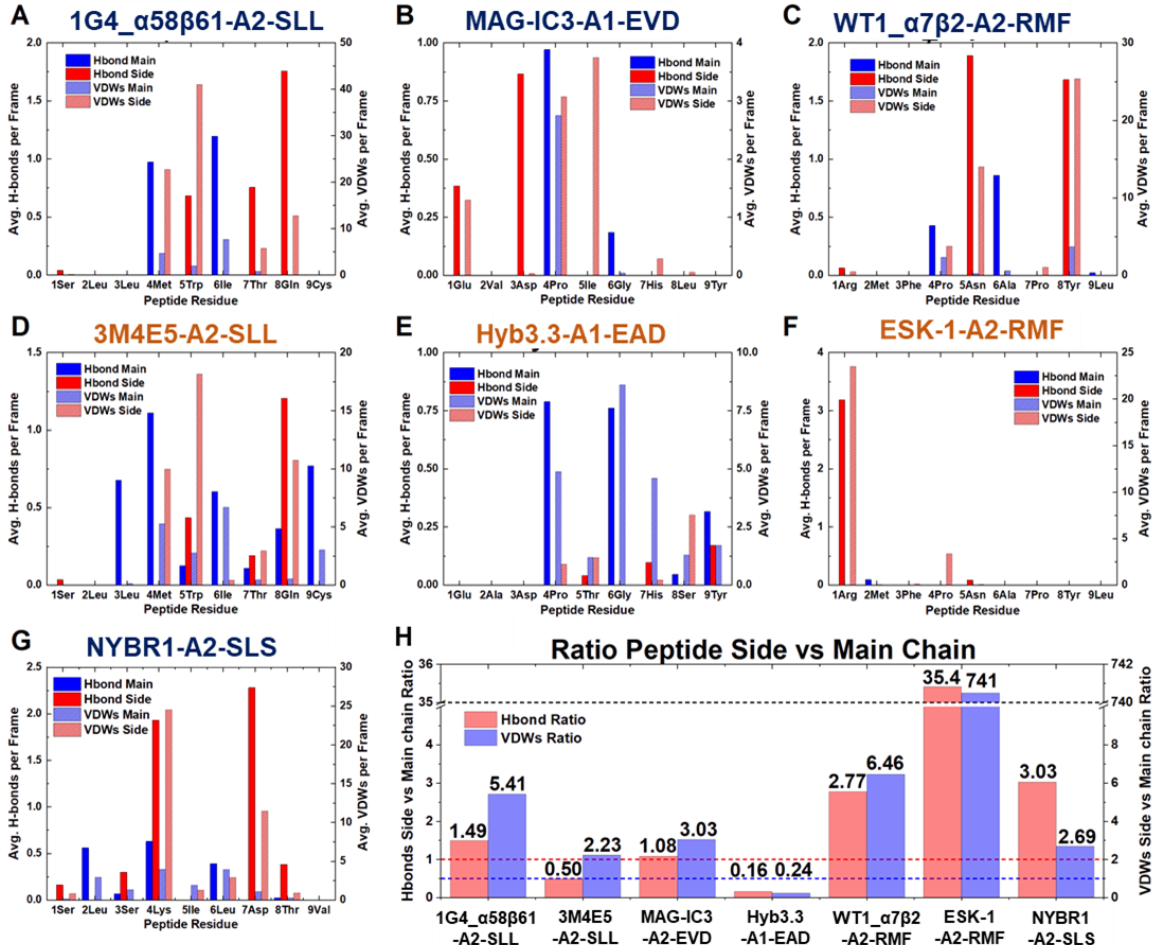
⁵Cardiff University School of Medicine, Heath Park, Cardiff, CF14 4XN, UK.

*These authors contributed equally to this study.

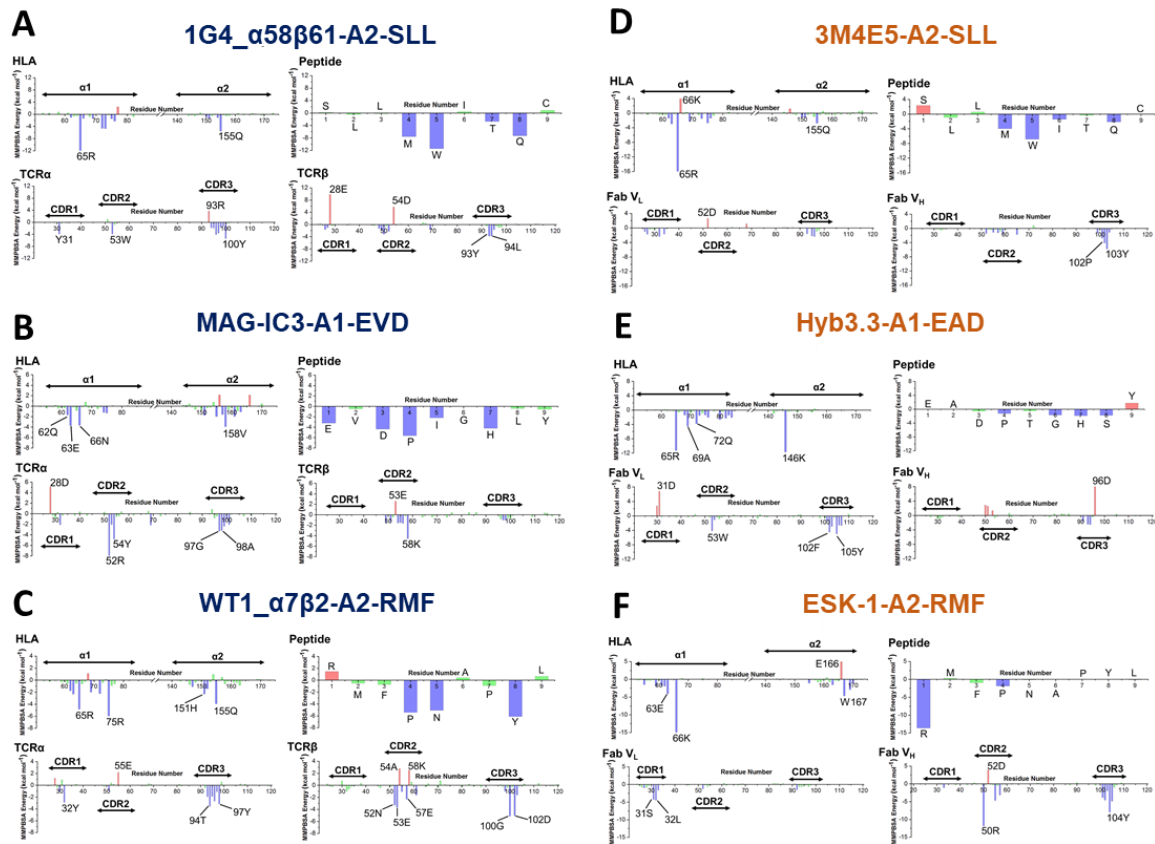


Supplementary Figure S1: – Binding affinity analysis of affinity-enhanced TCRs and TCR-mimics.

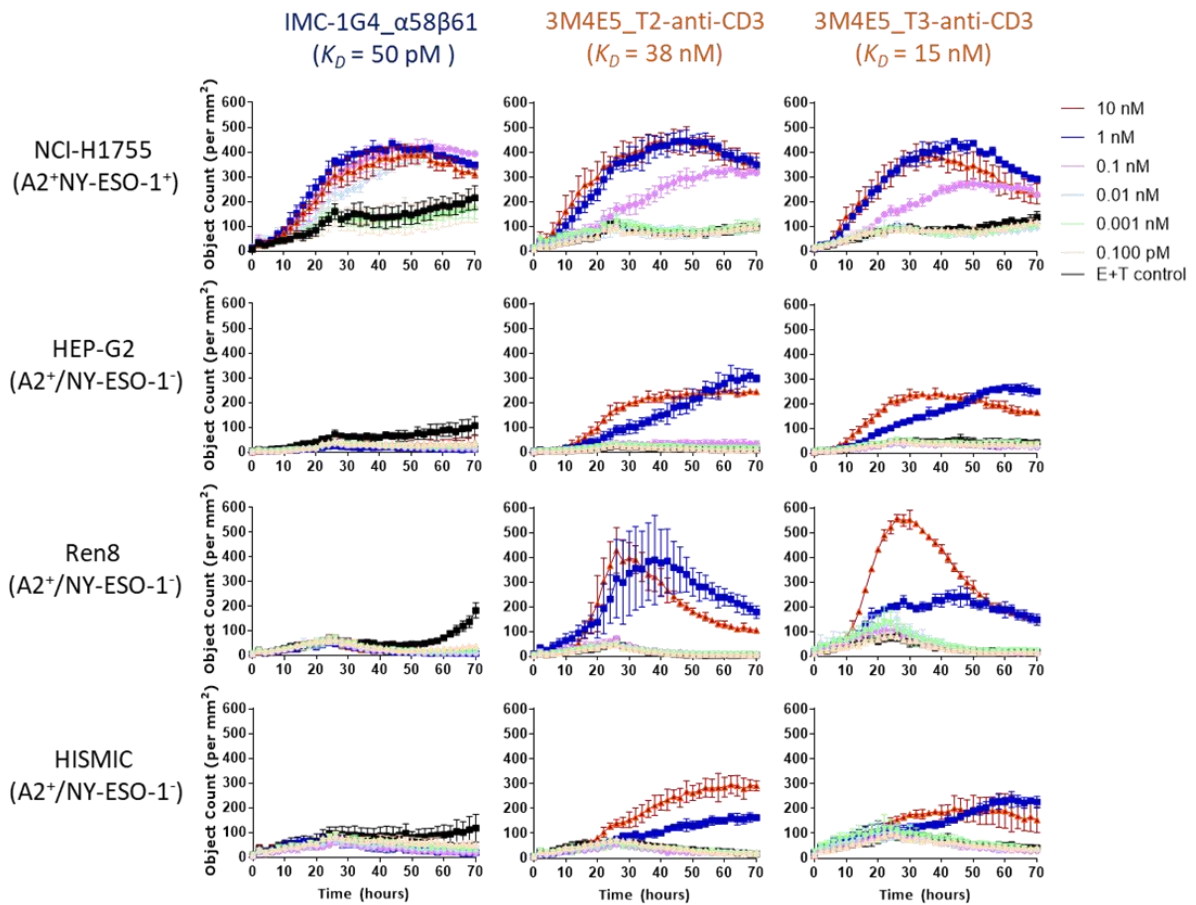
The interaction between each affinity-enhanced TCR, or TCR-mimic with cognate pHLA molecules were analysed using surface plasmon resonance. Binding affinities of the affinity-enhanced TCRs and TCR-mimics were determined using single cycle kinetic analysis, or equilibrium binding analysis (for ESK-1). Five injections of each reagent were performed using 3:1 dilution between injections. Raw data (dotted line) and fits (solid line) are shown for each plot. Representative data from three independent experiments are shown. (A-B) A2-SLL affinity-enhanced TCRs (1G4_α58β61 and 1G4_α5β100), (C-E) A2-SLL TCR-mimics (3M4E5, 3M4E5_T2 and 3M4E5_T3), (F-G) A1-EVD affinity-enhanced TCRs (MAG-IC3 and MAG-IC5), (H) Hyb3-A1-EAD, (I-K) A2-RMF affinity-enhanced TCRs (WT1_α7β2, WT1_α27β2 and WT1_α42β2), (L) ESK-1-A2-RMF.



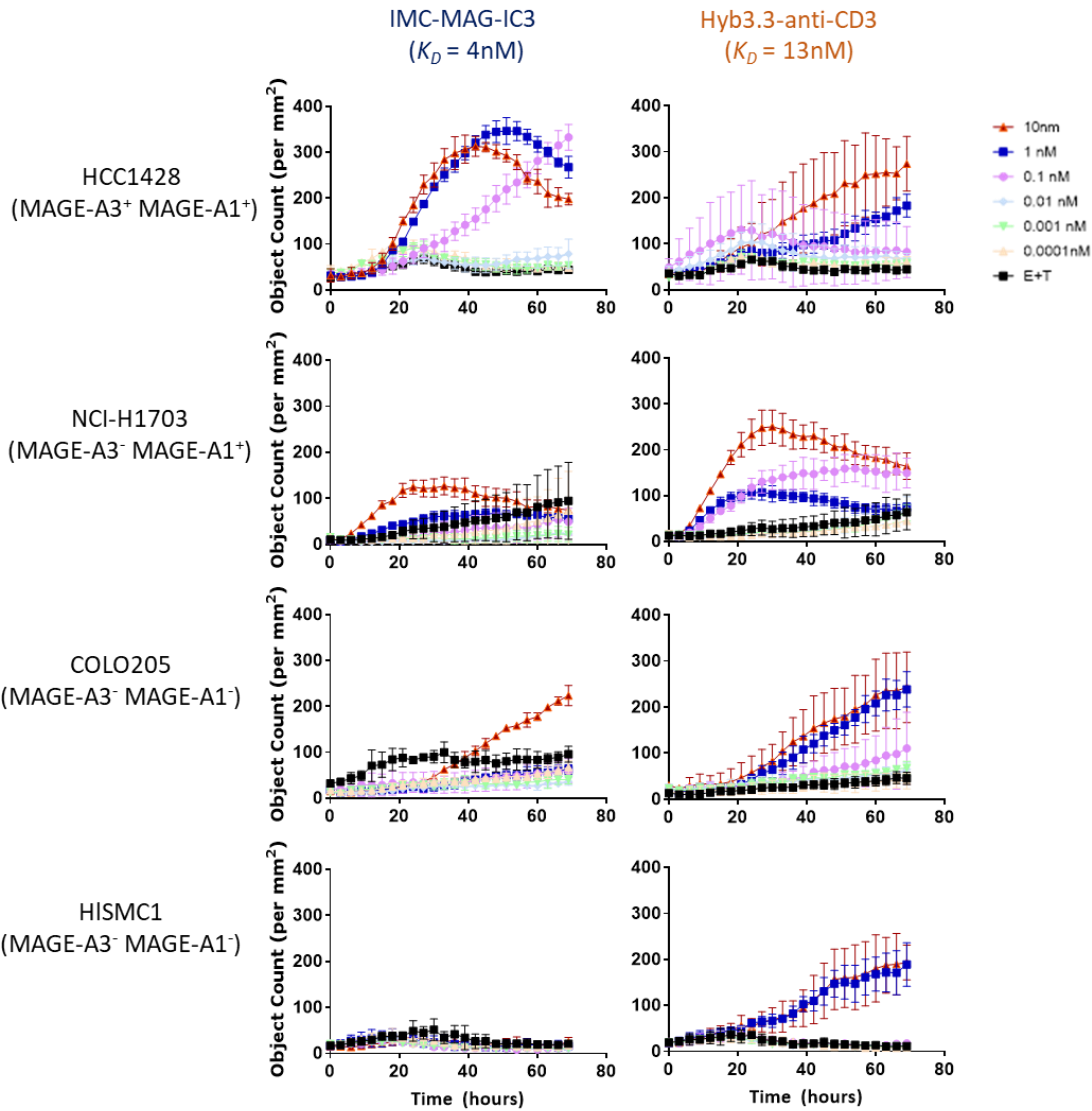
Supplementary Figure S2: – Molecular dynamic analysis of the TCR and TCR-mimic-peptide interface. Molecular dynamics was performed to access the average number of side chain or main chain peptide contacts over time. Average number of hydrogen bonds and vdWs interactions formed per frame between the peptide and the TCR/Fab over the course of our long time scale MD simulations. Per-residue peptide contributions are divided into main (blue and light blue) and side (red and light red) chain contributions. A) 1G4_α58β61-A2-SLL, B) MAG-IC3-A1-EVD, C) WT1_α7β2-A2-RMF, D) 3M4E5-A2-SLL, E) Hyb3.3-A1-EAD, F) ESK-1-A2-RMF. G) NYBR1 TCR, H) Ratio of total peptide side chain against peptide main chain hydrogen bonds (red) and vdWs interactions (blue) for each TCR/Fab-pHLA. Ratios are stated above each bar to 3 significant figures. The red and blue dotted lines represent the y-axis location for a ratio of 1 for Hydrogen bonds and vdWs interactions respectively. A black dotted line denotes the position of an axis break used to allow ESK-1 Fab to be plotted on the same graph.



Supplementary Figure S3: Per-residue decomposition of the binding free energy obtained from MMPBSA calculations for key regions of (A) 1G4_α58β61-A2-SLL, (B) MAG-IC3-A1-EVD, (C) WT1_α7β2-A2-RMF, (D) 3M4E5-A2-SLL, (E) Hyb3.3-A1-EAD, (F) ESK-1-A2-RMF. A more negative value indicates increased favourability towards binding. Bars are coloured as follows: blue, less than -1 kcal mol^{-1} ; green, between -1 and 1 kcal mol^{-1} ; and red, greater than 1 kcal mol^{-1} .

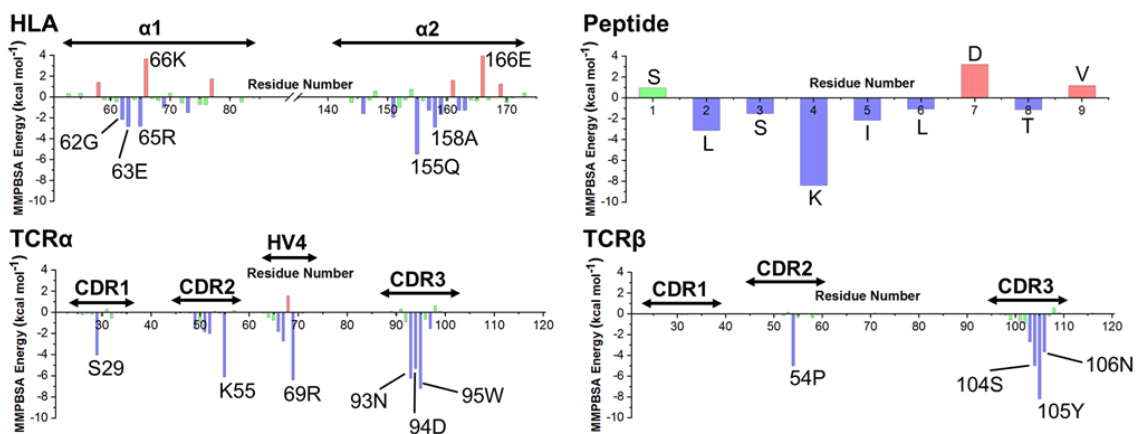


Supplementary Figure S4: The activity of IMC-1G4_α58β61, 3M4E5_T2-anti-CD3 and 3M4E5_T3-anti-CD3 was tested against HLA-A*02:01+/NY-ESO-1+ (NCI-H1755) and HLA-A*02:01+/NY-ESO-1- (HEP-G2, Ren8 and HISMIC) cells lines using Incucyte killing assays. Experiments were performed according to the manufacturer's instructions (Sartorius, UK). 100,000 PBMCs per well were added to 10,000 target cells per well. Target cells were incubated on plates overnight before addition of ImmTAC molecules, or TCR-mimic bispecific reagents followed by PBMCs. Cells were scanned every 3 hours for 70 hours. Data are plotted using cell death over time. Data from 1 out of 2 donors shown.



Supplementary Figure S5: The activity of IMC-MAG-IC3 and Hyb3.3-anti-CD3 was tested against HLA-A*01:01+/MAGE-A3+ (HCC1428), HLA-A*01:01+/MAGEA1+ (HCC1428 and NCI-H1703) and HLA-A*01:01+/MAGE- (COLO205 and HISM1) cells lines using Incucyte killing assays. Experiments were performed according to the manufacturer’s instructions (Sartorius, UK). 100,000 PBMCs per well were added to 10,000 target cells per well. Target cells were incubated on plates overnight before addition of ImmTAC molecules, or TCR-mimic bispecific reagents followed by PBMCs. Cells were scanned every 3 hours for 70 hours. Data are plotted using cell death over time. Data from 1 out of 2 donors shown.

NYBR1-A2-SLS



Supplementary Figure S6: Per-residue decomposition of the binding free energy obtained from MM-PBSA calculations for key regions of NYBR1-A2-SLS. A more negative value indicates increased favourability towards binding. Bars are coloured as follows: blue, less than -1 kcal mol⁻¹; green, between -1 and 1 kcal mol⁻¹; and red, greater than 1 kcal mol⁻¹.

Supplementary Table S1. Summary of cancer-targeting TCR-mimic antibodies reported in the literature.

TCR-mimic	Target/peptide	HLA-A*	In vivo data	Structure	Refs
3M4E5	NY-ESO-1/SLIMWITQC	02:01	3M4E5_T2	3M4E5	34,177,224
ESK-1	WT1/RMFPNAPYL	02:01	ESK-1	ESK-1	36,178,225
Hyb3.3, G8	MAGE-A1/EADPTGHSY	01:01	Hyb3.3	Hyb3.3	35,181,226
PR20	PRAME/ALYVDSLFFL	02:01	PR20	No	176
38	LMP2A/CLGGLLTMV	02:01	38	No	214
G2D12	GP100/KTWGQYWQV	02:01	No	No	227
1A9, G1	GP100/IMDQVPFSV	02:01	No	No	227
2F1	GP100/YLEPGPVTA	02:01	No	No	227
GPA7	GP100/ITDQVPFSV	02:01	GPA7	No	228
4A9	hTERT/ILAKFLHWL	02:01	No	No	229
3H2	hTERT/RLVDDFLLV	02:01	No	No	229
M2B1	MUC1/LLTTLTVV	02:01	No	No	230
7D4	MAGE-A3/FLWGPRALV	02:01	No	No	231
RL4B/3.2G1, 1B10	hCG β /GVLPALPQV	02:01	RL4B/3.2G1	No	232,233
3F9	hCG β /TMTRVLQGV	02:01	No	No	233
1B8	Her2/ KIFGSLAFL	02:01	No	No	234
CAG10	MART-1/EAAGIGILTV	02:01	CAG10	No	235
Fab-D2	TARP/FLRNFSML	02:01	No	No	236
T1-116C, T1-29D	p53/RMPEAAPV	02:01	T1-116C	No	237,238
T2-108A	p53/GLAPPQHLIRV	02:01	No	No	238
TA2	Tyrosinase/YMDGTMSQV	02:01	No	No	182
RL6A	p68/YLLPAIVHI	02:01	RL6A	No	239
RL21A	MIF/FLSELTQQL	02:01	RL21A	No	240
8F4	Proteinase 3/VLQELNVTV	02:01	No	No	241
#131	HA-1H/VLHDDLLEA	02:01	No	No	211

Supplementary Table S2. Histidine tautomerisation state assignments for all MD simulations.

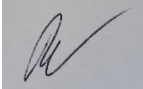
Structure	HID^a Tautomerisation State
α 58 β 61 TCR	Chain A: 3, 70, 74, 93, 114, 145, 260. Chain E: 151.
3M4E5 Fab	Chain A: 3, 70, 74, 93, 114, 145, 260. Chain D: 51.
MAG-IC3 TCR	Chain A: 3, 70, 93, 260. Chain E: 153.
Hyb3.3 Fab	Chain A: 3, 70, 93, 260. Chain D: 172, 212. Chain E: 95B , 189.
α 7 β 2 TCR	Chain A: 3, 70, 74, 93, 114. Chain B: 51. Chain D: 77. Chain E: 208.
ESK-1 Fab	Chain A: 3, 70, 93, 114. Chain D: 192.
NYBR1 TCR	Chain A: 3, 70, 74, 93, 151, 260. Chain E: 158.

^aHID corresponds to a histidine residue which is singly protonated on its N δ 1 nitrogen, with all other histidine residues simulated as singly protonated on their N ϵ 2 nitrogen.

Chapter 5:

Drivers for High Affinity TCR-pHLA Interactions: Insights from Structure and Dynamics

The focus in **Chapter 4** was to determine protein engineering principles for developing TCRs that are highly specific towards their given pHLA target. In this chapter, we now turn to determining engineering principles for affinity enhancement of TCRs. In order to do this, we performed structural analysis, MD simulations and free energy calculations on four case studies of TCR affinity maturation available from the literature. We first demonstrated the limitations of contact analysis (either from X-ray crystal structures or MD simulations) to rationalise changes in affinity. We then evaluated how the flexibility of the TCR had changed over the course of affinity maturation, identifying instances in which the apo high affinity (HA) TCR was more rigid than its corresponding apo wild-type (WT) TCR. This is an energetically favourable effect for the HA-TCR, as the entropic penalty associated with solute rigidification upon binding would be reduced (relative to the WT-TCR). This beneficial effect was largely found to be “compensated” for however, by the HA-TCR binding the pHLA bound structure tighter than the corresponding WT-TCR does. Finally, we used free energy calculations to identify at the per-residue level what changes/mutations to the TCR have driven affinity enhancement across all four case studies. Our free energy calculations were able to correctly predict the affinity hierarchy for all four case studies, and comparison of the WT and HA energetic footprints show the binding hotspots on the TCR and pHLA to be largely preserved. This has important implications in the design of specific TCRs, as preservation of the energetic footprint used by the thymically selected WT-TCR is arguably more likely to create a non-cross-reactive TCR. In this study, I performed all of the structural analysis, MD simulations and free energy calculations. I also wrote the draft manuscript.

This declaration concerns the article entitled:									
Drivers for high affinity TCR-pHLA interactions: insights from structure and dynamics									
Publication status (tick one)									
draft manuscript	<input checked="" type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input type="checkbox"/>
Publication details (reference)	Rory M. Crean, Chris R. Pudney, David K. Cole. and Marc W. van der Kamp. (Draft Manuscript).								
Candidate's contribution to the paper (detailed, and also given as a percentage).	<p>The candidate contributed to/ considerably contributed to/predominantly executed the...</p> <p><u>Formulation of ideas (60 %)</u> Ideas were formulated by RMC, CRP, DKC and MvdK.</p> <p><u>Design of methodology: (50 %)</u> Methodology for simulations was formulated by RMC, MvdK.</p> <p><u>Experimental work: (100 %)</u> Structural analysis, MD simulations and free energy calculations carried out by RC.</p> <p><u>Presentation of data in journal format (70 %)</u> Complete manuscript draft by RMC. Manuscript review and redraft by MvdK. (Changes by CRP and DKC to come)</p>								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed						Date	31/07/2019		

Drivers for high affinity TCR-pHLA interactions: insights from structure and dynamics

Rory M. Crean,^{1,2} Christopher R. Pudney,^{*,1,3} David K. Cole,^{*,4} and Marc W. van der Kamp.^{*,5}

¹Department of Biology and Biochemistry, ²Doctoral Training Centre in Sustainable Chemical Technologies, ³Centre for Therapeutic Innovation, University of Bath, Bath, BA2 7AY, UK.

⁴Immunocore Ltd., Abingdon OX14 4RY, UK

⁵School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD, UK.

*Correspondence: Dr Christopher R. Pudney, E-mail: c.r.pudney@bath.ac.uk, Dr David K.

Cole, E-mail: david.cole@immunocore.com, and Dr Marc W. van der Kamp Email:

marc.vanderkamp@bristol.ac.uk,

Keywords. Cancer immunotherapy, peptide-human leukocyte antigen (pHLA), T-cell receptor (TCR), molecular dynamics (MD).

Abstract

Immuno-oncology (IO) approaches that utilise T-cell receptors (TCRs) are becoming highly attractive because of their potential to target virtually all cellular proteins, including cancer specific proteins, *via* the recognition of peptide human leukocyte antigen complexes (pHLA) that are presented at the cell surface. However, because natural TCRs generally recognise cancer derived pHLAs with very weak affinities (high μM to low mM range), efforts have been made by several companies and academic laboratories to enhance their affinities, in some cases by several million-fold. Here, we investigate whether shared mechanisms drive the enhancement in affinity by studying the crystal structures of several published, and one new, affinity enhanced TCRs compared with structures of their natural progenitor TCRs. Additionally, we performed in depth molecular dynamic simulations to better understand the nature of the affinity enhancements. These data demonstrate that affinity enhancements can be achieved within the natural TCR-pHLA binding mode *via* relatively subtle modifications to the interface contacts. However, the individual energetic components of the TCR-pHLA interaction that governed the affinity enhancements were distinct and highly variable for each TCR under investigation. Our data demonstrate that that native TCR binding mode has the potential to bind pHLA with antibody-like (up to low pM) affinities *via* a range of different energetic mechanisms. This finding raises an important biological questions:

why are natural TCRs selected to bind to pHLA in the μM affinity range? Finally, this first comprehensive analysis of affinity enhanced TCRs has important implications for the future rational design of affinity enhanced TCRs for cancer therapy.

Introduction

$\alpha\beta$ T-cell receptor (TCRs) recognition of short antigenic peptide fragments presented at the cell surface by human leukocyte antigens (pHLA) governs T-cell immunity. These peptide fragments represent virtually all intracellular proteins, allowing TCRs to access a much larger pool of potential therapeutic targets than monoclonal antibodies (mAbs), which primarily bind to extracellular antigens.³² This advantage has encouraged the development of soluble engineered TCRs as therapeutics for viral and cancerous diseases.^{43,174,175} These soluble TCRs have been designed as bispecific T-cell engagers, making use of their pre-existing antigen recognition site to bind a specific pHLA molecule on a target cell, and an immune effector function to recruit and activate T-cells.²⁶ This approach of utilising a soluble bispecific TCR to target cancer has been shown to induce tumour regression³³ and clinical trials are currently under way for multiple diseases.

However, unlike mAbs, which can utilise somatic hypermutation to generate affinities for their target antigen in the nM-pM range, naturally occurring TCRs bind pHLA with relatively weak affinities ($\sim\mu\text{M}$) and short half-lives (on the timescale of seconds).¹⁸ These characteristics are undesirable for therapeutic molecules. Thus, a number of approaches^{46,47,175,242–244} usually focussed on introducing affinity enhancing mutations within the six complementarity-determining region (CDR) loops that comprise the TCR binding site (see **Figure 32A+B**), have been used to improve the binding characteristics of the TCR-pHLA interaction. One approach, using phage display, has been shown to yield TCRs with affinities up to the low pM range, and subsequently much longer half-lives (in the hours to days range) for their target pHLA complex.¹⁷⁵ Soluble bispecific TCRs generated using this phage display approach (ImmTAC molecules) can re-direct T-cell responses against cancer cells, leading to tumour regression³³, presenting as a few as 10 antigen specific pHLA per APC¹⁹.

To ensure coverage against all possible peptide antigens, the $\sim 2 \times 10^7$ TCRs (produced per human) are highly cross-reactive¹²⁸, with prior studies showing that a single TCR is able to bind over a million different peptide antigens presented on a single HLA.¹⁶ This cross-reactivity has been rationalised as being induced by specific residue “hotspots” on the pHLA, meaning if all or most of these residues are present, the TCR will bind to a reasonable degree.²¹² The dangers of such cross-reactivity (binding endogenous pHLA molecules to induce an autoimmune response) for TCRs are largely abrogated by thymic selection, in which TCRs with affinities too high for any endogenous pHLA molecule are deleted. Whilst high affinity (HA) TCRs have obvious advantages from a therapeutic

perspective, the mutations made to increase their affinity come with the risk of inducing cross-reactivity with endogenous pHLA molecules.²² It is therefore of significant interest to understand if affinity matured TCRs tend to preserve the “energetic footprint”/residue hotspots used by the thymically selected wild-type (WT) TCR, as this is likely to reduce cross-reactivity with a self-pHLA molecule. Further, the development of computational “assays” that would allow one to determine the extent to which mutations have altered this energetic footprint could be of value in rational design approaches.

With many possible mechanisms for affinity enhancement (such as improved electrostatics, burial of hydrophobic matter, expulsion of unfavourable water molecules and a reduction in the entropic cost of solute binding, by rigidification of the protein(s)), one would like to know if there are common methods by which TCRs can enhance affinity. It would be very challenging to ascertain the above information experimentally, whereas computational methods can provide detailed insight at the atomic level regarding the features of affinity enhancement. Further, the development of computational “assays” that would allow one to determine the extent to which mutations have altered this energetic footprint could be of value in rational design approaches.

Our study aims to provide fundamental insight into what drives affinity maturation in TCR molecules targeting pHLA. Such increased understanding is of direct importance in the rational design of high affinity and antigen-selective TCRs. Herein, we consider four separate case studies of TCR affinity maturation from the literature, and perform a combination of structural analysis, molecular dynamics (MD) simulations and binding free energy calculations to determine the drivers for affinity enhancement in all four cases. All case studies chosen, crystal structures of both the WT TCR and the corresponding HA variant(s) of the TCR in complex with the pHLA were compared, providing input for structural analysis and good starting points for MD simulations. Our first case study (**Figure 32C**) is composed of the WT and four HA variants targeting the NY-ESO-1₁₅₇₋₁₆₅ cancer-testis antigen-derived peptide (sequence: SLLMWITQC), presented by HLA-A*02:01 (A2-SLL).^{43,45,245} The remaining three case studies are pairs of a single WT and HA variants with two of the three pairs (DMF5 and MEL5, **Figure 32C**) targeted towards the MART-1₂₆₋₃₅ peptide bound to HLA-A*0201, with peptide sequence ELAGIGILTV.^{175,246} The final pair of TCRs (Tax A6, **Figure 32C**) recognise the HTLV-1₁₁₋₁₉ peptide, presented by HLA A*0201 with sequence LLFGYPVYV.¹⁷⁵

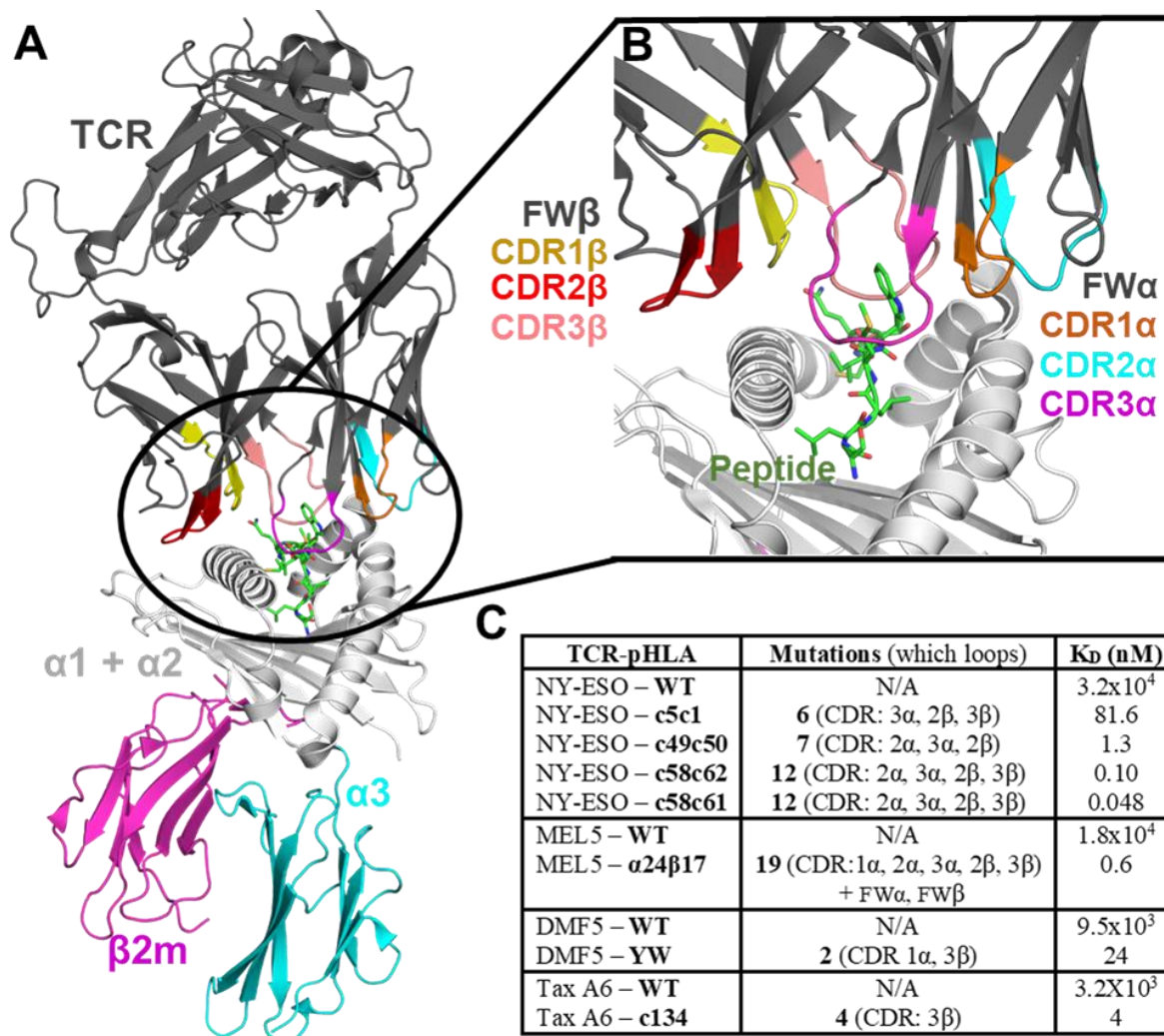


Figure 32: (A) Exemplar structure of a TCR-pHLA complex, with a different portions of the TCR and pHLA labelled. (B) Zoom in on the binding TCR-pHLA showing the 6 CDR loops responsible for binding the pHLA. (C) Table of all TCR-pHLA complexes investigated in this study. Note the α and β -chain framework (FW α and FW β) are non-hypervariable loops that flank the CDR loops. (A complete list of all mutations made onto each CDR loop can be found in **Table S1**).

Results and Discussion

Structural comparisons of WT and affinity matured TCRs show preservation of the WT binding mode. We analysed the crystal structures of all TCR-pHLA complexes investigated herein (**Table 5**), focussing first on how the TCR engaged with the pHLA molecule. We first measured how the relative orientation of the TCR variable domains (V α and V β) may have changed as a result of affinity maturation by measuring their

TRangle²¹ parameters. To compare how similar the variable domain orientations are for two different TCRs, one can calculate the d_{TRangle} (see Methods) between the two TCRs, with values $<10^\circ$ suggesting the orientations are highly similar.²¹ Comparison of the d_{TRangle} values (for WT vs HA, see **Table 5**) structures show no significant change in the relative orientation upon affinity maturation. We also calculated the “docking angle” for each TCR (**Table 5**), which measures the angle at which the TCR engages the pHLA, with canonical TCR docking angles in the range of 40-85°.⁵ For all cases, we observe no significant change in the docking angle upon affinity maturation. Interestingly, we note that the docking angles of both Tax A6 and DMF5 sit just below the previously described canonical docking angle range (**Table 5**). The above observations suggest (at least in these cases) that affinity maturation has not significantly impacted the core scaffold of the TCR

Table 5: Structural analyses of all TCR-pHLAs complexes under investigation. The d_{TRangle} is a measure of the TCR variable domains orientational similarity for 2 different structures (see Methods).⁴¹ For this, we compare the WT to each HA TCR. The docking angle (which describes the angle the TCR engages the pHLA) and d_{TRangle} were determined using the STCRDAB server. Buried solvent accessible surface area (BSASA) was determined using the LCPO algorithm within CPPTRAJ. Hydrogen bonds (H-bonds) are defined for donor-acceptor distances $\leq 3.5 \text{ \AA}$ and donor-hydrogen-acceptor angles 135-225°. Van der Waals (vdWs) interactions are considered for all heavy atoms within 4 \AA of each other.

TCR-pHLA	NY-ESO					MEL5		DMF5		Tax A6	
	WT	c5c1	c49c50	c58c62	c58c61	WT	$\alpha 24\beta 17$	WT	YW	WT	c134
$d_{\text{TRangle}} (^\circ)$	N/A	2.8	2.9	3.7	3.0	N/A	3.3	N/A	2.5	N/A	1.8
Docking Angle (°)	69.7	65.8	66.1	66.0	66.6	48.4	43.0	34.2	32.6	34.8	34.2
BSASA (\AA^2)	2200	2010	2197	2257	2203	1930	2382	2065	2206	2197	2320
H-bonds											
Total	12	15	14	13	12	17	12	9	10	13	13
Peptide	6	6	6	6	6	7	6	5	6	7	5
HLA	6	9	8	7	6	10	6	4	4	6	8
vdWs contacts											
Total	234	228	245	231	231	201	214	158	205	202	221
Peptide	109	107	116	108	111	68	53	54	68	94	92
HLA	125	121	129	123	120	133	161	114	137	108	129

or how it docks to bind the pHLA.

We next considered the number of hydrogen bonds (H-bonds) and van der Waals (vdWs) interactions between the TCR and pHLA (**Table 5**). No clear relationship between affinity and the number of H-bonds and/or vdWs interactions formed was observed. As neither measure considers the relative strength or occupancy of different H-bonds or vdWs

contacts, these results are unsurprising. We also note that water-bridged H-bonds (which were not considered in **Table 5**, as resolving water in crystal structures is resolution dependent) may play an important role in driving affinity. For example, the crystal structure of the HA MEL5 TCR (MEL5 α 24 β 17) contains 9 observed bridged water H-bonds between the TCR and pHLA, which cannot be compared to the WT MEL5 TCR, as its X-ray resolution is not sufficient (3 Å) to place water molecules.²⁰⁶ Comparisons in the change of the buried solvent accessible surface area (BSASA) show an increase upon affinity maturation in the three pairs (MEL5, DMF5 and Tax A6) of WT and single HA variants (**Table 5**). This observation is consistent with an increase in the number of vdWs contacts formed for each of the HA TCRs (**Table 5**), which suggests that a larger amount of interfacial water is expelled upon binding. In the case of the NY-ESO TCRs, no clear trend was established between the change in BSASA and affinity.

Contact analysis of molecular dynamics simulations. Our X-ray analysis showed no obvious structural differences between WT and their counterpart HA TCRs, perhaps because one cannot estimate the strength/occupancy of H-bonds/vdWs from single structures and because bridged water interactions cannot be fairly compared. To overcome these limitations, we turned to molecular dynamics (MD) simulations to obtain insight into how the number of contacts between the TCR and pHLA change over time and consider the role (if any) of molecular flexibility in driving affinity. In order to do this, we performed ten 100 ns long MD simulations of all TCRs in both their apo and pHLA bound forms (totalling 22 μ s of MD simulation). The use of many independent replicas (such as the 10 performed here) is important for obtaining reliable and reproducible results.¹⁴⁷

From our MD simulations with pHLA bound, we calculated the average number of H-bonds (including water bridged H-bonds) and vdWs contacts formed between the TCR and pHLA (**Figure 33A+B**). For MEL5 and DMF5 there was a small increase in both the average number of H-bonds and vdWs contacts between the TCR and pHLA for the higher affinity complexes, contrasting with the NY-ESO and Tax A6 TCRs which showed no clear relationship between affinity and number of contacts. To measure the extent to which contacts to individual pHLA residues were preserved upon affinity maturation, we calculated the total number of H-bonds and vdWs contacts formed between the TCR and each pHLA residue. Of the 10 pHLA residues most contacted by the WT NY-ESO TCR, 8 or 9 are also preserved in the top 10 for all NY-ESO variants (in terms of both the average number of H-bonds and vdWs contacts, see **Tables S4+S5**). For the Tax A6 TCRs, the affinity matured variant preserved between 9-10 of the top 10 WT contacts (**Tables S6+S7**), whilst for the DMF5 and MEL5 TCRs, between 7-9 of the top 10 WT contacts were preserved in the affinity matured variants (**Tables S6+S7**). These results demonstrate that the affinity matured TCRs studied here mostly preserve the contacts formed between the WT TCR and the pHLA.

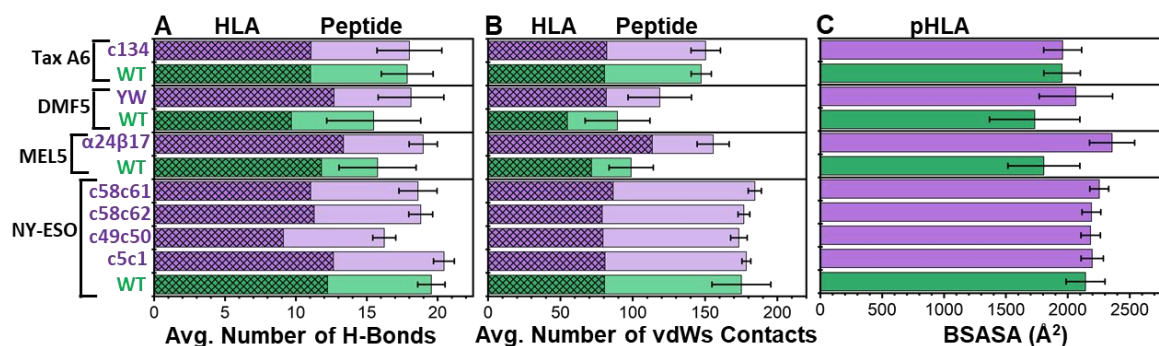


Figure 33: Differences in the average number of H-bonds, vdWs contacts and the BSASA from our MD simulations. All WT TCRs are coloured in green and all affinity matured TCRs are coloured in purple. Bars for the H-bonds and vdWs contacts are divided in two based on HLA-TCR interactions (darker colour and hashed bars) and peptide-TCR interactions (lighter colour). The totals obtained are from 10 independent 100 ns long MD simulations of each TCR-pHLA complex (using the last 90 ns of each simulation). Error bars plotted for **A** and **B** are the standard deviation of the averages from the 10 replicas. Error bars plotted for **C** are the standard deviation obtained from combining all snapshots from all replicas together.

We also applied our MD simulations to determine how the average BSASA differed for all TCR-pHLA complexes (**Figure 33C**). Further, the variance of the BSASA can be used to provide insight into how the overall rigidity of the binding interface differs for each TCR-pHLA complex (a smaller variance would suggest a less flexible binding interface). Whilst the DMF5 and MEL5 TCRs show an increase in the BSASA for the affinity matured variant (consistent with the crystal structure analysis, **Figure 32**), the Tax A6 TCRs show essentially no change in BSASA for the affinity matured TCRs (inconsistent with the crystal structure analysis). For the NY-ESO TCRs, a small general increase in the BSASA was observed for the affinity matured variants, however these results were all within error of one another. Interestingly, all NY-ESO TCRs showed a reduction in the variance of the BSASA, suggesting the binding interface has become overall more rigid (standard deviations of 157 \AA^2 for the WT as compared to between $75\text{-}90 \text{ \AA}^2$ for all HA TCRs). A reduction in variance for the higher affinity variant was also observed for the MEL5 TCRs (from 291 to 182 \AA^2) and to a lesser extent the DMF5 TCRs (from 365 to 297 \AA^2), whilst the Tax A6 TCRs show essentially no difference (from 147 to 154 \AA^2).

Reductions in flexibility are not required for affinity maturation. Changes in rigidity/flexibility of the unbound (apo) TCRs could occur over the course of affinity maturation (due to multiple mutations). Rigidification of the CDR loops (and/or the scaffold as a whole) could act as a mechanism by which affinity is enhanced (by reducing the entropic penalty associated with rigidification upon binding). We therefore made use of our MD simulations of all TCRs in their apo and pHLA bound states to calculate the

changes in root mean square fluctuation (Δ RMSF) upon affinity maturation (**Figure 34**). Further, we evaluated the significance of the Δ RMSF differences observed by performing a two-sample *t*-test ($p < 0.05$). Our results demonstrate the importance of using many replicas and performing subsequent statistical tests to determine if the observed differences in Δ RMSF were in fact significant or not.

For the NY-ESO TCRs, we observed a large decrease in the flexibility of the CDR3 α loop for the three variants that contain mutations in this loop (c5c1, c58c61 and c58c62, see **Figure 34A+B**). This increase in rigidity can be rationalised by the substitution of a glycine residue for a more conformationally restricted amino acid (G97D). Furthermore, the carboxyl side-chain of this mutated residue was able to form an interloop hydrogen bond with T99 (S99 in the WT TCR), which would further rigidify the loop. For all three other mutated CDR loops in the NY-ESO TCRs (CDR2 α , CDR2 β and CDR3 β), no significant changes are observed between the apo TCRs. An increase in the flexibility of the HV α loop is observed for TCR variants c49c50, c58c62 and c58c61, which is likely induced by mutations made in the CDR2 α loop given the close proximity between the two loops (see **Figure 32**) and that fact that c5c1 does not contain CDR2 α mutations. Comparison of the crystal structures of the CDR2 α mutated and WT loops shows the mutated loops to be translated further away from the HV α loop in order to form additional contacts with the pHLA, meaning the flexibility of the HV α loop is likely increased due to a reduction in the number/strength of contacts formed to the CDR2 α loop.

Whilst there was a clear example (CDR3 α loop) of affinity maturation increasing the rigidity of the apo TCR, it is important to consider that a higher affinity TCR is likely to bind the pHLA with increased rigidity relative to the WT. This would mean the beneficial increases in rigidity seen for the apo simulations could therefore be offset by a more rigid TCR when bound to the pHLA. We therefore calculated the Δ RMSF values for pHLA bound simulations of the NY-ESO TCRs (**Figure 34C+D**). Comparison of the apo and pHLA bound Δ RMSFs show that for the CDR3 α loop, the increased rigidity in the apo TCR was largely compensated for by increased rigidity when bound to pHLA, suggesting no beneficial entropy gain for the higher affinity TCR variants with CDR3 α mutations. Whilst mutations in the CDR2 α and CDR3 β loops appear to have an insignificant impact on the flexibility of either the apo or pHLA bound states, the CDR2 β and HV β loops are more rigid in the c5c1 and c58c61 variants pHLA bound form, which share the same set of CDR2 β loop mutations and a single HV β loop mutation (T69I). This data would therefore suggest these mutations to have an entropically unfavourable impact on the overall affinity (which is likely offset by an improved enthalpic contribution, discussed later).

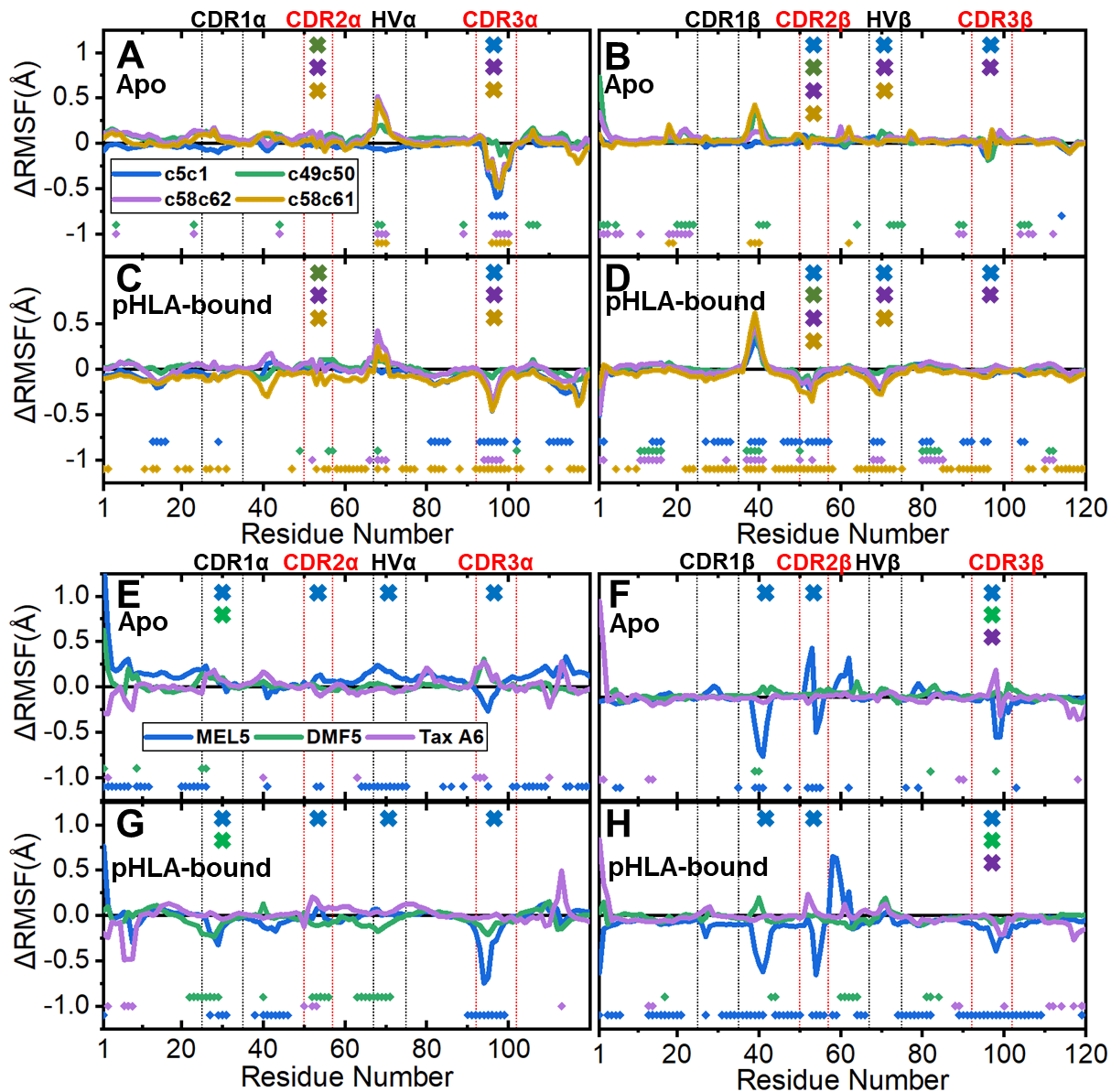


Figure 34: Differences in flexibility between the affinity matured TCRs variable regions and their counterpart WT TCRs. (A–D) Δ RMSF values (HA variant RMSF – WT RMSF) for all NY-ESO TCRs, with panels A and B corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels C and D corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. (E–H) Δ RMSF values (HA variant RMSF – WT RMSF) for the three pairs of HA and WT TCRs, with panels E and F corresponding to the CDR α and CDR β of the Apo TCRs respectively, and panels G and H corresponding to CDR α and CDR β of the TCRs in complex with pHLA respectively. A more negative Δ RMSF value indicates increased rigidity for the HA variant relative to the WT. The points towards the bottom of each graph indicate residues with significantly different Δ RMSF values as determined by a two-sample *t*-test ($p < 0.05$). Crosses are used to indicate the locations of mutated regions of the TCRs (using the same colouring as line graph). Complete RMSF plots for all TCRs simulated are provided in **Figures S1-S4**.

The affinity matured MEL5 TCR contains a total of 17 mutations from the WT spread over seven loops and a further two mutations located in the remainder of the variable domain scaffold. Comparison of the apo Δ RMSFs (**Figure 34E–F**) indicated only the HV α and CDR2 β loops at the TCR binding interface to change significantly upon affinity maturation. The large (both positive and negative) scale changes seen in the CDR2 β loop is likely the result of the large scale rearrangement of this loop upon affinity maturation²⁰⁶. The increased rigidity observed in the residues between the CDR1 β and CDR2 β loops in both the apo and pHLA bound simulations can be rationalised by the L44Q mutation, which is located ~ 15 Å from the closest pHLA residue. Whilst this mutation may well have had a beneficial effect on the pHLA binding affinity, it is possible this mutation (and others) were selected because they instead stabilised the TCR such that it could be better expressed on phage particles during affinity selection. This could ultimately lead to mutations being taken forward that do not impact the affinity of the TCR-pHLA complex. The pHLA bound simulations of the MEL5 TCRs show several regions of increased rigidity for the higher affinity variant, including the CDR1 α , CDR3 α , and CDR3 β loops. This observation of a generally more rigid binding interface for the higher affinity TCR is consistent with the previously discussed reduction in the variance of the BSASA for the higher affinity MEL5 TCR.

Prior thermodynamic analysis on both the WT and HA MEL5 TCR suggested an improvement in the entropy term of the binding free energy upon affinity maturation (from a $T\Delta S^\circ$ of ~ 8.3 kcal mol⁻¹ to ~ 18.1 kcal mol⁻¹).²⁰⁶ The Δ RMSF data discussed above suggest this favourable effect is likely not (at least primarily) driven through changes in rigidity, and is would instead be the result of an improved of entropy of solvation term (which would be consistent with an large increase in the BSASA for the higher affinity MEL5 TCR-pHLA complex, see **Figure 33C**).

For the pair of Tax A6 TCRs (which differ by four point mutations in the CDR3 β loop), we observed little significant changes in flexibility for both the apo and pHLA bound simulations. This lack of difference (in the pHLA bound simulations) is consistent with the BSASA calculations described above and would therefore suggest changes in flexibility are largely insignificant in affecting binding affinity. The WT and affinity matured DMF5 TCRs only differ by two point mutations (D26Y and L98W on the CDR1 α and CDR3 β loops respectively). Comparison of the WT and affinity matured apo TCR simulations (**Figure 34E–F**) showed virtually no differences in flexibility between the WT and the HA variant. In the pHLA bound simulations (**Figure 34G–H**), no notable effect can be observed for the CDR3 β mutation. In contrast however, the CDR1 α loop had increased rigidity in the HA variant. This increased rigidity in the CDR1 α loop is likely also responsible for the observed increased rigidity seen in the neighbouring CDR2 α and HV α loops.

From the four examples discussed, it is clear that changes in flexibility/rigidity during TCR maturation can play a role in increasing or decreasing affinity in some cases (NY-ESO TCR, MEL5 TCR and to a lesser extent DMF5), but not all (Tax A6). Entropy-enthalpy compensation should also be considered: increased rigidity of a TCR in complex with pHLA (an entropic effect disfavouring binding, which may or may not be fully compensated for by increased rigidity in the apo TCR) can aid the formation of favourable contacts between the TCR and pHLA (an enthalpic effect favouring binding).

Energetic hot spots are preserved over the course of affinity maturation. To study how the CDR loop mutations have enhanced the affinity between the TCR and pHLA, we performed binding free energy calculations using the Molecular Mechanics Generalized Born Surface Area (MMGBSA) method.³⁹ This approach (described in more detail in the Methods) uses MD simulations to sample conformations of the complex, receptor and ligand and subject these snapshots to an empirical calculation in order to determine the binding free energy (ΔG_{bind}).

Comparison of our calculated $\Delta\Delta G_{\text{bind}}$ values with the experimentally determined results showed that the differences in affinity between the WT and HA TCRs were identified correctly. In the case of the set of five NY-ESO TCRs, we obtained an R^2 value of 0.78 between simulation and experiment (**Figure 35**). In the cases of the three pairs of WT and affinity matured TCRs, the increase in affinity was correctly predicted for each pair (**Figure 35**). These results indicate that our simulations are able to identify the (atomic level) differences between the WT and HA TCR-pHLA complexes that result in increased affinity.

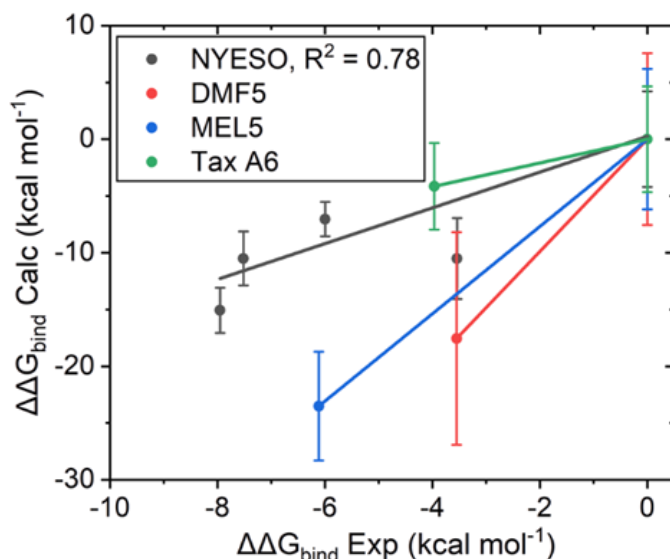


Figure 35: Experimental vs computational $\Delta\Delta G_{\text{bind}}$ values obtained from our MMGBSA calculations for all TCR-pHLA systems studied. For the three pairs of TCRs with a single WT and affinity matured TCR, lines are drawn to guide the eye. Error bars plotted are the standard deviation obtained from the 25 replicas performed per complex (see Methods).

Decomposing the calculated binding energies onto the per residue level can indicate which interactions are the main drivers for the increased binding affinity in the HA TCR variants. Whilst our primary focus was on the differences between WT and HA TCR variants (i.e. $\Delta\Delta G_{\text{bind}}$), we showed that WT energetic hotspots across the TCR-pHLA interface are largely conserved upon affinity maturation (**Figure S5**), in line with the contacts analysis above.

Mutations can directly or indirectly drive affinity enhancement. Overall analysis of the per-residue decomposition indicated that the NY-ESO TCR loop mutations had a largely additive effect on TCR-pHLA binding, i.e. the contribution of mutations in one loop was not affected by mutations in other loops. This additivity could be expected, because the overall TCR conformation was well preserved across all the NY-ESO TCRs studied. For example, the X-ray C α RMSD (vs. the WT TCR) of the TCR variable domains for all affinity matured variants is $< 0.4 \text{ \AA}$.

The two different CDR2 α loop mutations (see **Figure 36A–C**) both appear to have enhanced affinity via the same mechanism: large hydrophobic (and aromatic) groups were introduced where they can be effectively buried at the interface (either S53W or S52F + S53W). The G97D mutation in the CDR3 α loop was predicted to be unfavourable for all three cases in which it occurred (**Figure 36A+D**). This is not surprising as the mutation results in the partial burial of a negatively charged residue upon binding (which will incur a large desolvation penalty). However, the mutation had two clear benefits. The first was the formation of an internal hydrogen bond within the CDR3 α loop to T99 (S99 in the WT TCR), which helped to rigidify the Apo HA CDR3 α loops relative to the WT (see above and **Figure 3A**). Second, the mutation occurred at the same time as the S96L and S99T mutations, which more than compensate for the negative effect introduced by G97D (likely partly because the loop is more rigid, so the other residues can form strong and persistent interactions with the pHLA). This raises an important point: whilst one mutation (in isolation) can have a negative effect on binding affinity, this can be offset by inducing improvements in the contribution to binding affinity of other residues nearby. Further analysis of the impact of CDR3 α mutations suggested enhanced interactions with HLA residue R65 and CDR2 β residue D54. This improvement may again be in part due to the decreased mobility of the loop, which allows for an increase in the total average number of hydrogens bonds formed between the TCR and R65 (WT and non CDR3 α mutated c49c50 combined H-bond occupancies of 2.3 and c5c1, c58c62 and c58c61 occupancies of 2.9-3.2, **Table S5**).

For the CDR2 β mutations, the substitution of the methyl side chain of A50 for a larger hydrophobic side chain (A50V or A50I) was primarily responsible for the increased binding affinity. Furthermore, the G49A mutation (seen only in c58c61) further increased the favourability towards binding, which was in contrast to the more polar G49S mutation in c49c50 and c58c62 (**Figure 36A+E+F**). These results would suggest affinity enhancement is driven in the same way as described earlier for the CDR2 α mutations (by the burial of hydrophobic side chains).

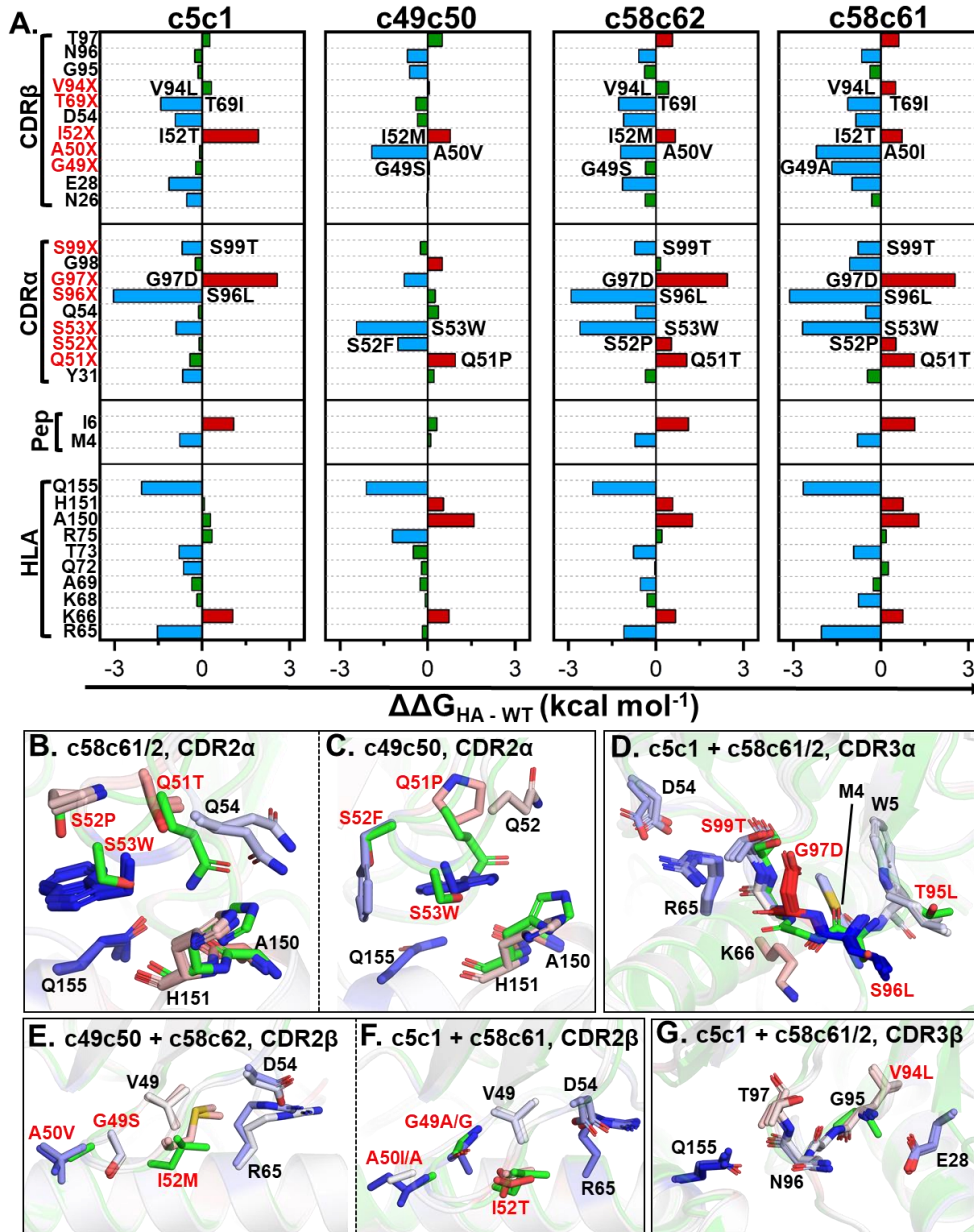


Figure 36: NY-ESO variants show largely additive energetic effects upon affinity maturation. **A.** Per-residue ΔG differences between the high affinity (HA) variants and WT-NY-ESO (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR are coloured blue if < -0.5 kcal mol⁻¹ (favourable for binding) and red if > 0.5 kcal mol⁻¹ (unfavourable for binding), with values in between coloured green. **(B–G)** Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-NY-ESO structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant. Figures are divided to focus on the different regions of the TCR subjected to affinity maturation (CDR2 α , CDR3 α , CDR2 β and CDR3 β), and subdivided when mutations are not consistent between HA TCRs. (c58c61/2 means both c58c61 and c58c62 are shown).

The above observations are in contrast with the CDR3 β mutations that appeared to have increased affinity largely through indirect effects, by increasing the favourability towards binding of the CDR1 β loop residue E28. This improvement in E28 (seen only in TCRs with CDR3 β loop mutations) is likely the result of increased preorganisation of E28 for binding through an increased strength hydrogen bond between the side chain carboxyl group and the backbone of residue V/L94 (average occupancy in WT and c49c50 simulations is between 0.50-0.56, compared to c5c1, c58c62 and c58c61 where it is between 0.83-0.86).

HLA residue Q155 showed an improvement in affinity for all NY-ESO affinity matured TCRs. This includes c5c1 and c49c50 that did not contain any CDR2 α or CDR3 β mutations, respectively, and are the loops responsible for engaging Q155 (see **Figure 36B+C+G**), suggesting mutations in either loop (CDR2 α or CDR3 β) can improve the favourability of Q155 towards binding.

Differences between the WT and HA Tax A6 TCRs all occurred on the CDR3 β loop (A99M, G100S, G101A and R102Q), and prior structural analysis of the differences between the WT and HA complexes suggested the increased affinity to be due to an increased number of contacts between the TCR and pHLA.²⁰⁴ Mutations A99M and G100S were primarily responsible for the enhanced affinity (**Figure 37A+B**). In line with the prior structural characterisation, we see an increase in the total average number of contacts made between the TCR and the pHLA residues A149 and A150, which sit below the CDR3 β loop residues A99M and G100S. Specifically, the total average number of H-bonds formed to both A149 and A150 doubled from 0.4 to 0.8 for both residues (**Table S6**), alongside an increase in the total average number of vdWs contacts formed (**Table S7**).

In the case of MEL5, a total of 19 mutations on 7 loops gave rise to the approximate 3×10^5 fold increase in affinity, which was found to be primarily entropically driven.²⁰⁶ It is therefore unsurprising to see that alongside the reduced mobility of the affinity matured TCR (see **Figure 34**), several of the most favourable of mutations (CDR α D27F and CDR β G52P and I53F) increased the total amount of buried hydrophobic matter at the binding interface (**Figure 37A+E-G**). This would likely be an entropically favourable process due to the expulsion of ordered water molecules that surround these hydrophobic or aromatic groups upon binding. The only mutation which showed a large negative effect on affinity was V93D CDR α , which can be rationalised in the same manner as seen for NY-ESO CDR3 α G97D (**Figure 36A+D**). That is, D93 (V93 in the WT) forms an interloop hydrogen bond with K96R, which would help to rigidify the loop (as seen in **Figure 37E+G**) and therefore reduce the entropic penalty associated with binding.

Interestingly, of the 19 mutations present in MEL5 $\alpha 24\beta 17$, only 9 showed substantial energetic differences. Of the remaining 10 mutations, two positions (CRD1 α R28L and CDR3 β T100M) made direct and favourable interactions with the pHLA but were of similar strength in both the WT and HA TCR. The other eight mutations did not make

direct contact with the pHLA and are instead likely to be involved in regulating the flexibility, stability and/or conformational sampling of the TCR.

Two point mutations (D26Y on CDR1 α and L98W on CDR3 β) on DMF5 gave rise to an approximate 400-fold enhancement in affinity.⁴⁷ Analysis of the effect of the D26Y mutation (**Figure 37A+C**) suggested that not only is the mutation itself directly favourable, but it also enhances the contribution to affinity of K66 on HLA, through the formation of a hydrogen bond between the two residues. The reduced favourability towards binding observed for the other nearby charged residues (peptide E1 and HLA E58) was likely a desolvation effect induced by the burial of the tyrosine side chain (which would displace some of the solvent around these charged residues). In the case of L98W (**Figure 37A+D**), the burial of a large hydrophobic residue unsurprisingly leads to increased contribution to the binding affinity.

The affinity matured TCRs discussed here provide examples of mutations whose effects are largely localised around the mutation site, as well as mutations that cause significant effects on the contributions to binding affinity of residues distal from the mutation site. In some cases, such non-local effects can be induced due to a change in the long-range electrostatics (in particular when a charged residues is introduced/removed). Alternatively, if the motions of different CDR loops are strongly correlated to one another, one may see a change in the conformational sampling of both loops upon mutation of either loop. A prior study on the degree of coupling between the CDR loops in WT Tax A6 and DMF5 found Tax A6 CDR loops to be largely uncoupled, whereas the dynamics of DMF5 loops were strongly correlated to one another.¹²³ This would help to rationalise the observation that energetic changes in Tax A6 are largely localised to the mutation sites, in contrast with DMF5 mutations (**Figure 37B-D**).

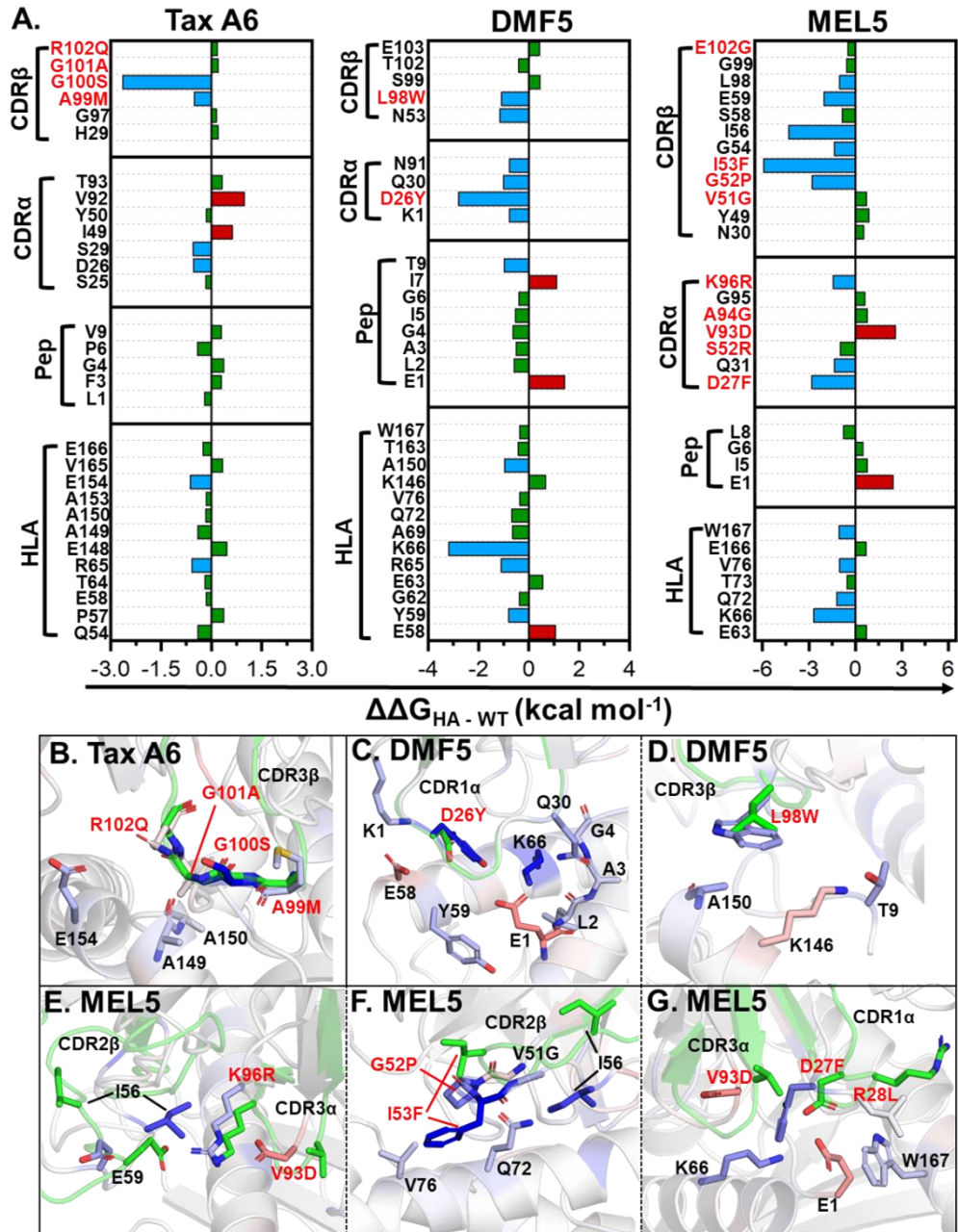


Figure 37: Changes in Energetics at the TCR-pHLA Interface upon affinity maturation. (A) Per-residue $\Delta\Delta G$ differences between the three high affinity (HA) variants and their counterpart WT TCRs (i.e. $\Delta\Delta G$), with positions mutated indicated throughout in red. $\Delta\Delta G$ differences between the WT and the given HA TCR pair are coloured blue if < -0.5 kcal mol⁻¹ (favourable for HA) and red if > 0.5 kcal mol⁻¹ (unfavourable for HA), with all values in-between coloured green. (B–G) Colour mapping of the above per residue $\Delta\Delta G$ values onto all carbon atoms of the high affinity variants (with the WT-TCR structure shown in green for reference). Colour mapping is performed from blue to white to red with blue indicating a favourable change and red indicating an unfavourable change for the affinity matured variant respectively. Figures are divided up to show the regions which show the major changes upon affinity maturation.

Conclusions

Here, we used structural analysis, MD simulations and free energy calculations to extensively characterise four case studies of TCR affinity maturation. Our data demonstrates the limitations of contact analysis only (both on X-ray structures and from MD simulations) for rationalising changes in affinity for TCRs. In some cases (NY-ESO and MEL5), rigidification of the apo TCRs occurred upon affinity maturation, which led to an entropic advantage for binding affinity. However, this entropic advantage was by and large negated for by the HA TCRs binding the pHLA tighter than their counterpart WT TCRs do (entropically unfavoured). However, this effect was not significant in the other two cases. These results are in agreement with a large scale study on the relationship between antibody affinity maturation and rigidity, where in some cases rigidification was observed to help drive affinity enhancement, however it was by no means a global trend nor a requirement.²⁴⁷ Our free energy calculations reproduce the experimental affinity relationships for all four cases studies and demonstrate the energetic footprint was by and large preserved upon affinity maturation. This ability to accurately predict affinity hierarchies (which we are currently investigating in more detail) has promise in rational design efforts. We have demonstrated the use of atomistic simulation to characterise the energetic footprint of different HA TCRs, allowing comparison with their thymically selected (WT) counterparts. This could be used to “filter” potential therapeutic candidates prior to extensive experimental specificity/safety validation. Our analysis indicates that the introduction and burial of large aromatic or hydrophobic side chains is a common occurrence in affinity maturation. This observation ties in with the fact that the most commonly identified residues in protein-protein binding sites are Trp, Met and Phe.²⁴⁸ However, relying too much on hydrophobic interactions (vs. more directional and thus specific electrostatic interactions) to drive affinity enhancements could have negative effects on the specificity of the TCR.

Methods

Molecular Dynamics Simulations. Starting structures for all simulations were obtained from previously solved X-ray crystal structures from multiple studies (see **Table S2+S3** for a complete list of structures used).^{43–45,47,105,204–206,245,249} Molecular dynamics (MD) simulations of apo TCR structures were initiated from either the apo X-ray structure coordinates (if available) or from the TCR-pHLA bound structure if not. MolProbity⁷⁵ was used to determine the optimum tautomerisation states for all histidine residues (tautomerisation states used for all simulations are provided in **Table S2+S3**) and make any required Asn/Gln side chain flips (under the criteria of optimising the hydrogen bonding network). Protonation states of all titratable residues were assigned using PropKa 3.0¹⁵⁸ for pH 7, leading to all residues treated in their standard protonation states. The results of these calculations were visually inspected, and care was taken throughout to

ensure consistency in tautomerisation and Asn/Gln flips where appropriate (i.e. same settings for the pHLA with different TCR variants). All structures were then solvated in an octahedral water box (retaining any crystal waters), ensuring all protein atoms were at least 10 Å away from the box boundary, with Na⁺ or Cl⁻ counter ions added as necessary to ensure an overall neutral charge. MD simulations were performed using GPU accelerated Amber16²²², with the ff14SB force field⁶⁰ and TIP3P water model used to describe protein and water molecules respectively. All systems investigated were equilibrated to 300 K and 1 atm in the NPT ensemble (protocol described in the Supporting Information). Subsequently, production MD simulations were performed for 100 ns each with 10 replicas performed per TCR-pHLA or apo-TCR (replicas assigned different random velocity vectors during equilibration procedure, see SI Methods). Production MD simulations were run using a 2 fs time step with the SHAKE algorithm applied to any bond containing a hydrogen atom. An 8 Å direct space non-bonded cut-off was applied with long range electrostatics evaluated using the particle mesh Ewald⁷² algorithm. Temperature was regulated using Langevin temperature control (collision frequency of 1 ps⁻¹), whilst pressure was controlled with a Berendsen barostat (setting the pressure relaxation time to 1 ps).

MD Trajectory and Crystal Structure Analysis. Trajectory and X-ray structure analysis was primarily performed with CPPTRAJ¹⁶⁰, using frames collected every 10 ps for analysis of MD simulations. H-bonds (including water bridged H-bonds) between atoms of the TCR and pHLA were defined as formed if the donor acceptor distance was within 3.5 Å and the donor hydrogen acceptor angle was 180 ± 45° (crystal structure hydrogen atom positions were energy minimised prior to hydrogen bond analysis). If two heavy (non-hydrogen) atoms were within 4 Å of one another, a vdWs contact was considered to be formed between the two atoms. The total number of H-bonds and contacts formed between the TCR and a particular pHLA residue were then summed (see **Tables S4–S7**). RMSF and RMSD calculations were performed on the Cα atoms of the relevant residues with RMS fitting performed to a consistent set of TCR residues in the variable domains that are not highly flexible (N-terminal residues and all CDR loop residues were omitted, see SI Methods for further details), therefore ensuring a fair comparison between apo and pHLA bound simulations. RMSFs, H-bonds and vdWs contacts were calculated by discarding the first 10 ns of simulation time (meaning 10 replicas of 10-100 ns used) to allow for structure equilibration. Buried solvent accessible surface area (BSASA) was determined using the LCPO algorithm²⁵⁰, available with the “molsurf” command within CPPTRAJ¹⁶⁰. The six TRangle terms and the TCR-pHLA docking angle was determined using the STCRDab webserver.²⁵¹ The TRangle terms were then used to calculate the d_{TRangle} as shown in Equation (14).²⁵¹

$$d_{\text{TRangle}} = \sqrt{\sum (\theta_{i,WT} - \theta_{i,HA})^2} \quad (14)$$

Where $\theta_{i,WT}$ is the i^{th} TRangle parameter for the WT TCR and $\theta_{i,HA}$ is the i^{th} TRangle parameter for HA TCR.

MMGBSA Methodology. The molecular mechanics generalized Born surface area (MMGBSA) method is a binding free energy calculation method which has been widely used to predict relative binding free energies.^{85,252} The approach uses a combination of MD simulations (for sampling to obtain many snapshots) and empirical calculations (on the obtained snapshots) to predict ΔG_{bind} . It is important to note that these calculations should not be relied upon to provide accurate absolute binding free energies (i.e. ΔG_{bind}), instead they can be used to provide relative binding free energies between similar ligands (i.e. $\Delta\Delta G_{\text{bind}}$) and identify residues that contribute (favourably or unfavourably) towards binding.⁸⁵ In the MMGBSA approach, the different contributions to affinity are calculated individually and summed together to obtain ΔG_{bind} (see Equation (Eq 7)).

$$\Delta G_{\text{bind}} = \Delta E_{\text{el}} + \Delta E_{\text{vdw}} + \Delta G_{\text{pol}} + \Delta G_{\text{npol}} - T\Delta S \quad (7)$$

Where ΔE_{el} and ΔE_{vdw} are obtained directly from the molecular mechanics force field terms and describe the gas phase interaction energy. The polar and non-polar contributions to the solvation free energy are described by ΔG_{pol} and ΔG_{npol} respectively. ΔG_{pol} is calculated by solving the GB equation, whilst ΔG_{npol} is obtained from a function that assumes a linear relationship between the solvent accessible surface area and ΔG_{npol} . Finally, $T\Delta S$ describes the change in entropy of the solute upon binding, most often calculated through normal mode analysis (NMA). NMA is computationally expensive (for large systems like TCR-pHLA) and also tends to produce large errors that do not improve the accuracy of the calculation.²⁵³ Furthermore, as it is not possible to decompose the results from NMA to a per-residue level, we did not perform NMA for our MMGBSA calculations. As shown in **Figure 35**, we were able to obtain good agreement with experiment without this term.

MMGBSA Procedure. MMGBSA calculations were performed using MMPBSA.py.MPI³⁹, using 25 independent (random velocity vectors assigned upon heating) 4 ns long MD simulations (separate to the above described 100 ns long simulations), starting from the X-ray structure, as this approach has previously been shown to provide converged and accurate relative binding free energies for both small molecule drugs⁹¹ and pHLA binding⁹². These simulations were run under the same conditions as the aforementioned longer timescale simulations (and are described in full in the SI). MD Simulations were performed on the complex structure only (often referred to as the single trajectory approach), with these simulations used to obtain snapshots of the free receptor and ligand. The single trajectory approximation is used throughout the literature, as it tends to significantly reduce the energetic noise and improve the predictability of the obtained results.⁸⁴ From each replica, 300 equally spaced snapshots were taken from the last 3 ns of each MD simulation for MMGBSA calculations, giving a total of 7500 frames per complex. MMGBSA calculations used the GB-Neck2 (i.e. igb

= 8) solvation model and an implicit salt concentration of 150 mM. The obtained results were decomposed into their per-residue contributions to the total free energy, with the values obtained used to calculate the differences between the wild type (WT) and HA variants (as plotted in **Figure 36+Figure 37**).

Author Contributions

RMC performed structural analysis, simulations and analysed data. RMC wrote the manuscript. All authors conceived and directed the project. All authors critiqued the manuscript.

Acknowledgments

R.M.C.'s studentship is funded by the EPSRC. M.W.v.d.K. 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.

Conflict of interest

D.K.C. is an employee of Immunocore LTD. The authors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest.

Supporting information for: Drivers for high affinity TCR-pHLA interactions: insights from structure and dynamics

Rory M. Crean,^{1,2} Christopher R. Pudney,^{*,1,3} David K. Cole,^{*,4} and Marc W. van der Kamp.^{*,5}

¹Department of Biology and Biochemistry, ²Doctoral Training Centre in Sustainable Chemical Technologies, ³Centre for Therapeutic Innovation, University of Bath, Bath, BA2 7AY, UK.

⁴Immunocore Ltd., Abingdon OX14 4RY, UK

⁵School of Biochemistry, University of Bristol, Biomedical Sciences Building, University Walk, Bristol, BS8 1TD, UK.

*Correspondence: Dr Christopher R. Pudney, E-mail: c.r.pudney@bath.ac.uk, Dr David K. Cole, E-mail: david.cole@immunocore.com, and Dr Marc W. van der Kamp Email: marc.vanderkamp@bristol.ac.uk,

Keywords. Cancer immunotherapy, peptide-human leukocyte antigen (pHLA), T-cell receptor (TCR), molecular dynamics (MD).

Supporting Information Methods

MD Equilibration Procedure

The following procedure was used to prepare all systems simulated for production MD simulations at 300 K and 1 atm. Furthermore, the equilibration protocol used is identical for both our “long” timescale (5 x 100 ns) and “short” time scale (25 x 4 ns) MD simulations. All dynamics steps applied the SHAKE algorithm to constrain all bonds containing hydrogen. Replicas simulations were initiated from the second heating step of the following protocol (with each replica therefore assigned different random velocity vectors at this stage).

First hydrogens atoms and solvent molecules were energy minimised (using 500 steps of steepest descent followed by 500 steps of conjugate gradient minimisation). To prevent the movement of non-hydrogen and non-solvent atoms during the minimisation, 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 300 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 restraint on only the C α carbon atoms and subjected to another round of energy minimisation (500 steps of steepest descent followed by 500 steps of conjugate gradient). Retaining these positional restraints, the system was heated from 25 K to 300 K over the course of 50 ps (NVT ensemble, 1 fs time step). Simulations were then performed in the NPT ensemble (1 atm, 300 K, 2 fs time step) by first gradually reducing the 5 kcal mol⁻¹ Å⁻¹ C α 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. The end structure from this run was then used as the starting structure for production MD simulations. 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 again used Langevin temperature control (collision frequency of 1 ps⁻¹) and a Berendsen barostat (1 ps pressure relaxation time).

RMS Fitting Procedure

To ensure fair comparison between the WT and HA TCR structures, RMS fitting (for RMSF calculations) was performed using the same set of residues in the TCR variable region. Residues excluded from the RMS fitting procedure were the first five N-terminal residues and all CDR loop residues (due to their high mobility, which would therefore provide a poor fit). Residues used for RMS fitting from Chain A were therefore: 6-22,33-46,55-65,73-94,103-113. Residues used for RMS fitting from Chain B were: 6-21,30-47,54-65,73-91,102-111. RMS fitting of MD simulation snapshots was first performed to the crystal structure, with this RMS fitted trajectory used to create an average structure. Following this, all snapshots were then re-fitted to the average structure for the subsequent RMSF calculation.

Supporting Information Tables

Table S1: CDR loop mutations for all TCRs under investigation. Wild-type (WT) residues subject to mutations are in bold, whilst mutated residues in the affinity matured variants are underlined and coloured red.

TCR	CDR1 α	CDR2 α	FW α	CDR3 α	CDR1 β	FW β	CDR2 β	CDR3 β
NY-ESO WT c5c1 c49c50 c58c62 c58c61	N/A	Res: 51-53 QSS QSS <u>PFW</u> <u>TPW</u> <u>TPW</u>	N/A	Res: 95-99 TSGGS <u>LLDGT</u> TSGGS <u>LLDGT</u> <u>LLDGT</u>	N/A	N/A	Res: 48-52 VGAGI <u>VGAGT</u> <u>VSVG</u> <u>VSVG</u> <u>VAIQI</u>	Res: 94-96 VGN <u>LGN</u> VGN <u>LGN</u> VGN
Tax A6 WT c134	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Res: 98-102 LAGGR <u>LMSAQ</u>
DMF5 WT YW	Res: 25-27 SDR SYR	N/A	N/A	N/A	N/A	N/A	N/A	Res: 97-99 SLS SWS
MEL5 WT α 24 β 17	Res: 27-30 DRGS <u>FLGS</u>	Res: 50-54 IYSNG <u>TYREG</u>	Res: 69-71 SQY <u>SQH</u>	Res: 93-97 VAGKS <u>DGGRL</u>	N/A	Res: 43-45 GLQ GPQ	Res: 50-53 SVGI <u>WGPF</u>	Res: 99-104 TGELF <u>MGGWO</u>

Table S2. PDB IDs and histidine tautomerisation state assignments for all NY-ESO MD simulations.

TCR-pHLA System ^a	HID Tautomerisation States ^b	HIE Tautomerisation States ^c
NY-ESO - WT PDB: 2BNR ²⁴⁵	HLA: 3, 70, 74, 93, 114, 145, 260. β2m: CDRα: CDRβ: 151.	HLA: 151, 188, 191, 192, 197, 263. β2m: 13, 31, 51, 84. CDRα: 112. CDRβ: 27, 45, 134, 164, 204.
NY-ESO – c5c1 PDB: 2PYE ⁴⁵	SAME AS WT	SAME AS WT
NY-ESO – c49c50 PDB: 2F53 ⁴³	SAME AS WT	SAME AS WT
NY-ESO - c58c62 PDB: 2P5W ⁴⁵	SAME AS WT	SAME AS WT
NY-ESO - c58c61 PDB: 2P5E ⁴⁵	SAME AS WT	SAME AS WT

^a, All Apo-TCR simulations used the same tautomerisation states as those in the TCR-pHLA simulations.

^b, HID corresponds to a histidine residue which is singly protonated on its N δ 1 nitrogen.

^c, HIE corresponds to a histidine residue which is singly protonated on their N ϵ 2 nitrogen.

Table S3. PDB IDs and histidine tautomerisation state assignments for the MD simulations of the Tax A6, DMF5 and MEL5 WT and HA TCRs.

TCR-pHLA System^a	HID Tautomerisation States^b	HIE Tautomerisation States^c
DMF5 – WT PDB: 3QDG ²⁴⁹	HLA: 3, 70, 74, 93, 114, 151, 191, 192, 260. β2m: 51. CDRα: CDRβ: 32, 50, 138, 168.	HLA: 145, 188, 197, 263. β2m: 13, 31, 84. CDRα: CDRβ: 155, 208.
DMF5 – YW PDB: 4L3E ⁴⁷	SAME AS WT	SAME AS WT
Tax A6 – WT PDB: 1AO7 ¹⁰⁵	HLA: 3, 70, 74, 93, 114. β2m: CDRα: CDRβ:	HLA: 145, 151, 188, 191, 192, 260, 263. β2m: 13,31, 51, 84. CDRα: CDRβ: 29, 47, 139, 156, 169, 209.
Tax A6 – c134 PDB: 4FTV ²⁰⁴	SAME AS WT	SAME AS WT
MEL5 – WT PDB: 3HG1 ⁴⁴	HLA: 3, 70, 74, 114. β2m: CDRα: CDRβ: 5, 207.	HLA: 93, 145, 151, 188, 191, 192, 197, 260, 263. β2m: 13, 31, 51, 84. CDRα: 71. CDRβ: 137, 154, 167.
MEL5 – α24β17 PDB: 4JFF ²⁰⁶	SAME AS WT	SAME AS WT

^a, All Apo-TCR simulations used the same tautomerisation states as those in the TCR-pHLA simulations.

^b, HID corresponds to a histidine residue which is singly protonated on its N δ 1 nitrogen.

^c, HIE corresponds to a histidine residue which is singly protonated on their N ϵ 2 nitrogen.

Table S4: Average number of vdWs contacts formed between the TCR to each pHLA residue for all NY-ESO simulations. For each TCR-pHLA complex the 10 pHLA residues with the greatest number of contacts are coloured red, with their rank provided in brackets.

Residue	TCR-pHLA Complex				
	WT	c5c1	c49c50	c58c62	c58c61
HLA: 19	0.0	0.0	0.2	0.3	0.7
HLA: 62	1.4	0.7	1.5	0.5	0.6
HLA: 65	14.6 (4)	19.2 (3)	16.3 (4)	14.3 (4)	19.5 (3)
HLA: 66	4.2	6.0 (9)	2.4	6.1 (9)	5.6 (10)
HLA: 68	2.4	2.1	4.2	3.7	3.7
HLA: 69	2.5	3.7	2.4	4.2	3.5
HLA: 71	0.0	0.0	0.1	0.1	0.0
HLA: 72	11.6 (6)	13.3 (4)	16.9 (3)	15.4 (3)	14.8 (4)
HLA: 73	4.0	5.9 (10)	4.7 (10)	6.0 (10)	6.0 (9)
HLA: 75	1.7	0.5	3.6	1.8	2.0
HLA: 76	0.8	1.9	1.4	1.9	1.9
HLA: 146	0.1	0.0	0.0	0.2	0.1
HLA: 149	0.0	0.0	0.0	0.3	0.2
HLA: 150	7.1 (9)	5.0	4.0	2.4	2.4
HLA: 151	4.9 (10)	5.0	6.5 (8)	5.2	4.8
HLA: 152	0.3	0.0	0.0	0.0	0.0
HLA: 154	2.1	2.1	1.4	4.0	4.9
HLA: 155	16.4 (3)	11.9 (6)	11.5 (6)	9.9 (6)	11.7 (6)
HLA: 158	0.0	0.0	0.7	0.4	0.6
HLA: 163	0.0	1.0	0.0	0.7	1.0
Pep: 4	21.1 (2)	27.8 (2)	21.5 (2)	28.3 (2)	28.0 (2)
Pep: 5	44.4 (1)	42.7 (1)	44.4 (1)	42.4 (1)	42.8 (1)
Pep: 6	9.4 (7)	7.8 (7)	9.5 (7)	7.7 (7)	7.7 (7)
Pep: 7	7.1 (8)	7.4 (8)	6.2 (9)	7.3 (8)	6.7 (8)
Pep: 8	12.5 (5)	13.1 (5)	12.0 (5)	12.7 (5)	12.7 (5)

Table S5: Average number of hydrogen bonds formed between the TCR to each pHLA residue for all NY-ESO simulations. For each TCR-pHLA complex the 10 pHLA residues with the greatest number of contacts are coloured red, with their rank provided in brackets.

Residue	TCR-pHLA Complex				
	WT	c5c1	c49c50	c58c62	c58c61
HLA: 19	0.1	0.3	0.2	0.2	0.4
HLA: 43	0.0	0.0	0.1	0.1	0.0
HLA: 58	0.0	0.1	0.0	0.1	0.1
HLA: 61	0.0	0.0	0.1	0.0	0.1
HLA: 65	2.3 (2)	3.2 (1)	2.3 (2)	2.9 (1)	3.2 (1)
HLA: 66	0.5	0.7	0.4	0.7 (10)	0.7
HLA: 68	0.4	0.6	0.5 (10)	0.4	0.6
HLA: 70	0.0	0.1	0.0	0.0	0.1
HLA: 71	0.0	0.0	0.0	0.1	0.0
HLA: 72	1.7 (4)	1.1 (8)	2.2 (3)	1.0 (7)	0.9 (8)
HLA: 73	0.6	1.6 (5)	0.7 (9)	1.6 (4)	1.6 (5)
HLA: 75	0.4	0.3	0.4	0.3	0.2
HLA: 80	0.0	0.0	0.0	0.1	0.0
HLA: 146	0.0	0.3	0.3	0.4	0.3
HLA: 149	0.1	0.1	0.1	0.2	0.2
HLA: 150	1.5 (6)	1.1 (7)	0.2	0.5	0.6
HLA: 151	1.0 (8)	0.8	0.3	0.7	0.7 (10)
HLA: 154	0.2	0.3	0.0	0.4	0.4
HLA: 155	2.1 (3)	1.8 (4)	1.2 (5)	1.2 (5)	1.6 (4)
HLA: 163	0.1	0.0	0.0	0.0	0.0
Pep: 1	0.0	0.1	0.0	0.0	0.1
Pep: 4	1.1 (7)	1.4 (6)	1.0 (6)	1.2 (6)	1.3 (6)
Pep: 5	1.0 (9)	0.9 (10)	0.9 (8)	0.9 (9)	0.9 (9)
Pep: 6	1.7 (5)	1.8 (3)	1.8 (4)	1.8 (3)	1.8 (3)
Pep: 7	1.0 (10)	1.0 (9)	1.0 (7)	1.0 (8)	1.0 (7)
Pep: 8	2.5 (1)	2.6 (2)	2.5 (1)	2.6 (2)	2.6 (2)

Table S6: Average number of hydrogen bonds formed between the TCR to each pHLA residue for our simulations of WT and HA DMF5, MEL5 and Tax A6. For each TCR-pHLA complex the 10 pHLA residues with the greatest number of contacts are coloured red, with their rank provided in brackets.

Residue	TCR-pHLA Complex					
	DMF5		MEL5		Tax A6	
	WT	YW	WT	$\alpha 24\beta 17$	WT	c134
HLA: 19	0.0	0.0	0.1	0.0	0.0	0.0
HLA: 43	0.0	0.0	0.2	0.0	0.0	0.0
HLA: 55	1.1 (5)	1.0 (5)	0.0	0.0	0.9 (9)	1.1 (8)
HLA: 58	1.0 (6)	0.3	0.1	0.8	0.5	0.5
HLA: 59	0.2	0.0	0.0	0.0	0.0	0.0
HLA: 61	0.3	0.1	0.1	0.3	0.0	0.0
HLA: 62	0.1	0.0	0.0	0.0	0.0	0.0
HLA: 65	1.6 (3)	1.5 (4)	2.1 (1)	2.0 (2)	2.5 (1)	2.6 (1)
HLA: 66	0.9 (9)	0.6 (10)	0.0	0.7	1.0 (7)	1.3 (5)
HLA: 68	0.0	0.1	0.5 (8)	1.0 (9)	0.0	0.0
HLA: 72	2.4 (1)	2.0 (1)	1.5 (2)	2.7 (1)	0.2	0.2
HLA: 73	0.2	0.1	0.6 (7)	0.0	0.0	0.0
HLA: 75	0.5	0.5	0.2	0.0	0.0	0.0
HLA: 145	0.1	0.0	0.0	0.0	0.0	0.0
HLA: 146	0.2	0.1	0.0	0.0	0.0	0.0
HLA: 149	0.2	0.0	0.0	0.0	0.4	0.2
HLA: 150	0.1	0.0	0.0	0.1	0.4	0.8
HLA: 151	0.1	0.0	0.0	0.2	0.4	0.8
HLA: 154	0.0	0.1	0.3	1.3 (6)	0.3	0.2
HLA: 155	0.9 (7)	0.7 (8)	0.5 (9)	0.8 (10)	1.0 (6)	1.5 (4)
HLA: 157	0.0	0.0	0.0	1.1 (8)	0.1	0.1
HLA: 158	0.2	0.2	0.0	0.2	0.3	0.2
HLA: 161	0.1	0.1	0.0	1.1 (7)	0.2	0.3
HLA: 162	0.0	0.0	0.0	0.1	0.0	0.0
HLA: 163	0.5	0.5	0.4 (10)	0.6	0.9 (8)	0.9 (10)
HLA: 166	0.8	0.4	0.9	0.1	1.4 (4)	1.1 (9)
HLA: 167	0.3	0.4	0.0	0.5	0.5	0.5
HLA: 170	0.0	0.1	0.4	0.0	0.1	0.1
Pep: 1	1.8 (2)	1.7 (3)	0.6 (6)	1.5 (4)	0.0	0.0
Pep: 2	0.5	0.7 (9)	0.2	0.4	0.7	0.5
Pep: 3	0.1	0.0	0.0	0.0	0.0	0.0
Pep: 4	1.3 (4)	1.8 (2)	1.2 (3)	1.4 (5)	2.2 (3)	1.9 (3)
Pep: 5	0.1	0.2	0.1	0.0	1.1 (5)	1.1 (7)
Pep: 6	0.0	0.1	0.2	0.1	0.9 (10)	1.2 (6)
Pep: 7	0.8 (10)	0.7 (7)	0.9 (5)	1.6 (3)	2.3 (2)	2.1 (2)
Pep: 9	0.9 (8)	0.8 (6)	0.4	0.2	0.0	0.0

Table S7: Average number of vdWs contacts formed between the TCR to each pHLA residue for our simulations of WT and HA DMF5, MEL5 and Tax A6. For each TCR-pHLA complex the 10 pHLA residues with the greatest number of contacts are coloured red, with their rank provided in brackets.

Residue	TCR-pHLA Complex					
	DMF5		MEL5		Tax A6	
	WT	YW	WT	$\alpha 24\beta 17$	WT	c134
HLA: 43	0.0	0.0	0.4	0.0	0.0	0.0
HLA: 55	3.0	3.5	0.0	0.0	2.7	3.6
HLA: 56	0.0	0.2	0.0	0.0	0.0	0.0
HLA: 58	0.1	2.9	0.3	2.0	0.1	0.5
HLA: 59	1.2	2.3	0.0	0.3	0.4	1.0
HLA: 61	0.0	0.9	0.0	0.0	0.0	0.0
HLA: 62	0.0	3.2	1.6	0.2	0.0	0.0
HLA: 63	0.0	0.3	0.0	0.0	0.0	0.0
HLA: 65	7.1 (4)	8.5 (3)	13.9 (1)	19.9 (2)	17.1 (3)	17.1 (2)
HLA: 66	4.1 (9)	8.3 (4)	1.8	4.0	8.3 (6)	9.1 (6)
HLA: 68	0.0	0.0	2.2	1.6	2.7	2.8
HLA: 69	5.6 (7)	7.1 (5)	3.6 (9)	3.1	5.7 (8)	5.7 (9)
HLA: 70	0.1	0.5	2.3	1.4	0.0	0.0
HLA: 72	10.4 (1)	12.2 (1)	8.0 (4)	25.4 (1)	2.2	2.5
HLA: 73	1.3	1.5	2.9	1.7	2.2	2.1
HLA: 75	1.3	1.7	0.9	6.6 (8)	0.0	0.0
HLA: 76	1.7	2.1	1.3	4.2 (10)	0.0	0.0
HLA: 145	0.0	0.1	0.0	0.0	0.0	0.0
HLA: 146	0.0	0.8	0.0	0.0	0.0	0.0
HLA: 149	0.0	0.7	0.0	0.0	0.8	0.2
HLA: 150	0.8	4.8 (9)	0.0	0.1	5.1 (9)	8.4 (8)
HLA: 151	0.2	0.0	1.0	0.7	3.5	4.6 (10)
HLA: 152	0.0	0.0	0.0	0.0	0.5	1.2
HLA: 154	0.1	0.6	4.2 (8)	6.3 (9)	1.2	1.9
HLA: 155	7.6 (3)	5.8 (8)	9.2 (2)	10.4 (3)	13.3 (4)	12.9 (4)
HLA: 157	0.0	0.0	0.0	8.0 (6)	0.5	0.1
HLA: 158	2.1	1.8	1.4	2.2	2.3	1.6
HLA: 159	1.9	1.6	0.7	1.4	0.9	0.8
HLA: 161	0.0	0.0	0.0	2.2	0.0	0.1
HLA: 162	0.0	0.0	0.0	0.3	0.0	0.1
HLA: 163	1.7	2.0	3.5 (10)	2.8	2.4	2.6
HLA: 166	1.0	1.8	4.3 (7)	0.4	4.2 (10)	3.6
HLA: 167	1.4	2.7	0.3	4.2	2.2	2.0
HLA: 170	0.7	1.7	1.5	0.7	2.1	1.0
Pep: 1	6.5 (5)	6.1 (6)	0.8	4.2	2.8	3.2
Pep: 2	2.3	2.0	0.4	1.2	1.6	1.2
Pep: 3	1.7	1.7	1.2	2.1	0.2	0.0
Pep: 4	8.4 (2)	9.1 (2)	5.5 (5)	8.7 (5)	9.5 (5)	8.9 (7)
Pep: 5	5.9 (6)	6.1 (7)	8.1 (3)	6.9 (7)	25.5 (1)	26.8 (1)

Pep: 6	1.9	0.1	2.3	2.9	3.0	2.7
Pep: 7	4.2 (8)	4.3 (10)	5.1 (6)	9.3 (4)	7.4 (7)	9.3 (5)
Pep: 8	2.7	0.0	1.8	3.0	17.2 (2)	17.1 (3)
Pep: 9	4.0 (10)	0.0	0.2	2.1	0.0	0.0

Supporting Information Figures

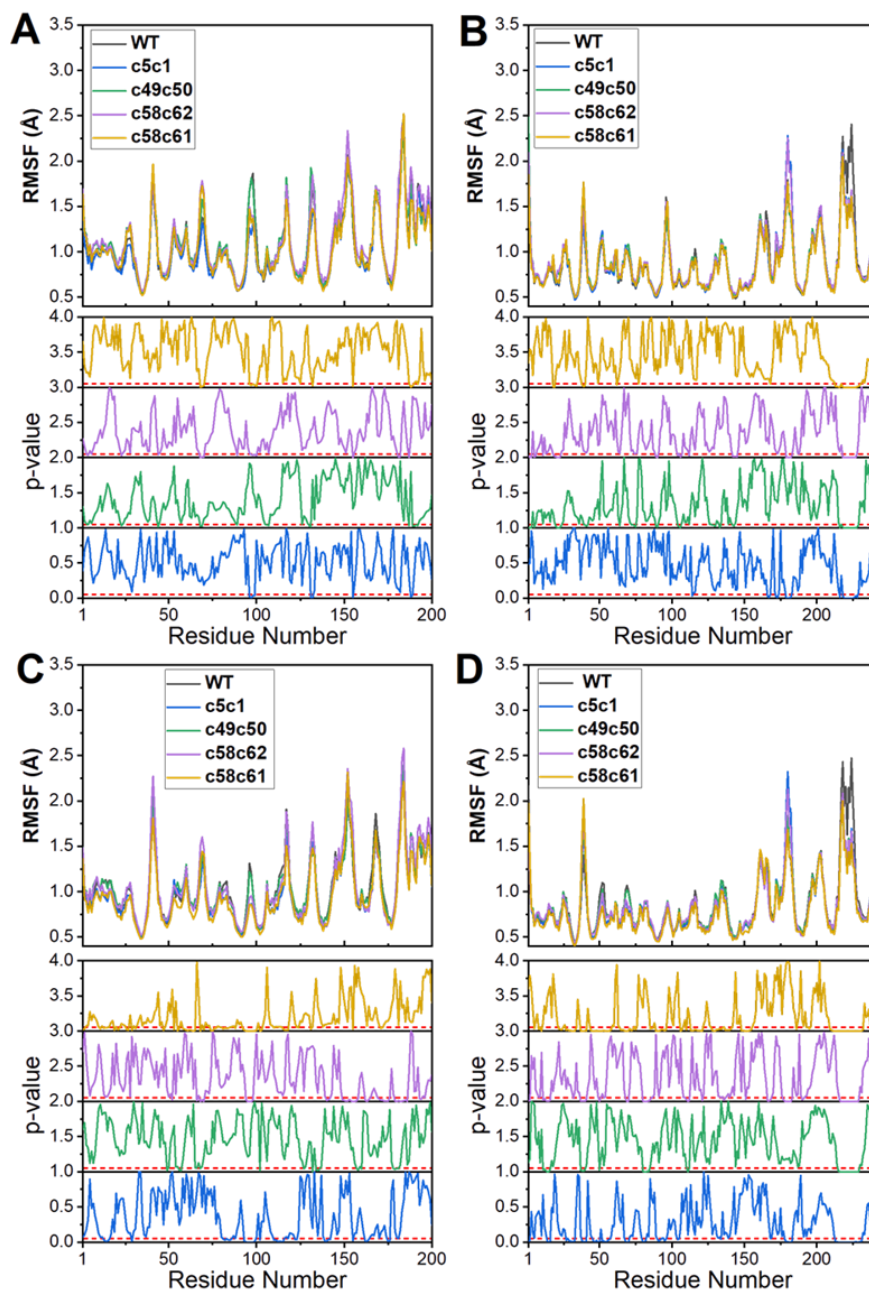


Figure S1: α RMSF values for all Apo (A + B) and pHLA bound (C + D) NY-ESO TCRs simulated. The α -chain RMSFs for the apo and pHLA bound simulations are plotted in panels A and C respectively, whilst the β -chain RMSFs for the apo and pHLA bound simulations are plotted in panels B and D respectively. Below each plot is the p-value obtained from a two-sample *t*-test between the WT and each affinity matured NY-ESO TCR (following the same colour scheme). A red dotted line is plotted at a p-value of 0.05, which is the cut-off used to determine significance.

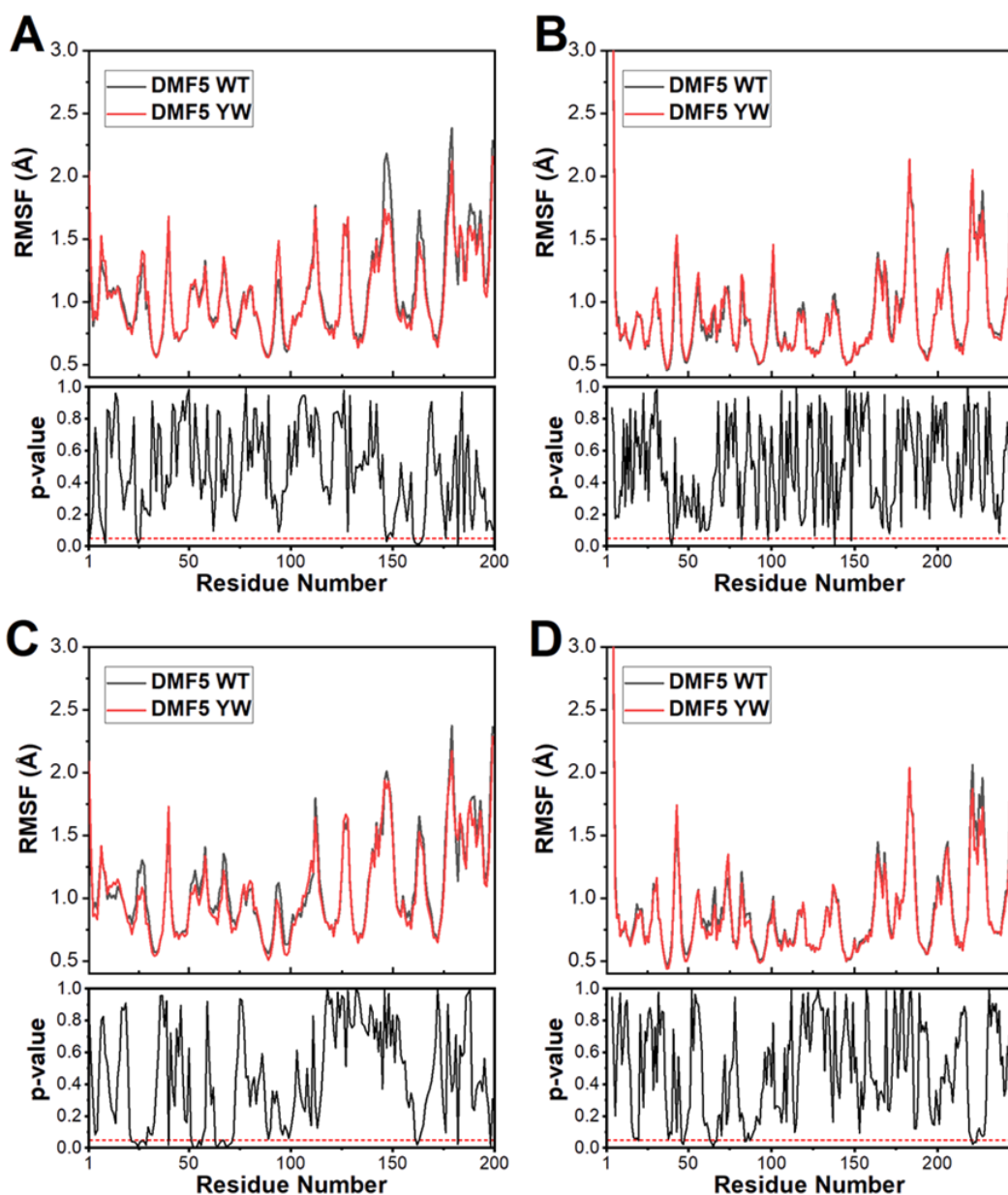


Figure S2: $C\alpha$ RMSF values for both Apo (A + B) and pHLA bound (C + D) DMF5 TCRs simulated. The α -chain RMSFs for the apo and pHLA bound simulations are plotted in panels A and C respectively, whilst the β -chain RMSFs for the apo and pHLA bound simulations are plotted in panels B and D respectively. Below each plot is the p-value obtained from a two-sample t -test between the WT and the high affinity DMF5 TCRs. A red dotted line is plotted at a p-value of 0.05, which is the cut-off used to determine significance.

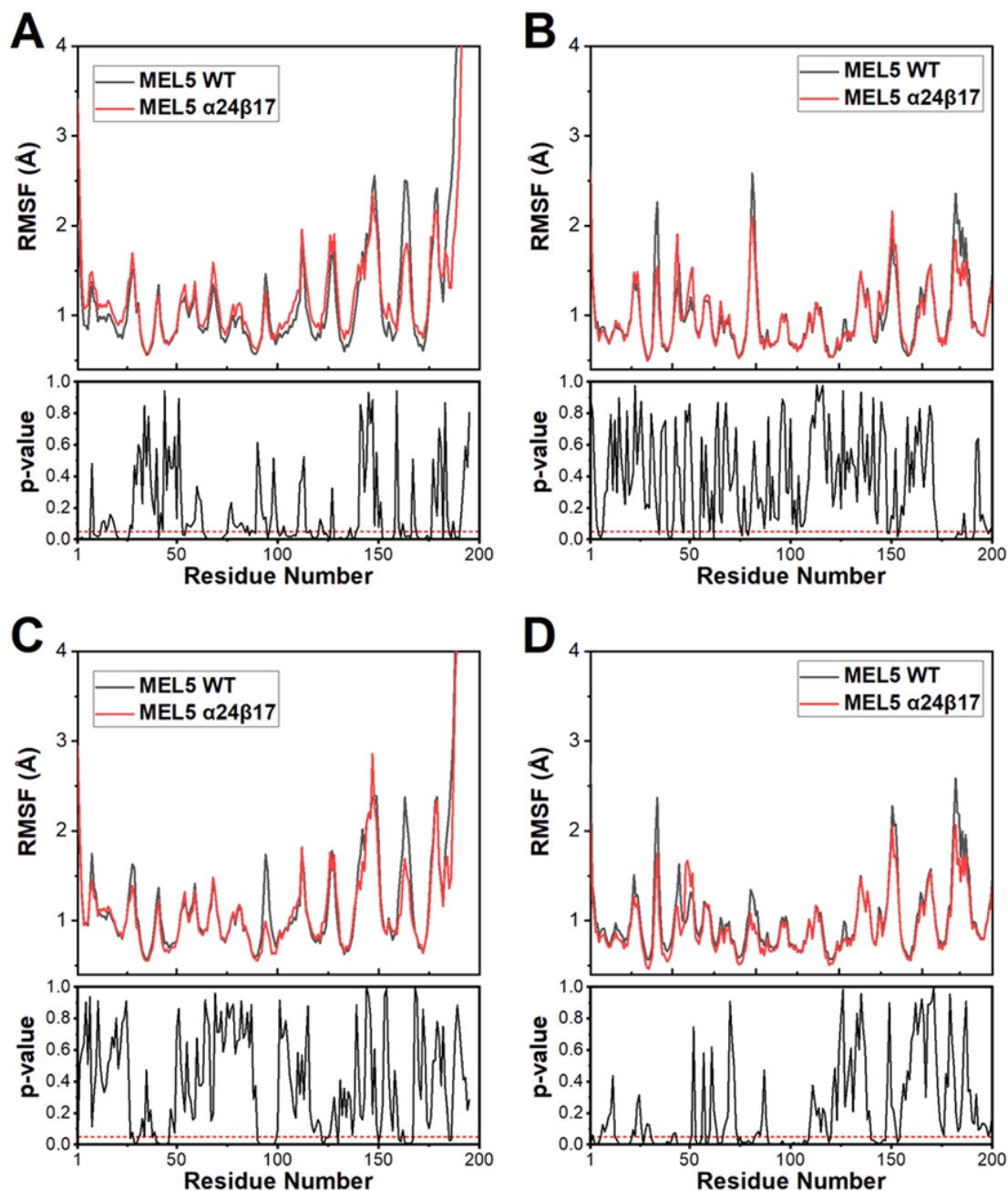


Figure S3: $C\alpha$ RMSF values for both Apo (**A + B**) and pHLA bound (**C + D**) MEL5 TCRs simulated. The α -chain RMSFs for the apo and pHLA bound simulations are plotted in panels **A** and **C** respectively, whilst the β -chain RMSFs for the apo and pHLA bound simulations are plotted in panels **B** and **D** respectively. Below each plot is the p-value obtained from a two-sample *t*-test between the WT and the high affinity DMF5 TCRs. A red dotted line is plotted at a p-value of 0.05, which is the cut-off used to determine significance.

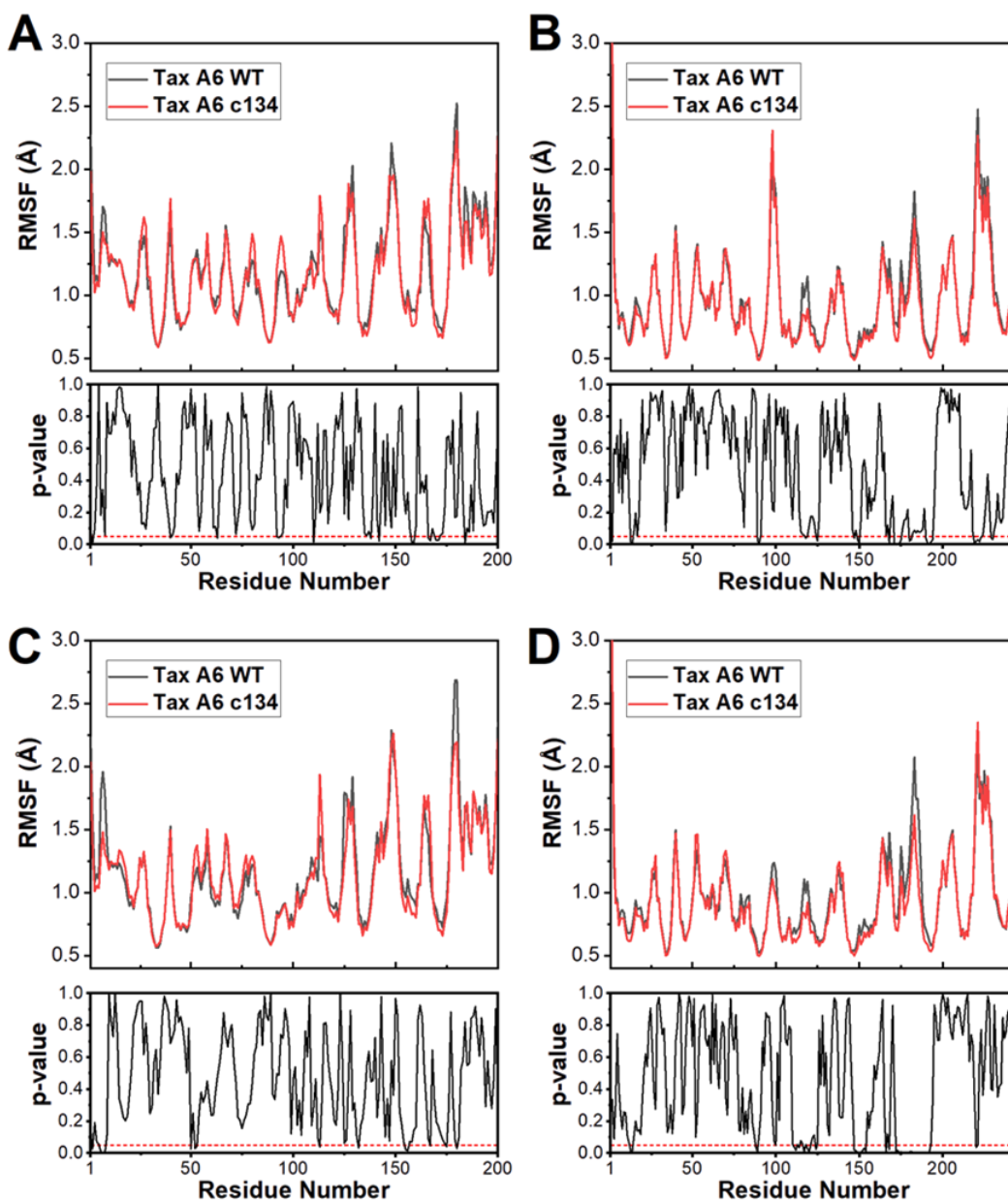


Figure S4: $C\alpha$ RMSF values for both Apo (A + B) and pHLA bound (C + D) Tax A6 TCRs simulated. The α -chain RMSFs for the apo and pHLA bound simulations are plotted in panels A and C respectively, whilst the β -chain RMSFs for the apo and pHLA bound simulations are plotted in panels B and D respectively. Below each plot is the p-value obtained from a two-sample *t*-test between the WT and the high affinity DMF5 TCRs. A red dotted line is plotted at a p-value of 0.05, which is the cut-off used to determine significance.

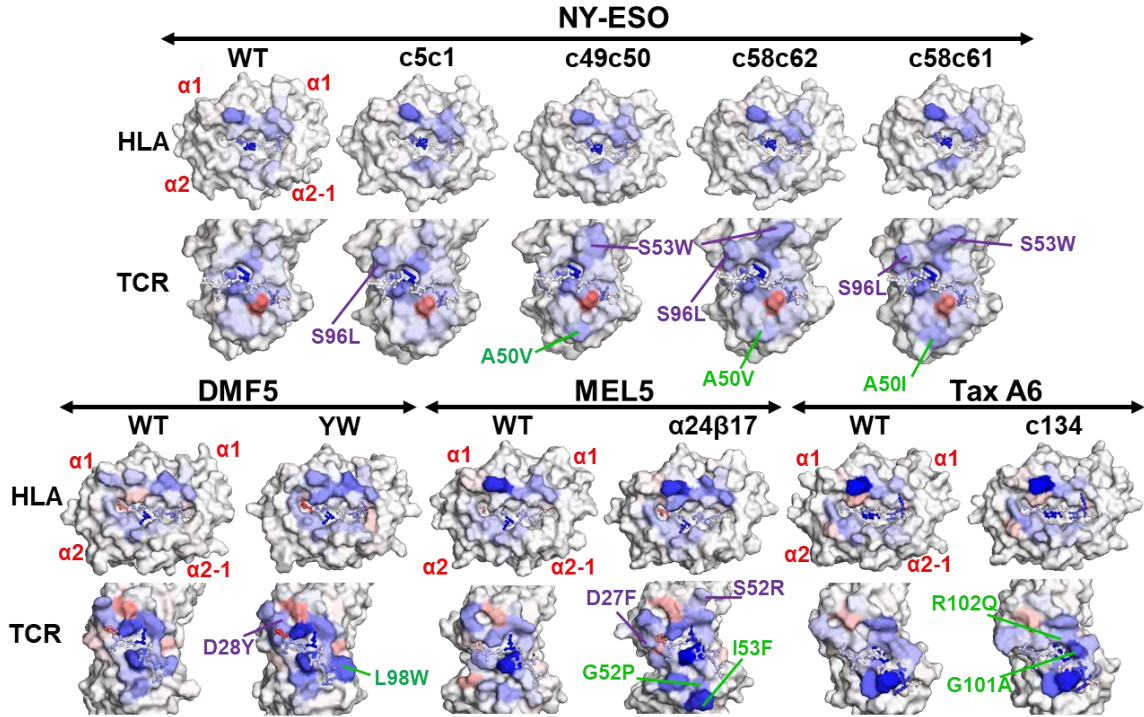


Figure S5: Changes in the energetic footprint between the WT and affinity matured TCRs. For all TCR-pHLA complexes, the HLA (top) and TCR (bottom) structures are plotted as surfaces with the peptide shown in both structures as sticks. All plots are colour mapped according to the MMGBSA per residue decomposition results, going from blue (favours binding) to white (no preference) to red (disfavours binding). Separate scaling is used for each of the 4 sets of TCRs. All pHLA and TCR structures are shown in the same orientation, such that the peptide N-terminus is left and the C-terminus right. Several mutations sites are indicated on the high affinity variants (purple labels: CDR α mutations; green labels: CDR β mutations).

Chapter 6:

Conclusions and Future Work

The work from this thesis focussed on gaining fundamental insights into TCRs and their therapeutic targets, pHLA molecules. **Chapter 3** demonstrated how different peptide cargo can tune the molecular flexibility of the entire HLA molecule, including areas distal from the peptide binding site. Whilst it is clear that different peptides will modulate the TCR recognition process both directly (through peptide-TCR interactions) and indirectly (through HLA-TCR interactions), how different peptide cargo could alter the interaction between the pHLA and the CD8 co-receptor (required for an immune response, see **Section 1.3**) remains an open question. Furthermore, HLA Class I (or II) pathways involve many protein-protein interactions prior to presentation on the cell surface^{8,9}, meaning the peptide cargo could play an important role in tuning these interactions as well. One should also consider how impactful these changes in peptide cargo are on the overall conformational dynamics of the pHLA molecule, and therefore how impactful it is in modulating the pHLA molecules interactions with other proteins. For example, although we have identified several statically significant differences in flexibility for regions of the HLA (which are peptide cargo dependent), the effect/impact of these differences may still be too subtle to play an important role in the regulation of immunological interactions/pathways. Only through further experimentation and simulation could one attempt to determine the significance/role of peptide dependant tuning of the pHLA molecule in immunological pathways.

To further the work outlined in **Chapter 3**, an increased understanding of how the peptide communicates conformational changes to the HLA could be achieved by studying many more peptides in increased detail. Our MD simulations of six relatively different peptide sequences (as much of the variation in the sequences was in the N and C-termini, which are the residues which primarily contact the HLA, see **Section 1.4**) consistently observed the C-terminal peptide residues to be primarily responsible for communication with the HLA. One could therefore ask if the C-terminal residues of all peptides that binds the HLA isoform studied in **Chapter 3** are consistently responsible for the majority of the communication between the peptide and HLA, and if so whether this extends to all HLA Class I molecules. Regardless of whether the above questions are true or not, studies of this nature would help to inform on how to tune the conformational dynamics of the pHLA molecule. This could be of use in vaccine design using peptide-mimetics.²⁵⁴ One could argue that it would be most beneficial to replicate as best as possible the dynamics induced by the antigenic peptide onto the HLA with the peptide-mimetic. Follow on simulations should also consider including a lipid bilayer into the MD simulations, in order to be closer to the biologically relevant state.

As stated above, the impact of peptide dependant tuning of the pHLA molecules conformational dynamics has yet to be explored. One such area that could be focussed on is how different peptide cargo modulate the strength of the interaction between the pHLA and the CD8 co-receptor (which binds the pHLA molecule far from the peptide binding site). A study like this would also have the benefit of providing insight into the conformational transitions that occur to produce an immune response from the T-cell. However, studying the formation of a productive immune synapse (interface between T-cell and APC) does pose several technical challenges, with the first being correctly representing the biologically relevant state(s). Prior studies have shown multiple simultaneous TCR-pHLA binding events are required to occur between a given T-cell and the antigen presenting cell (APC) in order for an immune response to actually be triggered.¹⁶⁸ These studies have found at least three simultaneous TCR-pHLA complexes needed to be formed (with higher numbers possible and perhaps even more common).¹⁶⁸ This could therefore make simulating the biologically relevant state(s) challenging. A previous MD simulation study consisting of just one TCR-pHLA complex, one CD8 co-receptor and two lipid bilayers consisted of 329,265 atoms,²⁵⁵ which is already large by all-atom MD simulation standards. Given the likely requirement to simulate potentially multiple copies of the TCR-pHLA and CD8 co-receptors, this number will become even larger. Whilst simulations of this size are possible with all-atom MD (a recent 1 μ s long all-atom study on a HIV capsid containing over 64 million atoms demonstrates this²⁵⁶), one would likely need specialist equipment for (reasonably long) simulations of this size which are not currently available to most academic labs. One could instead consider applying a coarse-grained force field to make the project more tractable. The loss of accuracy when switching from all-atom MD to coarse-grained models is likely to be problematic, especially as protein-protein interactions are known to be poorly described by coarse-grained models (protein-protein interactions tend to be over stabilised and end up aggregating with one another).²⁵⁷

A potentially better suited part of the immune system to study is the process known as “peptide editing”, whereby chaperones are used to load peptides onto HLA Class I molecules before they are transported to the cell surface.²⁵⁸ Recently solved crystal structures of the chaperone TAP-binding protein-related (TAPBPR) in complex with pHLA molecules demonstrates they play an active role in both dislodging low affinity peptides and loading high affinity peptides onto the HLA.^{149,259} TAPBPR inserts a loop into the HLA F-pocket (C-terminal binding site of the peptide), which can dislodge low affinity peptides (for the given HLA) and be dislodged by high affinity peptides (for the given HLA). Alike HLA Class I molecules, TAPBPR is also membrane bound and binds to a large portion of the pHLA, including regions distal from the peptide binding groove. In **Chapter 3** we found several of these regions show significantly different flexibility dependant on the peptide cargo, making this protein-protein interaction a good system to study in order to evaluate the impact of different peptide cargo. Furthermore, whilst crystal structures have captured snapshots of this exchange mechanism, MD simulations could provide a much greater deal of insight into how these large scale conformational

changes occur, further enhancing our fundamental understanding of this immunological mechanism. These large scale conformational changes would be hard to study with “normal” MD simulations as they are likely to occur on the μ s-ms timescale (making them very computationally expensive to sample with normal MD simulations). Enhanced sampling techniques such as Gaussian accelerated MD²⁶⁰ (GaMD) or Hamiltonian replica exchange²⁶¹ (HREX) could therefore be used to accelerate the observation of rare events, making the project more feasible.

Work in **Chapters 4** and **5** focussed on determining engineering principles for generating high specificity and high affinity TCRs. In terms of specificity, TCRs that bind with a broad energetic footprint and make interactions with several peptide residues (particularly with the peptide side chains) tend to show increased specificity. Additionally, the affinity of TCRs can be enhanced by numerous mechanisms, with our findings suggesting that the burial of large hydrophobic or aromatic residues are commonly used to enhance affinity. The above observations are relatively straight forward to rationalise, and arguably of more importance is the validation of computational approaches that allow us to investigate these phenomena. Of particular note is the MMPB/GBSA decomposition procedure, which provides a measure of each residue’s “favourability” towards being in the bound vs unbound state. Care should be taken when interpreting the meaning of a “favourable” or “unfavourable” residue (in particular for the purpose of rational design). This is because whilst an individual residue may be determined by the decomposition to be unfavourable, it could have an overall positive impact on the binding affinity, by increasing the favourability towards binding for its neighbouring residues. Examples of this can be found in **Chapter 5** (see **Figure 30+Figure 31**). For instance, the NY-ESO TCR mutation G97D was calculated to be unfavourable for that residue (likely due to desolvation of the charged side chain), but overall had a favourable effect on the binding affinity, by rigidifying the apo CDR3 α loop and improving the quality of other CDR3 α residues interactions with the pHLA. An alternative to decomposition analysis is computational alanine scanning, for which there are two versions available for the MMPB/GBSA approach (and more generally to most computational binding free energy calculation methods).³⁹ In the first version, a whole new set of MD simulations are run on the alanine variant and MMPB/GBSA calculations are run on these new trajectories. In the second version, the side chain of the alanine variant is simply deleted from the previously performed WT MD simulations and MMPB/GBSA calculations are performed on these trajectories. Whilst both methods are slower than decomposition analysis (as only one MMPB/GBSA calculation is required for decomposition), the second alanine scanning version is notably faster than the first because it requires no additional MD simulations. All three methods ultimately provide different information and it is up to the end user to decide what is most relevant and most useful to them.

Further, the MMPBSA calculations in **Chapter 4** were able to rationalise the specificities of different TCRs or TCR-mimics, whilst the MMGBSA calculations performed in **Chapter 5**, showed how one could compare the energetic footprints of the WT and affinity matured TCRs. This demonstrates how MMPB/GBSA calculations could be used

to predict the specificity of a given TCR towards its pHLA target, and therefore it's potential use to filter which TCR variants are taken forward for experimental specificity/safety testing.

The work in **Chapter 5** highlighted the relatively common occurrence of large aromatic or hydrophobic side chains being buried into the binding site to enhance affinity, which is in line with observations that tryptophan, methionine and phenylalanine are the most commonly observed residues in protein-protein binding sites.²⁴⁸ If one wishes to take advantage of the hydrophobic effect (the entropically favourable expulsion of ordered water molecules at the binding interface to become “bulk-like” and therefore more disordered), computational calculations on the thermodynamics of water at the binding interface of the WT TCR-pHLA complex could be performed. Several such methods exist of differing levels of accuracy and speed (e.g. ordered from fastest/most approximate to slowest/most accurate: 3D-RISM²⁶², GIST²⁶³ or Grand Canonical Monte Carlo²⁶⁴ methods) and have been used extensively in small molecular drug discovery to rationalise quantitative structure–activity relationship models.^{265–267} The identification of unfavourable water binding sites could then lead one to perform experimental or computational mutagenesis studies on residues near to these regions (to form new contacts that are likely to lead to an increase in affinity as the water molecules they replace are relatively unstable). Further, favourable and highly ordered water molecules bound to the pHLA which currently do not contact the TCR are likely to be a good target for engineering a water bridged hydrogen-bond between the TCR and pHLA (because the entropic penalty associated with binding the water molecule has already been largely paid). Interestingly, a previous docking study on a TCR-pHLA complex showed an improved relationship with experimental affinity data by the inclusion of a water molecule that formed a bridged hydrogen bond.²⁶⁸

Our simulations in **Chapter 5** were able to reproduce the experimental affinity relationships observed between the WT and affinity matured variants. In work not shown in this thesis (as it is currently underway), we are testing the applicability of MMPB/GBSA calculations as a medium throughput screening approach for computational affinity maturation of TCRs. The amount of MD simulation time required (in **Chapters 4** and **5** we use 25 x 4 ns per TCR-pHLA) for MMPB/GBSA calculations make it far too computationally expensive for high throughput screening. Instead, a docking approach such as those detailed in **Section 1.8.1** would be far more suited for this. Following a docking calculations to predict higher affinity variants, the top X number of variants could be screened with MMPB/GBSA to reduce the number required to take forward experimentally (where X is a reasonably small number, say 30–100).

Given the above sections on the possible applications of this work and the methods used, a suggested protocol for the use of computation towards TCR engineering is provided below. Ideally, at the beginning of the project an experimentally validated pHLA target along with a WT-TCR which binds to this pHLA will have been identified and crystallised in complex with each other. If the crystal structure is not available, a necessary

combination of homology modelling and docking (perhaps with TCR specific docking algorithms⁵²) could be used to generate the starting model for rational design. At this point, the WT-TCR-pHLA complex should be thoroughly characterised with both MMPB/GBSA decomposition analysis and water thermodynamics calculations. The information gleaned from these calculations can be used to suggest residues on the TCR that one should screen for beneficial mutations, with considerations towards both affinity and specificity (as outlined in **Chapter 4**). High throughput computational docking should now be performed on these residues to identify possible mutations that enhance the binding affinity. The water thermodynamics calculations used above could be combined with water placement algorithms²⁶⁹ to include highly energetically favourable water molecules as part of the receptor in these docking protocols. Following the identification of possible beneficial mutations, the top portion can be subjected to experimental testing or MMPB/GBSA calculations to re-rank the results and reduce the number of variants subjected to experimental testing. If MMPB/GBSA calculations are performed on the variants, this information can later be used to predict which variants are more likely to have a better specificity profile and should be prioritised for experimental specificity/safety validation (after confirming they are of high enough affinity). Several iterative rounds of computational mutagenesis and screening may need to be performed to obtain the target affinities.

In conclusion, the work from this thesis has extensively characterised the TCR-pHLA interaction. This has provided a foundation for several areas of future work, in particular in the use of computation in the rational design of TCRs as therapeutics.

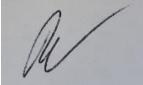
Appendix Chapter 1: Understanding the Role of Heat Capacity in Enzyme Catalysis

The final two chapters included in this thesis are publications from collaborative projects produced during my PhD thesis. These publications are however not focussed on or around TCRs (instead enzymes). To allow for a more straightforward focus in my thesis, both of the following chapters have instead been included as appendices only.

In this first chapter we used a combination of experiment and computation to gain insight into the heat capacity of catalysis (ΔC_p^\ddagger), which is a thermodynamic parameter that can be incorporated into the Eyring equation to describe a non-linear relationship between temperature and enzyme activity that has been observed in several enzymes and cannot be explained by unfolding. ΔC_p^\ddagger is a measure of the change in the distribution and frequency of vibrational (and rotational and translational) modes between the ground and transition state. As isotope effects are driven by changes in the frequency of vibrational modes between isotopologues, we experimentally determined ΔC_p^\ddagger for a number of different substrate isotopologues using the model enzyme glucose dehydrogenase (GDH), which catalyses the oxidation of several sugars.

We found small isotopic substitutions (which would have a direct effect on the value of ΔC_p^\ddagger on the order of a few $\text{J mol}^{-1} \text{K}^{-1}$) were able to lead to large scale changes in ΔC_p^\ddagger on the $\text{kJ mol}^{-1} \text{K}^{-1}$ level. This would suggest small changes in substrate vibrational modes are able to give rise to relatively large changes in the distribution/magnitude of vibrational modes between the ground and transition state. Our QM cluster model of the enzyme active site predicted $\sim \text{J mol}^{-1} \text{K}^{-1}$ changes in ΔC_p^\ddagger upon isotopic substitution, suggesting the majority of the change ΔC_p^\ddagger is sourced from the remainder (not the active site) of the enzyme.

My role in this project was to construct a QM cluster model of the enzyme reaction mechanism. This was used to 1) determine the chemical mechanism of the reaction (which we obtained to be stepwise with hydride transfer rate limiting, in agreement with the experimentally determined KIEs, and 2) simulate the effect of isotopic substitutions with the described cluster model, to predict changes in ΔC_p^\ddagger for different isotopologues.

This declaration concerns the article entitled:									
Uncovering the relationship between the change in heat capacity for enzyme catalysis and vibrational frequency through isotope effect studies									
Publication status (tick one)									
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>	In review	<input type="checkbox"/>	Accepted	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	<p>Jones, H.B.*, Crean, R.M.*, Matthews, C., Troya, A.B., Danson, M.J., Bull, S.D., Arcus, V.L., Van der Kamp, M.W. and Pudney, C.R., 2018. Uncovering the Relationship between the Change in Heat Capacity for Enzyme Catalysis and Vibrational Frequency through Isotope Effect Studies. <i>ACS Catalysis</i>, 8(6), pp.5340-5349. https://pubs.acs.org/doi/abs/10.1021/acscatal.8b01025</p> <p>* = Equal Contributions.</p>								
Candidate's contribution to the paper (detailed, and also given as a percentage).	<p>The candidate contributed to/ considerably contributed to/predominantly executed the...</p> <p><u>Formulation of ideas: (25%)</u> Laboratory experiments formulated by HBLJ and CRP Computational experiments formulated by RMC, MvdK</p> <p><u>Design of methodology: (25%)</u> Laboratory experiments designed by HBLJ and CRP Computational experiments designed by RMC and MvDK</p> <p><u>Experimental work: (25%)</u> Experimental work by HBLJ and CM Structure preparation for MD simulations by HBLJ MD simulations were carried out by MvdK QM cluster calculations were carried out by RMC and AT</p> <p><u>Presentation of data in journal format: (25%)</u> CRP, HBLJ, RMC and MvdK wrote the main text HBLJ made figures, wrote laboratory methodology and referenced text RMC made figures, formulated SI and wrote computational methodology and results.</p>								
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.								
Signed					Date	02/09/19			

Uncovering the relationship between the change in heat capacity for enzyme catalysis and vibrational frequency through isotope effect studies

Hannah BL Jones^{†Δ}, Rory M Crean^{#†Δ}, Christopher Matthews[†], Anna B Troya[†], Michael J Danson[†], Steven D Bull[‡], Vickery L Arcus^{§*}, Marc W Van der Kamp^{¶*} and Christopher R Pudney^{†*}

[†]Department of Biology and Biochemistry, [‡]Department of Chemistry, [#]Doctoral Training Centre in Sustainable Chemical Technologies, University of Bath, Bath BA2 7AY, United Kingdom

[§]School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand

[¶]School of Biochemistry, University of Bristol, Biomedical Sciences building, University Walk, Bristol BS8 1TD, United Kingdom.

Abstract: Understanding how enzyme catalysis varies with temperature is key to understanding catalysis itself, and ultimately, how to tune temperature optima. Temperature-dependence studies inform on the change in heat capacity during the reaction, ΔC_p^\ddagger , and we have recently demonstrated that this can expose links between the protein free energy landscape and enzyme turnover. By quantifying ΔC_p^\ddagger , we capture information on the changes to the distribution of vibrational frequencies during enzyme turnover. The primary experimental tool to probe the role of vibrational modes in a chemical/biological process is isotope effect measurements, since isotopic substitution primarily affects the frequency of vibrational modes at/local to the position of isotopic substitution. We have monitored the temperature-dependence of a range of isotope effects on the turnover of a hyper-thermophilic glucose dehydrogenase. We find a progressive effect on the magnitude of ΔC_p^\ddagger with increasing isotopic substitution of D-glucose. Our experimental findings, combined with molecular dynamics simulations and quantum mechanical calculations, demonstrate that ΔC_p^\ddagger is sensitive to isotopic substitution. The magnitude of the change in ΔC_p^\ddagger due to substrate isotopic substitution indicates that small changes in substrate vibrational modes are ‘translated’ into relatively large changes in the (distribution and/or magnitude of) enzyme vibrational modes along the reaction. Therefore, the data suggest that relatively small substrate isotopic changes are causing a significant change in the temperature-dependence of enzymatic rates.

KEYWORDS *Heat capacity, isotope effect, enzyme, catalysis, temperature-dependence*

Isotope effects are one of the most powerful tools to investigate chemical mechanism and the physical chemical underpinnings of catalysis.²⁷⁰ Enzyme mechanistic studies often employ kinetic and binding isotope effects, KIE and BIE, to access information on

specific steps in enzyme catalysis and to test the role of protein ‘dynamics’ in enzyme turnover.^{271–273} The power of isotope effects lies in the simplicity of their physical origin. That is, isotopic substitution decreases the frequency of all vibrational modes, but does not affect the electronic structure of the molecule.²⁷⁴ Therefore, where one observes an isotope effect, the physical chemical interpretation of the experimental observation is much less ambiguous compared to mutagenesis studies, for example.

In enzymatic studies, the temperature-dependence of the KIE is often used to infer the presence or absence of protein motions that affect the rate of enzymatic turnover.^{275,276} The microscopic interpretation of these studies is controversial (see e.g. ref. ²⁷⁷), but at least at a basic level these studies seem to validate the notion that protein motions can affect enzyme turnover (if not necessarily the catalytic step itself). Recently, the potential role of protein ‘dynamical’ effects on the reaction catalyzed by dihydrofolate reductase (DHFR) or HIV-1 protease, have been assessed based on simulations with mass modulated (isotopically substituted) enzymes.^{278–280} In both systems, a small contribution from these dynamical effects to reducing the free energy profile around the transition state of the order ~ 2 kJ mol⁻¹ was found. These authors interpret this difference as arising from the coupling between the reaction coordinate and the degrees of freedom of the system. However, the major contributor to reducing the free energy barrier arises from electrostatic effects.^{271,281–283} In addition, Åqvist used simulation to reveal the molecular origin of entropic effects in catalysis with respect to temperature, and illustrated the importance of considering not just the immediate active site, but also remote parts of the protein (and surrounding solvent).²⁸⁴ We further point out that others have argued that fast (sub-picosecond), local active site protein modes play a role in transition state formation (and thereby catalysis).²⁸⁵ A further alternative view is that so-called protein dynamical effects are coupled to the reaction coordinate, but provide a small contribution to barrier reduction relative to, e.g., the electrostatic contribution via preorganisation.^{271,281–283} Although not a catalytic ‘dynamical effect’, when strictly defined,²⁸⁶ differences in enzyme fluctuations (or vibrations) that cause a change in heat capacity along the reaction can affect the temperature-dependence of enzyme activity.²⁸⁷ Understanding this effect in detail may provide new tools to manipulate enzyme optimal temperatures.

We have recently demonstrated how capturing information on the change in heat capacity for activation ΔC_p^\ddagger , informs on differences in the distribution of frequencies of vibrational modes between the enzyme-substrate and enzyme-transition state complex.^{288,289} Typically, the temperature-dependence of enzyme rate constants is fitted to the Eyring equation (Eq 15).

$$k = (k_B T/h) e^{\Delta S^\ddagger/R} e^{-\Delta H^\ddagger/RT} \quad (\text{Eq 15})$$

This model assumes that ΔH^\ddagger and ΔS^\ddagger are temperature independent. However, if ΔH^\ddagger and ΔS^\ddagger are temperature-dependent, then this implies a non-zero value for ΔC_p^\ddagger . ΔC_p^\ddagger can be extracted from temperature-dependence studies of enzymes fitted using a model that

incorporates temperature-dependence of ΔH^\ddagger and ΔS^\ddagger , which we have termed macromolecular rate theory (MMRT, (Eq 16):^{290,291}

$$\ln k = \ln \frac{k_B T}{h} - \left[\frac{\Delta H_{T_R}^\ddagger + \Delta C_p^\ddagger (T - T_R)}{RT} \right] + \left[\frac{\Delta S_{T_R}^\ddagger + \Delta C_p^\ddagger (\ln T - \ln T_R)}{R} \right] \quad (\text{Eq 16})$$

where T_0 is an arbitrary reference temperature. ΔC_p^\ddagger is the difference in heat capacity between the ground and transition states. ΔC_p^\ddagger determines the change in ΔH^\ddagger and ΔS^\ddagger with temperature and thereby defines the non-linearity of the temperature-dependence of the Gibbs free energy difference between the ground state and the transition state (ΔG^\ddagger). Indeed, Roy *et al.* point to a temperature-dependent activation entropy as the source of non-linear temperature-dependence plots.²⁷⁷ Other models that move beyond (Eq 15) have been proposed, primarily relating to equilibria of different functional/non-functional states.^{292,293} Whilst we do not discount these models, it appears, based on the range of recently published work from different labs,^{287–289,291} that Eq 16 is useful and broadly accurate.

We expect that the dominant contribution to ΔC_p^\ddagger in enzymes is the difference in distribution and frequency^{290,291} of the large number of vibrational modes of the molecule and its closely associated solvent molecules in the ground and transition states. Alternatively, a negative value of ΔC_p^\ddagger implies that $\langle(\delta H)^2\rangle$ (the mean squared distribution of enthalpies) for the enzyme substrate complex is greater than $\langle(\delta H)^2\rangle$ for the enzyme-transition state complex at a given temperature.²⁹¹ We suggest that the magnitude of ΔC_p^\ddagger can therefore be used as a proxy for the changes in vibrational modes (distribution, frequency) during enzyme turnover.

The origin of isotope effects lies in the difference in the frequency of vibrational modes between isotope changes. Observing a relationship between (i) increasing isotopic substitution of the substrate, (ii) an isotope effect on the rate of turnover, and (iii) ΔC_p^\ddagger would suggest that the enzymatic ΔC_p^\ddagger is sensitive to (substrate) vibrational modes that affect the observed reaction rate. Bigeleisen considered the effect of isotopic substitution on heat capacity, though not for hydrogen.²⁹⁴ More recently, Tjahjono and Garland²⁹⁵ have directly measured the difference in apparent molar heat capacity, C_p^0 , for a series of model compounds with different levels of deuterium substitution. As with other reports,²⁹⁶ the authors find that the deuterium isotope effect on C_p^0 was always positive, i.e. C_p^0 increases with the increasing number of deuterium substitutions (Nd) and that the relationship was essentially linear: ΔC_p^0 (J mol⁻¹ K⁻¹) = 2.75Nd – 1.52. The increase in C_p^0 is attributed to the increased mass of the isotopologue and the concomitant decrease in frequencies of the affected bonds.

In a recent appraisal of a range of previously published experimental enzyme systems,²⁸⁹ we found that there was potentially an isotope effect on ΔC_p^\ddagger . Longbotham *et al*²⁹⁷ recently performed a study exploring a range of isotope effects on labelling the flavin in a model flavoenzyme and found their data could only be adequately fitted using Eq 16. These authors find a small isotope effect on ΔC_p^\ddagger outside experimental error for some, but not all, labelling patterns. Based on these intriguing findings we are inspired to explore the potential for isotope effects combined with Eq 16 to inform on changes in enzyme vibrational modes along the chemical reaction coordinate and their relationship to temperature-dependence. Herein, we use a hyperthermophilic enzyme, *Sulfolobus solfataricus* glucose dehydrogenase (*ssGDH*) as a model system to explore the contribution of enzymatic isotope effects on heat capacity differences during the reaction. Our data provide an experimental link between the temperature-dependence of enzyme turnover and (low frequency) vibrational modes. Importantly, the results raise the question of how subtle localized changes (through isotopic substitutions in the substrate) can lead to a significant change in the enzymatic heat capacity (and thus $\langle(\delta H)^2\rangle$) during the reaction.

Results and discussion

Hydride transfer in *ssGDH* is rate determining. *ssGDH* is a promiscuous hyperthermophilic enzyme that reduces nicotinamide adenine dinucleotide (phosphate) (NAD(P)⁺), with a variety of sugars.²⁹⁸ Milburn *et al*²⁹⁹ solved the X-ray crystal structure with both NADP⁺ and glucose/xylose bound using a catalytically inactive variant (T41A). Notionally the reduction of the nicotinamide (C4) occurs as a hydride transfer from C1 of the sugar, concomitant with a proton transfer from the C1 hydroxyl (Scheme S1) to a water molecule or hydroxide ion coordinated by a Zn²⁺ ion. We have performed molecular dynamics simulations of *ssGDH* (**Figure 38A**) in complex with both glucose and xylose (four independent runs of 50 ns, with all four active sites occupied). From our MD simulations, consistent with the X-ray crystal structures, we find that the hydride donor-acceptor (D-A) distance is very similar for glucose and xylose, but not identical, with the xylose D-A being ~0.2 Å longer (**Figure 38B**; Supporting information). Specifically, the averages are 2.77 +/- 0.287 Å and 2.93 +/- 0.310 Å; medians are 2.725 Å and 2.895 Å.

Quantum chemical cluster calculations (**Figure 38C–20G**) on a 148 atom model of the active site of *ssGDH* in complex with glucose (**Figure 38C**) suggest a stepwise chemical mechanism, whereby proton abstraction from a Zn²⁺ coordinated hydroxide forms a stable alkoxide intermediate on the sugar C1 prior to hydride transfer. When a water molecule is modelled as coordinated to the Zn²⁺ ion (**Figure 38C**), the D-A distance is 2.49 Å, consistent with our MD simulations (above, **Figure 38B**). When instead a hydroxide ion is modelled as coordinated to the Zn²⁺ ion, the proton on the sugar C1 hydroxyl is readily abstracted upon geometry optimization (**Figure 38D**) to form a reactive intermediate, with the resulting anion stabilised by hydrogen bonding interactions to the now Zn²⁺ bound water and T41. This reactive intermediate has a reduced D-A distance of 2.22 Å, which

reduces further as the glucose and NADP⁺ rings pucker to form a transition state (**Figure 38F**) with a predicted free energy barrier relative to the reactive intermediate of 32.6 kJ mol⁻¹ at 298 K. In synchrony with the hydride transfer, the Zn²⁺ bound water rotates away from glucose, returning to its original position of hydrogen bonding with Q150 and the C2 hydroxyl of glucono-D-lactone. In order to compare the predicted and experimental activation energies, one must consider the initial deprotonation into the bulk milieu (Reactant to Reactive Intermediate; **Figure 38D** and **E**, respectively), which cannot be determined reliably with the cluster based approach due to the importance of inhomogeneous long range interactions in modulating pK_a values.³⁰⁰ Instead, one can subtract the hydride transfer barrier from the experimental barrier to obtain a predicted pK_a of 7.2 (Full calculation details can be found in the Supporting Information), which is in good agreement with the experimental pH optimum of 8.³⁰¹ Finally, we note that there is no obvious role for water in the rate determining step (deprotonated water acts as a base for proton abstraction, whereas the subsequent hydride transfer is rate limiting, see below) and so at least this route is not giving rise to a primary solvent isotope effect (see below).

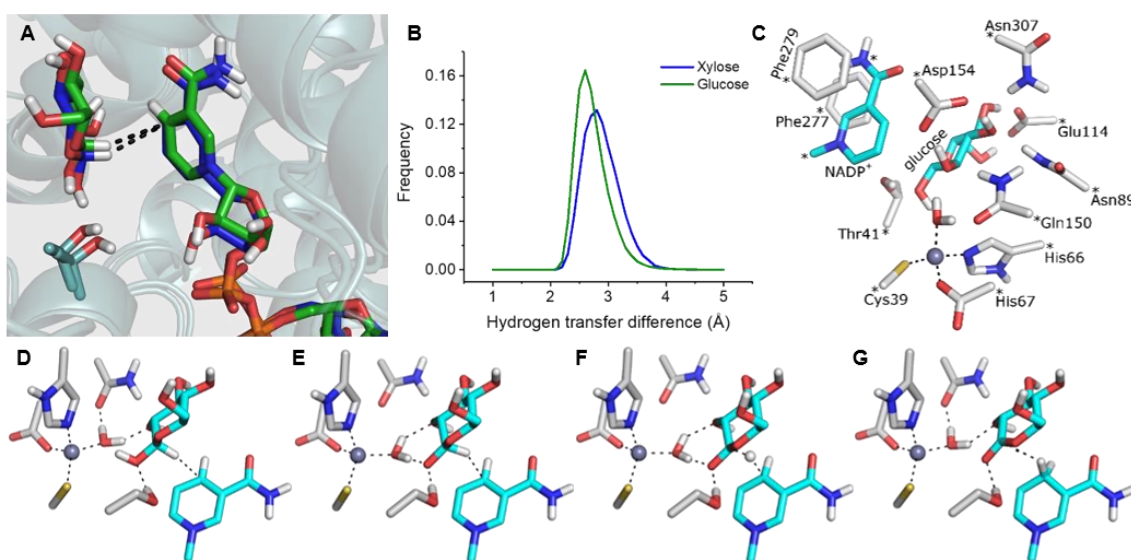


Figure 38: **A**, Representative structures of glucose (green) and xylose (blue) from our MD simulations demonstrating they have the same binding interface with NADP⁺. **B**, Normalized histograms (bin width 0.1 Å) of the hydrogen transfer distance of glucose and xylose from MD simulations of ssGDH. **C**, QM cluster model created of glucose in complex with NADP⁺, with asterisks indicating atoms fixed throughout the optimisation process. **(D–G)** Reaction mechanism obtained from the QM model, starting from the reactant (**D**), to the deprotonated reactive intermediate (**E**), the transition state (**F**), and finally the product. (**G**).

Our combined computational and kinetic data (discussed below) provide evidence for the mechanistic step observed from our kinetic data. The slightly larger D-A distance in the Michaelis complex for xylose *versus* glucose, calculated from MD simulations, correlates with a ~ 1.6 -fold larger k_{cat} value ($k_{\text{cat}} = 95$ and 58 s^{-1} at $90 \text{ }^\circ\text{C}$; **Figure 39**) and a ~ 5 -fold larger K_m ($K_m = 2.5 \pm 0.3$ and $0.5 \pm 0.2 \text{ mM}$ at $85 \text{ }^\circ\text{C}$; **Figure S1A** and **S1B**) for glucose *versus* xylose, respectively. This magnitude increase in rate seems reasonable for a hydride transfer given a $\sim 0.2 \text{ \AA}$ difference in D-A distance^{302,303} (**Figure 38B**) and implies that the observed rate is primarily capturing the hydride transfer step. Similarly, the difference in K_m is likely reflective of the difference in binding geometry and bonding (discussed below). Our QM cluster calculations provide a reasonable mechanism in which a Zn^{2+} coordinated hydroxide ion can deprotonate the glucose C1 hydroxyl to form a reactive intermediate. The direct nature of the enzyme assay (NADP^+ reduction to NADPH), our QM calculations and the observation of a significant primary kinetic isotope effect (KIE, see below) suggests that the kinetic data primarily reflect the hydride transfer step.

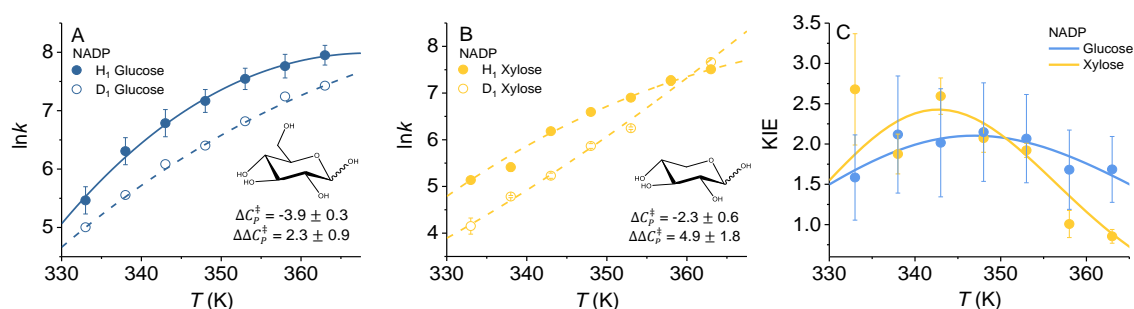


Figure 39: The temperature-dependence of NADP^+ reduction by glucose (**A**) and xylose (**B**). Solid and dashed lines show the fitted to Eq 16 for the protiated and deuterated Glucose/xylose (D^1), respectively. (**C**), the resulting KIE extracted from the $\ln k_{\text{cat}}$ ($\ln \text{min}^{-1}$) in panels **A** and **B**. The solid line is the modelled KIE based on the parameters extracted from Eq 16 (Solid lines in panels **A** and **B**).

Heat capacity changes during enzyme reaction.

We have measured the temperature-dependence of k_{cat} for both glucose and xylose, fitted to Eq 16, shown in **Figure 39A+B**. For both glucose and xylose, the temperature vs. $\ln(k_{\text{cat}})$ data are clearly curved and are therefore appropriately fitted to Eq 16 instead of Eq 15. *ssGDH* is extremely thermally stable³⁰¹ and we do not find evidence for unfolding on the timescales of our assays at any temperature. To capture the curvature in temperature-dependence plots accurately, and to capture the experimental system at its natural operating temperature ($77 \text{ }^\circ\text{C}$)³⁰¹, we have focused on the temperature range 60 – $90 \text{ }^\circ\text{C}$. From **Figure 39A+B** we find a significant difference in the magnitude of ΔC_p^\ddagger with $\Delta C_p^\ddagger = -3.9 \pm 0.3$ and $-2.3 \pm 0.6 \text{ kJ mol}^{-1} \text{ K}^{-1}$, for glucose and xylose, respectively.

Potentially, the difference in ΔC_p^\ddagger between glucose and xylose could arise through a difference in the chemical structure of the sugars, i.e. the additional CH_2OH group of glucose. For example, the hydroxyl of the hydroxymethyl group can form hydrogen bonding interactions with either E114 or H297. This additional interaction may cause a general rigidification of the glucose and the active site, and thus decrease the absolute heat capacity of the ground and transition state. We cannot, however, confidently project how this would change ΔC_p^\ddagger (i.e. how the heat capacity of the ground state is affected differently from the heat capacity of the transition state), and thus cannot assign the physical origin of the observed differences in ΔC_p^\ddagger between glucose and xylose.

Change in ΔC_p^\ddagger with substrate isotopic substitutions.

To explore the relation between ΔC_p^\ddagger and substrate vibrational states, we determine the effects of substrate isotope substitutions on ΔC_p^\ddagger for sugar dehydrogenation by *ssGDH*. **Figure 39A+B** show the temperature-dependence of the primary kinetic isotope effect (1° KIE) for hydride transfer for both glucose and xylose using D-glucose (D^1) and D-xylose (D^1), respectively. For both sugars, the KIE is temperature-dependent (**Figure 39C**) and similar in magnitude ($\sim 2 - 2.5$). Despite the similar magnitude of the KIE for both sugars, the isotope effect (IE) on the magnitude of ΔC_p^\ddagger is significantly different. The 1° IE on ΔC_p^\ddagger is very large for xylose [D-Xylose (D^1)], bringing the ΔC_p^\ddagger value to ~ 0 within error ($\Delta\Delta C_p^\ddagger = 4.9 \pm 1.8 \text{ kJ mol}^{-1} \text{ K}^{-1}$). However, for glucose the 1° IE is smaller: D-glucose (D_1) $\Delta C_p^\ddagger = -1.6 \pm 0.6 \text{ kJ mol}^{-1} \text{ K}^{-1}$ ($\Delta\Delta C_p^\ddagger = 2.3 \pm 0.9 \text{ kJ mol}^{-1} \text{ K}^{-1}$). We note that this trend in the data is recapitulated when omitting ‘outlier’ data points (**Figure S2**). Moreover, these data do not appear to be due to significant differences in the structure or bonding of the reactive complex, since the K_m values are the same within error for the protiated and deuterated substrate, $K_m = 3.2 \pm 0.4$ and $0.7 \pm 0.1 \text{ mM}$ at 85°C *c.f.* $K_m = 2.5 \pm 0.3$ and $0.5 \pm 0.2 \text{ mM}$ at 85°C for glucose *versus* xylose, respectively. The magnitude of $\Delta\Delta C_p^\ddagger$ for both xylose and glucose is surprising, as these differences are much larger than can be expected from effects on (substrate) heat-capacity for isotopic substitutions alone (discussed below). It indicates some (unknown) interaction between the reacting species and the enzyme enthalpy distribution that can be perturbed by isotopic substitution.

As discussed previously, the difference in ΔC_p^\ddagger between glucose and xylose could arise through a difference in the sugar-enzyme interactions. However, isotope effects arise from differences in the frequency of vibrational modes,²⁷⁴ and not from changes in electronic structure that can lead to additional interactions. At a phenomenological level, our data therefore provide evidence that ΔC_p^\ddagger can be significantly perturbed by the frequency of bond vibrations in the reacting species. **Figure 39C** shows the KIE as a curve resulting from the temperature-dependence parameters extracted from the fits in **Figure 39A+B**.

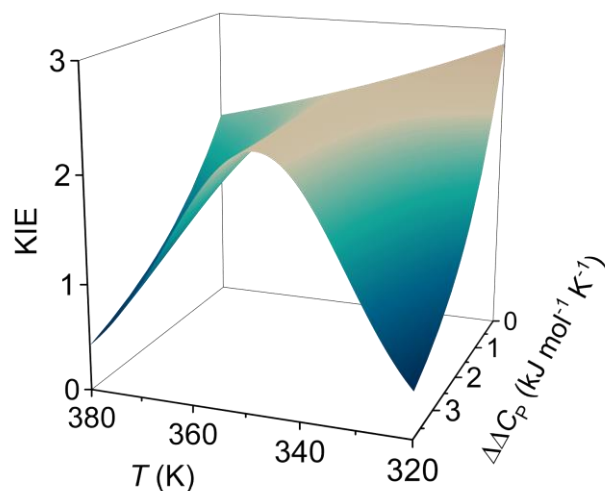


Figure 40: Numerical model showing how the magnitude of the glucose 1° KIE versus temperature is affected by differences in the isotope effect on ΔC_p^\ddagger . ΔH^\ddagger and ΔS^\ddagger values used for modelling given in Table 2.

The temperature-dependence of the KIE for both glucose and xylose show a qualitatively similar relationship; curvature with a maximum at 346 and 343 K, respectively. Typically, one observes a decrease in the magnitude of the KIE with respect to temperature when fitting data using the Eyring/Arrhenius equation (as shown in **Figure S3**). **Figure 40** shows a numerical model of the data shown in **Figure 39C** (using Eq 16), where the magnitude of ΔC_p^\ddagger for the isotopologue is varied to explore a range of $\Delta\Delta C_p^\ddagger$ values and the resulting effect on the KIE. From this model, we find that the curvature in the KIE plots is a direct result of the isotope effect on ΔC_p^\ddagger . That is, as $\Delta\Delta C_p^\ddagger$ tends towards zero (no isotope effect on ΔC_p^\ddagger), the KIE plot will become ‘normal’, showing a decrease in magnitude with increasing temperature (Figure 22). These data show that a consequence of *any* significant isotope effect on ΔC_p^\ddagger is that the KIE will tend towards unity and this is also true for all temperature-dependent KIEs fitted using e.g. Eq 16. The difference when accounting for plot curvature is that the KIE will approach unity at both low and high temperatures. It is therefore not surprising that both sugars show a KIE that tends towards 1 at low and high temperatures and this will be the case for all isotope effects on ΔC_p^\ddagger .

Effect of increasing isotopic substitution on ΔC_p^\ddagger .

To explore the relationship between substrate vibrational modes and ΔC_p^\ddagger further, we use glucose to find if there is an isotopic mass dependence on the magnitude of ΔC_p^\ddagger . We have increased the isotopic substitution of glucose using both per-deuterated glucose and in combination with deuterium oxide (D_2O) and monitored the temperature-dependence of k_{cat} as shown in **Figure 41A**. The temperature-dependence of the KIE is shown in Figure 23B with the corresponding labelled structures shown in **Figure 41D**. Given that the *ss*GDH mechanism involves a concomitant hydride and proton transfer (discussed above

and **Figure 38C-G**), we have essentially two possible 1° KIEs for the hydride (1°_H) and proton (1°_P) transfer. However, we note that our combined experimental and computational evidence above suggests that the proton transfer is fast relative to the hydride transfer, indicating that there would be no primary KIE for proton transfer and the experimental assay thus essentially captures the hydride transfer step.

Labelling of sites distal to the transferred hydride are a secondary (2°) KIE. In the present case, this 2° KIE will be composed of many microscopic 2° KIEs for each labelled position (shown in green in **Figure 41D**). Finally, to label the exchangeable OH groups we have performed our experiments in D_2O and this will lead to deuteration of *all* exchangeable protons (including amino acid side chains) that are immediately solvent accessible. Note that the enzyme itself is not incubated in D_2O , only the substrate and so the enzyme deuteration occurs only on the timescale of the steady-state assay (~1 min), see *Materials and Methods*. The resulting kinetic parameters are given in **Table 6** and for the substrate isotope effects.

Table 6. Kinetic parameters extracted from fits of Eq 15 and Eq 16 to the temperature-dependence data shown in **Figure 39A**.

	ΔH^{\ddagger} (kJ mol ⁻¹) ^a	ΔS^{\ddagger} (kJ mol ⁻¹ K ⁻¹) ^a	ΔC_p^{\ddagger} (kJ mol ⁻¹ K ⁻¹)	KIE ^a	KIE _{calc} ^{a,d}	$\Delta\Delta H^{\ddagger}$ (kJ mol ⁻¹ K ⁻¹) ^b	$\Delta\Delta H^{\ddagger}$ (kJ mol ⁻¹ K ⁻¹) ^c
						MMRT	Eyring
D-glucose	76.0 ± 2.3	1.33 ± 0.01	-3.8 ± 0.5	–	–	–	–
D_1	78.0 ± 2.1	1.33 ± 0.01	-1.6 ± 0.5	2.1 ± 0.6	2.8	2.0 ± 4.4	1.6 ± 11.7
D_5	86.1 ± 2.5	1.35 ± 0.01	-0.6 ± 0.5	4.0 ± 1.2	0.9	10.1 ± 4.8	7.2 ± 10.6
D_7	80.4 ± 2.0	1.34 ± 0.01	0.0 ± 0.5	2.4 ± 0.9	2.7	4.4 ± 4.4	3.8 ± 9.5
D_{12}	80.4 ± 1.5	1.33 ± 0.01	0.3 ± 0.4	3.7 ± 1.1	2.3	4.4 ± 4.4	0.9 ± 9.5

^a, Data at 348 K. ^b, from fits to Eq 16 (MMRT) at 348 K. ^c, From fits to Eq 15 (Eyring equation). ^d, Calculated from the QM cluster model.

One expects an increase in mass of the substrate to alter the frequency of the C-H stretch in both the ground and transition state, but whether the expected change still manifests in the presence of the active site amino acids is not obvious. Therefore, we have performed frequency calculations on the stationary points obtained from the above cluster model (**Figure 38C**), for each of the isotopically substituted states shown in **Figure 41** and **S1**. The resulting frequencies are given in **Table S1**. We find a large decrease in stretching frequencies of both the TS and ground state on deuteration of the transferred hydride. On additional isotopic substitution there is generally a small decrease in the C-H stretching

frequency for both the ground state and the TS. Again, one expects small changes in frequency on increasing mass through isotopic substitution and our calculations suggest this expected trend is preserved when the first shell of protein amino acids is also present (as in **Figure 38C**). It is interesting to note that where the amino acids are deuterated (solvent exchangeable positions as with our experiments conducted in D₂O), the frequencies show an additional and significant effect on the calculated frequencies in the order of $\sim 1 \text{ cm}^{-1}$. We would stress that clearly elucidating the relationship between the change in frequencies at the ground/TS, the protein and $\Delta\Delta C_p^\ddagger$ would require very extensive (QM/MM) simulation studies.

There is a significant effect of increased isotope substitution on both the magnitude of the observed rate but also the magnitude of ΔC_p^\ddagger (**Figure 41A** and **Table 6**). From **Figure 41B**, the KIE increases with increasing isotopic substitution, but not in a linear fashion. The absolute magnitude of the KIE depends on the temperature at which the value is reported, the different contributions from substitutions at different positions and the fact that the temperature-dependence of the isotope effects is different for different isotopic substitutions (**Table 6**). Therefore, one does not necessarily expect the values to follow an obvious e.g. linear trend. That said, the maximal KIE does tend to increase with increased isotopic substitution, except in the case of Gluc-D₇ (**Figure 41D**) and we note the relatively large error on these values. As with the 1° KIE, we observe curvature in the magnitude of the KIE for all our isotopic labelling patterns. **Figure S3** shows the resulting curve from both fits of the data to Eq 15 and Eq 16, and the corresponding extracted data are given in **Table 6**.

Given the complex nature of the isotopic labelling pattern for each isotopologue, we do not wish to overinterpret the microscopic contributions to the absolute magnitude of the KIEs. However, it is worth noting that comparison of Gluc-D12 with Gluc-D1 should reveal the combined influence of the secondary substitutions, where a ‘normal’ secondary KIE would be in the region ~ 1.1 . These effects should be additive, i.e. the individual KIEs should be multiplied: $2.1 \pm 0.6 \times 1.1 \times 1.1 \approx 2.5 \pm 0.6$ which differs from the result observed ($\sim 3.7 \pm 1.1$). The large value suggests a significant contribution from secondary effects ($\sim 1.3 - 1.4$). Huskey and Schowen have suggested that large enzyme secondary KIEs reflect strong ‘vibrational coupling’ of the secondary sites to the reaction coordinate at the transition state.³⁰⁴ More recent studies combining density functional theory (DFT) calculations of a model enzyme secondary KIE with high pressure measurements have similar findings.³⁰⁵ We note that the notion of vibrational coupling is itself problematic and we do not suggest that our data reflect such coupling, not least because no study has provided unequivocal evidence for so called vibrational coupling. Given our data we cannot confidently assign the origin of these apparently exalted secondary KIEs.

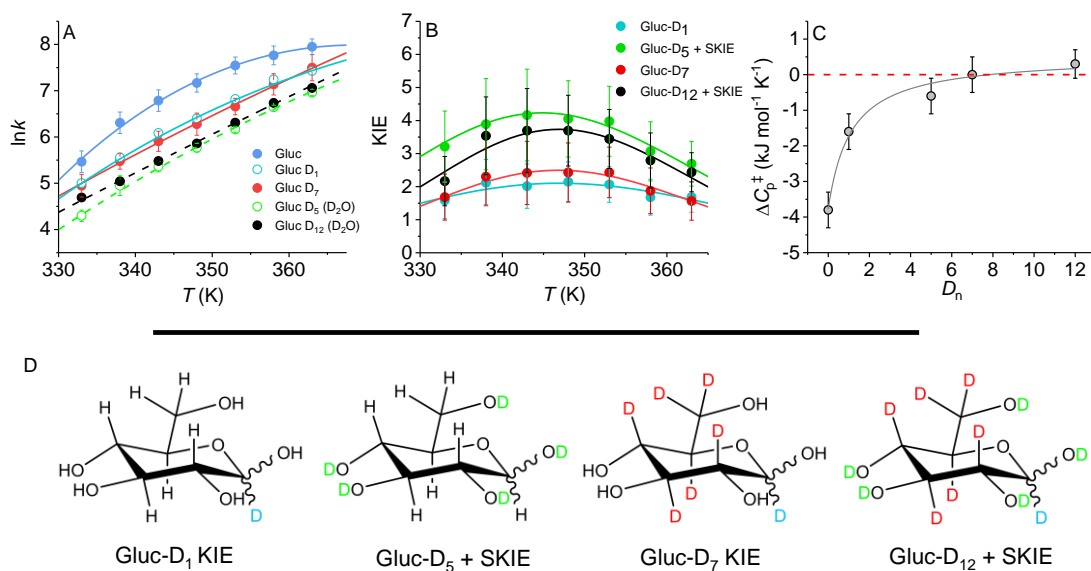


Figure 41: The effect of isotopic labelling on $\Delta\Delta C_p^\ddagger$. **(A)** The temperature-dependence of k_{cat} for each isotopologue of glucose. Solid lines are fits to Eq 16. **(B)** Resulting KIEs extracted from the fits in panel A. **(C)**, correlation between the increase in molecular mass (through isotopic substitution) and the extracted magnitude of $\Delta\Delta C_p^\ddagger$. The solid line is a simple rectangular hyperbola and is to aid the eye only to illustrate the trend. **(D)**, structures of each isotopologue used in the panel A.

The calculated KIEs from our QM cluster model (**Table 6**), suggest an expected secondary KIE of ~ 3.4 ($2.8 \times 1.1 \times 1.1$), which differs from the calculated value of 2.3 (**Table 6**). That is, the secondary KIE value is not as expected for either the experimental measurements or QM calculations. Given the excellent agreement between the experimentally measured and calculated 1° KIE (**Table 6**) the data may indicate that the limitations of the (static) QM cluster model and the importance of enzyme KIEs being calculated using as full a structural model, including conformational sampling, as possible. However, we note that the differences in the absolute magnitude of the experimental KIEs are in fact relatively small, particularly when taking into account the experimental error as shown in **Figure 41B**. As we state above, given this limitation, we prefer a more conservative interpretation of the labelling study, focusing on the increase in isotopic mass rather than the absolute contributions to the KIE from each labelled atom. **Figure 41C** shows the relationship between the extracted $\Delta\Delta C_p^\ddagger$ values (from **Figure 41A**) and the increase in isotopic mass of the substrate. From **Figure 41C**, the isotope effect on $\Delta\Delta C_p^\ddagger$ appears to increase with respect to the mass of the glucose. These data appear to show saturation behaviour, with $\Delta\Delta C_p^\ddagger$ trending towards zero with increasing isotopic mass. Whilst the initial change in $\Delta\Delta C_p^\ddagger$ is relatively large for just one mass unit increase (Gluc-

D₁. $\Delta\Delta C_p^\ddagger = 2.2 \pm 1 \text{ kJ mol}^{-1} \text{ K}^{-1}$), we find large ($\sim 1 \text{ kJ mol}^{-1} \text{ K}^{-1}$) changes in ΔC_p^\ddagger associated with further isotopic substitutions.

Frequency calculations from the QM cluster model indicate these additional increases in mass (D₅-D₁₂) should only change the frequencies of the reacting species (and immediate surrounding) at the ground and transition state by at most a few wavenumbers (**Table S1**). Moreover, the calculated ΔC_p^\ddagger values are 3 orders of magnitude smaller than measured experimentally (**Table S2**). Finally, the calculated KIEs from the QM cluster model show essentially no curvature (**Figure S4**) compared to the clear curvature observed experimentally (**Figure 41B**) and shown by comparative fitting of Eq 15 and Eq 16 in **Figure S3**. A simplistic conclusion from these data would be that our experimental data are not reflecting vibrational frequency changes on isotopic substitution. However, the difference between experimental and computational values is that the QM cluster model neglects nearly all the protein. Recent molecular dynamics simulations that correctly predict significant enzyme ΔC_p^\ddagger values ($\sim \text{kJ mol}^{-1} \text{ K}^{-1}$) have shown that the ΔC_p^\ddagger arises from energetic fluctuations across the whole molecule, including domains distant from the active site.²⁸⁷ It therefore seems likely that the large isotope effect on ΔC_p^\ddagger has a major component arising from changes in protein fluctuations further removed from the active site.

Conclusions

We have monitored the isotope effect on ΔC_p^\ddagger for a hypethermophilic enzyme, finding a very large primary isotope effect on the magnitude of ΔC_p^\ddagger for two different substrates (glucose and xylose). The size of the isotope effect on ΔC_p^\ddagger is very much larger than predicted based on a QM cluster model. Further we illustrate an additive effect of increase the isotopic mass of glucose on the magnitude of ΔC_p^\ddagger . Taken together, our data shows that the change in ΔC_p^\ddagger is coincident with an increase in isotopic mass. These significant changes in ΔC_p^\ddagger ($\sim 1\text{-}2 \text{ kJ mol}^{-1} \text{ K}^{-1}$) are accompanied by only small ($\sim 1 \text{ cm}^{-1}$) changes in vibrational frequency of the reacting species. If the change in ΔC_p^\ddagger arises from these small-scale frequency changes there would need to be some significant ‘amplification’ of the small local effect. Therefore, the key question arising from the mass modulation data presented here is what is behind the isotopic mass dependence on ΔC_p^\ddagger (**Figure 41C**). Specifically, if a significant change in enzymatic ΔC_p^\ddagger (and ΔH^\ddagger ; **Table 6**), on the $\sim \text{kJ mol}^{-1} \text{ K}^{-1}$ scale arises when there are only small changes (on the $\sim 1 \text{ cm}^{-1}$ scale) in the vibrational frequency of the ground and transition states of the reacting species, how might this occur?

A similar conceptual challenge arises from protein mass modulation studies where changing the mass of a protein by isotopic substitution (mass change of $\sim 10\%$) gives rise

to very large changes in the temperature-dependence of the rate (isotope effect on ΔH^\ddagger , $\Delta\Delta H^\ddagger$, changes by ~ 0 -10 kJ mol⁻¹). These very large changes are difficult to rationalize since the expected change in the frequency of protein vibrational modes (for a protein that is ~ 10 % increase in mass) is on the scale of only a few wavenumbers. Previous efforts to interpret such data have suggested that large changes in ΔH^\ddagger might be achieved where protein vibrational modes become ‘decoupled’ from the enzyme catalyzed chemistry.^{297,306,307} Ranasinghe *et al* have recently extended this rationale by suggesting that mass modulation not only affects protein motions coupled to the enzyme catalyzed chemical step, but also the electrostatics associated with longer time-scale events during turnover.³⁰⁸ We note that these works have not considered ΔC_p^\ddagger . Moreover, there have been a significant number of studies that suggest that protein ‘dynamics’ do not affect enzyme catalysis^{286,309-312} and are not in any way coupled to the reaction coordinate.

Our study provides a fresh perspective on current hypotheses that seek to understand the relationship between enzyme vibrational modes and (the temperature-dependence of) catalysis, incorporating a difference in enzyme heat capacity (and thus vibrational modes) along the reaction, ΔC_p^\ddagger . Our data point to a model that links small changes in the vibrational modes of the substrate (or reacting species) to large changes in enzyme fluctuations in different states along the reaction. ΔC_p^\ddagger reflects the change in the *distribution* (and/or magnitude) of protein vibrational modes between the ground and transition state. Therefore, we expect that the apparent disconnect between the scale of substrate isotopic mass changes and the (thermodynamic) heat capacity changes might be resolved by a deeper understanding of the distribution of these protein vibrational modes and how these modes are affected by subtle changes in substrate vibrations. For example, based on our present understanding of the physical origin of ΔC_p^\ddagger , we suggest a physical model where the isotopic changes in the substrate are translated to a shift in the conformational landscape (structural, energetic or both) of the enzyme, resulting in a difference in fluctuations between the reactant and transition state complexes.²⁸⁴

Materials and methods

ssGDH expression and purification.

ssGDH was expressed with AmpR in a pET3a plasmid. It was transformed into BL21 (DE3) *Escherichia coli* using heat shock and grown on LB agar with ampicillin (100 μ g/ml) at 37 °C. A 50 ml LB starter culture was used to inoculate 5 x 1L LB until an OD₆₀₀ of 0.5-0.6 was reached. Cells were harvested by centrifugation (4 °C, 8000 rpm, 10 min) before being lysed by sonication using a lysis buffer (pH 7) containing 100 mM HEPES, lysozyme, DNAase and a protease inhibitor cocktail tablet. Soluble and insoluble fractions were separated by centrifugation at 4 °C (25,000 rpm, 10 min). Due to the thermostability of ssGDH, the soluble fraction was purified by heating the sample to 70

°C for 50 min. To remove precipitated protein, samples were centrifuged (4 °C 13,000 rpm, 10 min) before being dialysed for 4 hours in 100 mM HEPES buffer (pH 7). Samples were further purified/concentrated through the use of Vivaspin centrifugal concentrators (MWCO = 30 kDa). The concentration of purified samples was measured by the absorbance at 280 nm ($\epsilon_{280} = 49,390 \text{ M}^{-1} \text{ cm}^{-1}$), obtained via the input of *ssGDH* amino acid sequence into the ExPASy ProtParam tool.³¹³ Approximately 8 ml of 35 mg/ml purified sample was obtained.

Substrates and coenzymes.

D-glucose, D-xylose, D₂O and NADP⁺ were obtained from Sigma Aldrich. D-glucose (1-D), D-glucose (1,2,3,4,5,6,6-D7) and D-xylose (1-D) were obtained from Cambridge Isotope Laboratories. In this manuscript the varying isotopes and D₂O combinations will be described with the following nomenclature: D-glucose (1-D) – D₁, D-xylose (1-D) – D₁, D-glucose + D₂O – D₅, D-glucose (1,2,3,4,5,6,6-D7) – D₇, D-glucose (1,2,3,4,5,6,6-D7) + D₂O – D₁₂.

Enzyme assays.

Steady-state *ssGDH* kinetic measurements were carried out using a lidded 1 ml quartz cuvette to prevent evaporation at high temperatures and a UV/ Vis spectro-photometer (Agilent Cary 60 UV-Vis spectrometer) in 100 mM HEPES (pH 8). Accurate concentrations of NADP⁺ were determined using NADP $\epsilon_{260} = 17,800 \text{ M}^{-1} \text{ cm}^{-1}$. Enzyme activity was measured for each condition at 85 °C by following the formation of NADPH at 340 nm using $\epsilon_{340} = 6220 \text{ M}^{-1} \text{ cm}^{-1}$ as a direct measurement of *ssGDH* steady-state rates; the data fitted well to Michaelis-Menten kinetics. Temperature-dependences were carried out from 60 °C – 90 °C at 5 °C intervals using initial velocity measurements at substrate concentrations maintained above 10x K_m to ensure saturation. The data were fitted to Eq 15 or Eq 16 as described in the manuscript using OriginPro 2016 (MicroCal). The measured pH for experiments performed in D₂O was adjusted accordingly to match that of the pH in H₂O.³¹⁴

Molecular dynamics simulations.

The *ssGDH* crystal structure 2CDB²⁹⁹ was prepared for simulation using scwrl4³¹⁵ to revert the T41A mutation and modeller³¹⁶ to model in the missing loop at positions 50-59 (based on chain A). To obtain similar starting points for the glucose and xylose complexes, this loop was used for all four chains and coordinates from 2CDB were also used for the xylose complex (where xylose was placed based on alignment with 2CDC²⁹⁹). The Amber16 suite of programs was used for periodic boundary simulation and analysis²²², with the ff14SB force-field for protein atoms⁶⁰, GLYCAM-06j for glucose/xylose⁶³, parameters from Ryde *et al.* for NADP⁺,³¹⁷ TIP3P for water and ZAFF³¹⁸ for the Zn²⁺ coordinated by Cys93, Cys96, Cys99 and Cys107. For the Zn²⁺ directly adjacent to the substrate binding site, restraints were used to maintain the crystallographically observed coordination with Cys39 and His66. After brief minimization of the complex and added water, the system was heated to 300 K and subsequently equilibrated to 1 atm in the NPT ensemble (with positional restraints on C α

atoms). After gradual release of C α positional restraints, 50 ns NPT production simulations were performed at 300 K and 1 atm. Histograms of the D-A distances were calculated over all four binding sites using 10-50 ns of four independent simulations for each substrate. (Further details of model setup, restraints and simulation procedures are included in the Supporting Information.)

QM Cluster Model.

The aforementioned X-ray crystal structure of *ss*GDH in complex with glucose and NADP⁺ (PDB ID 2CDB²⁹⁹) was used to create a 148 atom model of the active site (**Figure S5**). To preserve the overall structure of the active site, several atoms were fixed throughout the optimisation process and care was taken to ensure non-reacting groups stayed in the same local minima throughout the reaction. The T41A mutation was reversed *in silico* with the rotamer selected based on our MD simulations. All Calculations were performed using Gaussian16,³¹⁹ employing the M06-2X functional.³²⁰ Geometry optimisations and frequency calculations were performed *in vacuo* with the 6-31G(d,p) basis set. All models were optimised on an ultrafine integration grid and under tight convergence criteria. Single point energies were obtained using the 6-311++G(2d,2p) basis set, with the surrounding protein environment accounted for using the SMD solvation model with a dielectric constant of 4.³²¹ Thermal corrections to the obtained energies were taken from the aforementioned frequency calculations, employing a scale factor of 0.97.³²² Contributions to tunnelling on the rate of reaction were estimated using Wigner's correction.³²³ Further details about calculation methodology and the coordinates of all stationary points obtained can be found in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information. Molecular dynamics simulations, QM cluster model calculations, supporting data and cartesian coordinates of all stationary points. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Christopher R Pudney, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom. c.r.pudney@bath.ac.uk.

*Marc W van der Kamp, School of Biochemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom. marc.vanderkamp@bristol.ac.uk.

*Vickery L Arcus, School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand. varcus@waikato.ac.nz.

Author Contributions

HBLJ and CM performed experimental work. RMC, MWK and ABT performed computational simulations and calculations. All authors discussed and interpreted data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Δ , these authors contributed equally.

ABBREVIATIONS

D-glucose (1-D), D₁ ; D-xylose (1-D), D₁ ; D-glucose + D₂O, D₅ ; D-glucose (1,2,3,4,5,6,6-D₇) , D₇ ; D-glucose (1,2,3,4,5,6,6-D₇) + D₂O, D₁₂.

ACKNOWLEDGMENT

HBLJ's studentship is funded by the University of Bath. RMC's studentship is funded by the EPSRC. MWvdK is a BBSRC David Phillips Fellow (BB/M026280/1). We thank Katarzyna Świderek (Universitat Jaume I), Ian H. Williams (University of Bath) and Vicent Moliner (Universitat Jaume I) for helpful discussions and critical review of the manuscript.

Supporting information

Uncovering the relationship between the change in heat capacity for enzyme catalysis and vibrational frequency through isotope effect studies

Hannah BL Jones^{†Δ}, Rory M Crean^{#†Δ}, Christopher Matthews[†], Anna B Troya[†], Michael J Danson[†], Steven D Bull[‡], Vickery L Arcus^{§*}, Marc W Van der Kamp^{%*} and Christopher R Pudney^{†*}.

[†]Department of Biology and Biochemistry, [‡]Department of Chemistry, # Doctoral Training Centre in Sustainable Chemical Technologies, University of Bath, Bath BA2 7AY, United Kingdom

[§]School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand

[%] School of Biochemistry, University of Bristol, Biomedical Sciences building, University Walk, Bristol BS8 1TD, United Kingdom

*Christopher R Pudney, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom. c.r.pudney@bath.ac.uk.

*Marc W van der Kamp, School of Biochemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom. marc.vanderkamp@bristol.ac.uk.

*Vickery L Arcus, School of Science, Faculty of Science and Engineering, University of Waikato, Hamilton 3240, New Zealand. varcus@waikato.ac.nz.

Author Contributions

HBLJ and CM performed experimental work. RMC, MWK and ABT performed computational simulations and calculations. All authors discussed and interpreted data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. Δ, these authors contributed equally.

Supporting Information Materials and Methods

Molecular dynamics simulations: Model setup, restraints and simulation details

The ssGDH crystal structure 2CDB²⁹⁹ was prepared for simulation using scwrl4³¹⁵ to revert the T41A mutation and modeller³²⁴ to model in the missing loop at positions 50-59, based on the coordinates of the surrounding residues in chain A. The same loop conformation was subsequently transferred to the other three chains. To avoid a clash with this loop in chain C, conformer B (assigned 50% occupancy) was chosen for His118. (If not mentioned, conformer A was chosen for other residues with multiple conformers defined, including for the glucose O6.) Asn/Gln flips and His tautomers and were selected based on optimal hydrogen bonding contacts through the AmberTools facility reduce, with His66 and His319 singly protonated on Nδ1, others on Nε2. The exception was His297, which was protonated on Nδ1: test simulations indicated that this tautomer better retained the crystallographically determined enzyme-ligand conformation structure. All ionizable residues (with exception of the Cys residues coordinating Zn²⁺, see below) were modelled in their standard protonation states (in agreement with pK_a predictions by PropKa 3.1)¹⁵⁸. To obtain similar starting points for the glucose and xylose complexes, the same protein starting model was used for the xylose complex (where xylose was placed based on alignment with 2CDC²⁹⁹).

The Amber16 suite of programs was used for periodic boundary simulations and analysis,²²² with the ff14SB force-field for protein atoms,⁶⁰ GLYCAM-06j for glucose/xylose,⁶³ parameters from Ryde *et al.* for NADP+,³¹⁷ TIP3P for water and ZAFF³¹⁸ for the Zn²⁺ coordinated by Cys93, Cys96, Cys99 and Cys107. These Cys residues were thus modelled as thiolates (CYM). For the Zn²⁺ directly adjacent to the substrate binding site, the default Amber ion parameters were used. To avoid coordination changes around the Zn²⁺ ion, the following one-sided harmonic restraints were applied if: 1) the His66 NE2–Zn²⁺ distance was larger than 2 Å (force constant 70 kcal mol⁻¹ Å⁻²); 2) the Asp42 OD2–Cys39 N distance was larger than 1.95 Å (force constant 70 kcal mol⁻¹ Å⁻²); 3) the Asp42 OD2–Zn²⁺ distance was smaller than 4.2 Å (force constant 100 kcal mol⁻¹ Å⁻²).

In addition to the water molecules present in the crystal structure, a rectangular box of water was added using the AmberTools facility tleap such that all protein atoms were at least 11.5 Å away from the edge of the box (closeness parameter of 0.9). 20 water molecules (at least 5 Å away from the protein) were replaced with Na⁺ ions to neutralize the system. After minimization of solvent and hydrogens (100 steps steepest descent, 400 steps conjugate gradient), the full system was briefly minimized (400 steps) with positional restraints of 5 kcal mol⁻¹ Å⁻² on all Cα atoms. Maintaining these positional restraints on Cα atoms, velocities were assigned at 25 K and the system was heated to 300 K in 50 ps using Langevin temperature control in the NVT ensemble (with a 1 ps⁻¹ collision frequency). Subsequently, 100 ps of equilibration was performed in the NPT ensemble at 1 atm, using the Berendsen barostat (1 ps pressure relaxation time) and Langevin dynamics for temperature control (5 ps⁻¹ collision frequency). Thereafter, the

positional restraints on C α atoms were gradually removed in four consecutive 10 ps simulations under the same conditions (4, 3, 2, and 1 kcal mol⁻¹ Å⁻² as force constants). Production MD (without positional restraints) was then performed in the NPT ensemble at 1 atm and 300 K for 50 ns, saving coordinates every 10 ps. All MD simulations were performed with the default direct-space cut-off for non-bonded interactions and particle-mesh Ewald summation for long-range electrostatics, run with pmemd.cuda. Four independent simulations were run for each complex (GDH-glucose and GDH-xylose).

QM Cluster Model Calculations: Model Construction, pKa and KIE Calculations

The X-ray crystal structure of *ss*GDH in complex with glucose and NADP⁺ (2CDB)²⁹⁹ was used to create a 148 atom model of the active site. Truncations of the enzyme active site were made across non-polar bonds. All titratable residues included in the model were in their standard protonation state. The T41A mutation used to crystallise *ss*GDH was reverted in silico, with the rotamer (side chain alcohol group acting as a hydrogen bond acceptor to the alcohol group of the glucose C1) chosen based on its prevalence in our MD simulations. To preserve the overall structure of the active site, several atoms were fixed throughout the optimisation process and care was taken to ensure non-reacting groups stayed in the same local minima throughout the reaction.

All Calculations were performed using Gaussian16,³¹⁹ employing the M06-2X functional.³²⁰ Geometry optimisations and frequency calculations were performed in vacuo with the 6-31G(d,p) basis set. All models were optimised on an ultrafine integration grid and under tight convergence criteria. Single point energies were obtained using the 6-311++G(2d,2p) basis set, with the surrounding protein environment accounted for using the SMD solvation model with a dielectric constant of 4.³²¹ Thermal corrections to the obtained energies were taken from the aforementioned frequency calculations, employing a scale factor of 0.97.³²² Contributions to tunnelling on the rate of reaction were estimated using Wigner's correction.³²³

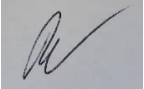
In order to simulate the deprotonation of the Zn coordinated water to a Zn coordinated hydroxide by the bulk milieu (reactant to reactive intermediate), a proton on the Zn coordinated water (reactant model) was removed and the structure re-optimised, giving the reactive intermediate. During optimisation, the Zn coordinated hydroxide abstracted the sugar C1 alcohol group's proton, forming an alkoxide on the sugar. This therefore led us to propose the stepwise mechanism as seen in Scheme S1. To calculate the barrier for the transition from reactant to reactive intermediate, the experimentally determined barrier at 298 K (73.5 kJ mol⁻¹) was subtracted from the computationally determined hydride transfer barrier (32.6 kJ mol⁻¹), with the subsequent energy difference used to predict a pKa of 7.2.

All polar hydrogen atoms in the cluster model were considered deuterated when modelling the SKIE (see **Figure S5**). Heat capacities of activation (**Table S2**) were calculated from the differences in constant volume heat capacity for the reactive intermediate and transition state models.

PLEASE NOTE: In the interest of saving paper/space, only the Supporting Information Methods are included and not the figures or tables. They can instead be found online free of charge at: <https://pubs.acs.org/doi/abs/10.1021/acscatal.8b01025>

Appendix Chapter 2: Probing the Importance of the Environment in Regulating Functionally Important Enzyme Dynamics

In this second appendix chapter we applied a combination of computational and experimental approaches to study the biomedically relevant enzyme monoamine oxidase B (MAO-B). The primary aim of this study was to understand how the environment can perturb/tune the functionally important dynamics of this enzyme. This was first studied experimentally with enzyme temperature dependence studies, which showed how the use of different substrates and/or a different membrane environment were able to significantly modify the heat capacity of catalysis (ΔC_p^\ddagger), suggesting both the substrate and membrane are able to tune functionally important dynamics linked to catalysis. We complemented our experimental approach with molecular dynamics (MD) simulations of MAO-B in a lipid bilayer. Our MD simulations were able to identify and characterise a new entrance for substrate/ligand binding, which is of particular relevance considering the that MAO-B can be inhibited for the treatment of both Parkinson's disease and depression. Furthermore, our MD simulations with substrate bound suggested the possibility of half-site reactivity. Finally, we used MD simulations in an attempt to understand how the previously described distal mutations removed activity from MAO-B. We demonstrated that that dynamics of the loop which control substrate/ligand entry were significantly altered in all of the variants as compared to the wild-type, suggesting substrate binding/product release was disrupted in the variants. My role in this project was to lead the setup, running and analysis of the MD simulations of MAO-B, with this aspect of the project performed in collaboration with Hannah Jones.

This declaration concerns the article entitled:			
Exposing the Interplay Between Enzyme Turnover, Protein Dynamics, and the Membrane Environment in Monoamine Oxidase B.			
Publication status (tick one)			
draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>
	<input type="checkbox"/>	In review	<input type="checkbox"/>
	<input type="checkbox"/>	Accepted	<input type="checkbox"/>
	<input type="checkbox"/>	Published	<input checked="" type="checkbox"/>
Publication details (reference)	<p>Jones, H. B. L. *, <u>Crean, R. M.</u> *, Mullen, A., Kendrick, E. G., Bull, S. D., Wells, S. A., ... Pudney, C. R. (2019). Exposing the Interplay Between Enzyme Turnover, Protein Dynamics, and the Membrane Environment in Monoamine Oxidase B. <i>Biochemistry</i>, 58(18), 2362–2372. https://doi.org/10.1021/acs.biochem.9b00213</p> <p>* = Equal contribution.</p>		
Candidate's contribution to the paper (detailed, and also given as a percentage).	<p>The candidate contributed to/ considerably contributed to/predominantly executed the...</p> <p><u>Formulation of ideas (40 %)</u> Computer simulation experiment ideas was formulated by RMC, MvdK & HBLJ. Laboratory experiment ideas were formulated by HBLJ and CRP.</p> <p><u>Design of methodology: (40 %)</u> Methodology for MD simulations and analysis was formulated by RMC, HBLJ and MvdK. Experimental methodology was formulated by HBLJ, CRP, AM, FM.</p> <p><u>Experimental work: (35 %)</u> MD simulations were carried out by RC and HBLJ. All MD simulation analysis except Caver was carried out by RC. Caver analysis was carried out by HBLJ and RC. FIRST calculations by HBLJ. EPR experiments were carried out by AM and HBLJ. All other experiments carried out by HBLJ.</p> <p><u>Presentation of data in journal format (35 %)</u> Main text by HBLJ, CRP, RC and MvdK SI formulation by RC. All authors edited and reviewed manuscript.</p>		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		Date	31/07/2019

Exposing the Interplay Between Enzyme Turnover, Protein Dynamics and the Membrane Environment in Monoamine Oxidase B

Hannah B. L. Jones^{†||}, Rory M. Crean^{#†||}, Anna Mullen[§], Emanuele G. Kendrick[†], Steven D Bull^Δ, Stephen A. Wells[‡], David R. Carbery^Δ, Fraser MacMillan[§], Marc W. Van der Kamp^{*%}, Christopher R. Pudney^{*†|}.

[†]Department of Biology and Biochemistry, ^ΔDepartment of Chemistry, [‡]Department of Chemical Engineering University of Bath, [#]Doctoral Training Centre in Sustainable Chemical Technologies, [|]Centre for Therapeutic Innovation, University of Bath, Bath BA2 7AY, United Kingdom.

[%]School of Biochemistry, University of Bristol, Biomedical Sciences building, University Walk, Bristol BS8 1TD, United Kingdom.

[§]School of Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom.

Abstract: There is an increasing realization that structure-based drug design may show improved success rates by understanding the ensemble of conformations and sub-states accessible to an enzyme and how the environment affects this ensemble. Human monoamine oxidase B (MAO-B) catalyzes the oxidation of amines and is inhibited for the treatment of both Parkinson's disease and depression. Despite its clinical importance, its catalytic mechanism remains unclear and routes to drugging this target would be valuable and relevant. Evidence of a radical in either the transition state or resting state of MAO-B is present throughout the literature, and is suggested to be a flavin semiquinone, a tyrosyl radical or both. Here we see evidence of a resting state flavin semiquinone, via absorption redox studies and electron paramagnetic resonance, suggesting that the anionic semiquinone is biologically relevant. Based on enzyme kinetic studies, enzyme variants and molecular dynamics simulations we find evidence for the crucial importance of the membrane environment in mediating the activity of MAO-B and that this mediation is related to effects on the protein dynamics of MAO-B. Further, our MD simulations identify a hitherto undescribed entrance for substrate binding, membrane modulated substrate access, and indications for half-site reactivity: only one active site is accessible to binding at a time. Our study combines both experimental and computational evidence to illustrate the subtle interplay between enzyme activity, protein dynamics and the immediate membrane environment. Understanding key biomedical enzymes to this level of detail will be crucial to inform strategies (and binding sites) for rational drug design for these drug targets.

KEYWORDS *EPR, Molecular Dynamics, Monoamine Oxidase B, semiquinone, enzyme, flavin, membrane*

Human monoamine Oxidase B (MAO-B) catalyzes the oxidative deamination of amines, by electron transfer, via its flavin adenine dinucleotide (FAD) cofactor (**Scheme 1**).³²⁵ It

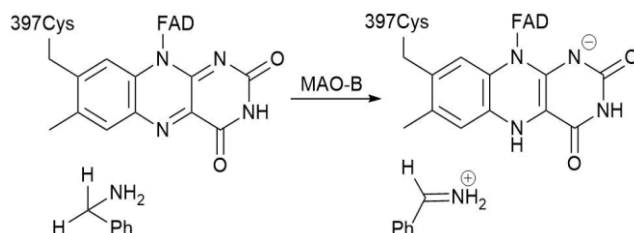
is located in the outer mitochondrial membrane,³²⁶ as a dimer, with bipartite substrate binding and active site cavities.³²⁵ MAO-B is the target of treatment for both depression and Parkinson's disease, with inhibitors of the enzyme first being approved as pharmaceuticals in the 1960's.^{327,328}

Despite the important medical applications associated with MAO-B its chemical mechanism remains unclear and there is debate over the role of MAO-B conformational change and protein dynamics. Reduction of the FAD has been shown to proceed by a tunneling mechanism via primary (1°) and secondary (2°) kinetic isotope effect (KIE) studies.³²⁹ These studies have suggested that MAO-B catalysis is not linked to fast (pico/nanosecond) dynamics.^{329,330} MAO-B catalysis has been investigated via experimental and computational studies, with at least seven different proposed mechanisms, including polar nucleophilic,³³¹ radical,³³² direct hydride transfer^{333–335} and two step hydride transfer³³⁶ (**Scheme S2**).

Although the direct single electron transfer (SET) radical mechanism (**Scheme S2**) has previously been discounted experimentally³³⁷ and by quantum mechanics/molecular mechanics (QM/MM),³³⁸ a separate radical mechanism has been proposed by Murray *et al.*³³⁹ This was established via a model small molecule reaction that mimics MAO-B to provide evidence for a neutral semiquinone flavin that can be formed aerobically. The authors suggest that a neutral semiquinone flavin is the reactive species for the oxidation of benzylamine (BZA).³³⁹ This led to the hypothesis of a radical mechanism whereby MAO-B forms a neutral semiquinone flavin via a proximal tyrosine radical (Y398). The presence of a stable anionic semiquinone flavin and tyrosyl radical intermediate are also reported in MAO-A and MAO-N.^{340,341}

In our hands, we find spectroscopic evidence for a stable semiquinone in resting state MAO-B (*vide infra*). Here, MAO-B was expressed and purified in *Pichia pastoris* as outlined by Newton-Vinson *et al.*³⁴² This protocol attributes observations of oxidized/semiquinone MAO-B FAD to reactive oxygen species (ROS) that form upon the disruption of the mitochondrial membrane, which they observe upon purification from bovine liver, but not from *P. pastoris*.³⁴²

Previous observation of a stable anionic semiquinone FAD in MAO-B^{343,344} led to a hypothesis of half-site reactivity. This mechanism posits that one monomer of the MAO-B dimer is inaccessible to oxygen and substrate, resulting in the formation of the stable semiquinone species, whilst the other contains oxidized FAD. Electrons are then shuttled



Scheme 1: General reaction catalyzed by MAO-B.

to the semiquinone species, allowing for the oxidation of the reduced FAD upon turnover.³⁴⁴ The authors suggest this might arise from electron shuttling between specific amino acids. Potentially, such a mechanism might require significant conformational change associated with turnover in each monomer. Indeed, conformational changes have been found to be associated with MAO-B turnover, with a molecular dynamics (MD) study demonstrating the potential for the membrane to regulate access to the active site entrance via two gating loops (residues 85-119 and 155-165).³⁴⁵ Other MD studies have been carried out on MAO-B, both with and without a membrane environment, with the focus on identifying or improving inhibitors for MAO-B³⁴⁶⁻³⁵⁰, or ascertaining how MAO-B binds to the membrane.³⁵¹

Herein, we find evidence for a resting-state anionic semiquinone, and through kinetic and computational studies, evidence for conformationally controlled enzyme activity at each MAO-B monomer. Crucially, we find that the membrane environment exposes novel substrate/product channels that could be potential new drug targets. We thereby link the membrane environment, substrate binding and MAO-B conformational change to enzyme turnover.

Results and Discussion

Evidence for a resting-state semiquinone in MAO-B.

The presence of purified MAO-B was established by SDS-PAGE (**Figure S6**; essentially a single band via size exclusion chromatography) and via electrospray ionization quadrupole time-of-flight (ESI-Q TOF) mass spectrometry in combination with the MASCOT server.³⁵² The absorption spectrum of purified MAO-B is shown in **Figure 42A**. The spectra share characteristics of an anionic semiquinone FAD (FAD⁻) with an absorption feature at ~415 nm.³⁵³ The preparation protocol of MAO-B was the same as that used by Newton-Vinson *et al.*,³⁴² with small differences outlined in the *Materials and Methods*. Multiple preparations were completed, including the final polymer partition step outlined by Newton-Vinson *et al.* and without EDTA present, with the ~415 nm spectral feature consistently present. The effect of incubation with BZA, under anaerobic conditions, on the MAO-B absorption feature at ~415 nm was monitored over time (**Figure 42A inset**). From **Figure 42A inset**, we find a decrease in absorption at 415nm with respect to time on incubation with BZA, suggesting the ~415 nm spectral feature is redox sensitive with a natural MAO-B substrate.

We note that the absorption spectrum shown in **Figure 42A** lacks a defined feature around ~450 nm where one would expect oxidized flavin as well as spectral features at ~475 nm that would also characterize a putative anionic semiquinone. We do not have a clear explanation for the lack of these features and the complexity of the absorption spectrum; the protein is purified to homogeneity, it is identified as MAO-B by mass spectrometry and is catalytically active with the natural substrates (**Figure 44A**, **Figure S7** and as described below). The electronic environment around both flavins is highly complex, with a large number of proximal Tyr residues (Y60, 398 and 435 positioned 3.1, 3.3 and 3.5 Å

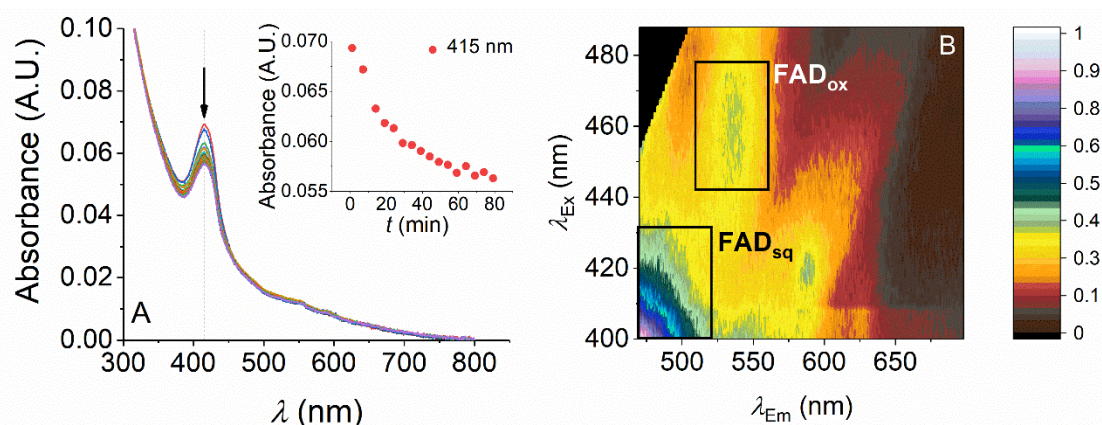


Figure 42: (A) The absorbance spectra of MAO-B after treatment with BZA over time. Inset, the effect on the 415 nm peak over time. (B) Fluorescence excitation/emission matrix resolves oxidized and semiquinone flavin states (highlighted in solid black boxes). Scale bar is relative intensity. Conditions: 50 mM HEPES 0.5 % w/v Triton X-100, 20 °C. For absorption experiments 40 mM BZA, anaerobic conditions as Materials and Methods.

from the alloxan moiety, respectively). Potentially, this gives rise to a complex absorption spectrum arising from different electronic environments for a sub-set of active site conformational states. This hypothesis requires MAO-B to be able to explore different conformational states and we consider this in more detail below. Given the complexity of the absorption spectrum, we turned to fluorescence and electron paramagnetic resonance (EPR) spectroscopy to provide more specific evidence for the oxidation state(s) of the MAO-B flavin.

Previously, MAO-B steady-state fluorescence spectroscopy has demonstrated that there are two different chromophores present in resting state MAO-B.³⁵⁴ The authors concluded that these two different species were consistent with oxidized and semiquinone flavin. To establish if we similarly have both oxidized and semiquinone flavin present (which is not obvious from the absorption spectrum), we have monitored the fluorescence excitation-emission matrix (**Figure 42B**). Similar to the previous report, the spectra resolve at least two different emission peaks suggesting the presence of two different oxidation states of the flavin with $\lambda_{\text{Ex}} \sim 400\text{-}420$ and 460 nm. Notably, the excitation/emission profile at $\lambda_{\text{Ex}} \sim 460$ and $\lambda_{\text{Em}} \sim 540$ nm (**Figure 42B**) indicates the presence of oxidized flavin.

The ~ 415 nm absorption feature could be potentially attributed to a tyrosyl radical, which has a characteristic absorbance peak at 410 nm,³⁵⁵ or a mixture of both an FAD semiquinone and tyrosyl radical, as seen as intermediates in MAO-A.^{340,341} Murray *et al.* postulated that the reactive semiquinone FAD was formed by a proximal tyrosyl radical (Y398). However, the UV-vis absorption spectrum of Y398F MAO-B also shows the spectral feature at ~ 415 nm (**Figure S8A**), indicating that this feature is not attributable

to a Y398 tyrosyl radical. The Y398F variant also shows a slight increase in K_m (**Figure S9**), similar in magnitude to previously reported changes in K_m for Y398F.³⁵⁶

A stringent approach to identifying the nature of the flavin oxidation state is EPR. The EPR X-band spectrum of the wild type MAO-B (*wt*MAO-B) clearly indicates that the signal arises from a semiquinone radical (**Figure 43**), in agreement with the UV-Vis and fluorescence spectroscopy (**Figure 42**). The measured spectra lack the defining features which would identify the signal as arising from a tyrosyl radical: the typical ‘wings’ or ‘shoulders’ around the central signal at around $g = 2$ ^{340,357} at X-band (**Figure 43**), and increased g -value anisotropy at higher frequency (and therefore resonant field), i.e. Q-band (data not shown). Furthermore, the signal persists in the Y398F variant, confirming that it is not caused by the proximal tyrosine. Further, computational simulation and fitting of the experimental X-band data of the WT and Y398F MAO-B (**Table 7**) suggests that the semiquinone radical species is anionic; the hyperfine environment of a neutral semiquinone radical would contain a contribution from an additional hydrogen atom.^{344,358} This is not the case for the signals seen in the X-band spectra.

Previous studies have illustrated the importance of the membrane environment in mediating the normal enzymatic activity of MAO-B.³⁴⁵ To probe if the putative semiquinoid species (inferred from spectroscopic studies above) was also stably present in the membrane environment, we have conducted spectroscopic studies in an artificial membrane environment using 1- α -phosphatidylcholine styrene maleic acid co-polymer (SMA) nanodiscs, prepared as reported previously.³⁵⁹ We find that the absorption feature

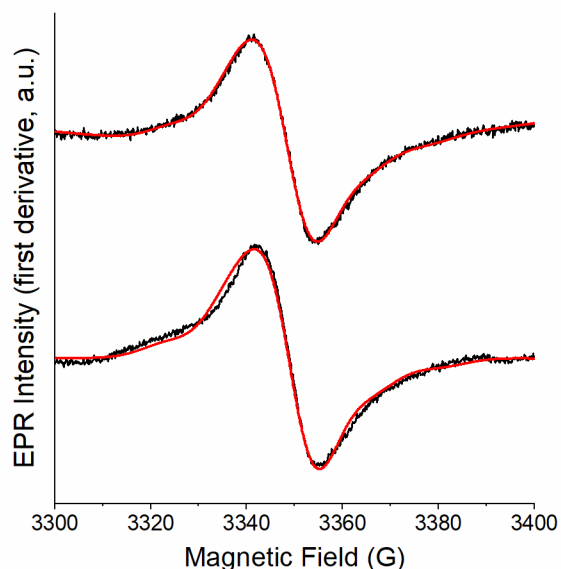


Figure 43: X-band cw-EPR spectra of WT (top, black) and Y398F (bottom, black) MAO-B, with their respective fitted simulations (red). EPR microwave frequency = 9.3916 GHz (WT) & 9.3926 GHz (Y398F), microwave power = 0.2 mW, modulation amplitude = 0.5 mT, temperature = 16 K.

at ~415 nm is present in both reduced Triton X-100 and nanodisc environments (**Figure S8A**), implying that the putative anionic semiquinone is not an artifact of the buffer system used, and that its presence is not affected by the specific membrane environment used. It is clear from these data (**Figure 42** and **Figure 43**) that the FAD of resting state MAO-B is able to stably occupy oxidation states other than fully oxidized FAD. Evidence for this is recurring in the literature, demonstrated by Raman³⁴³, fluorescence³⁵⁴, and EPR^{342,344} spectroscopy. We suggest that these different observations could be rationalized by an equilibrium of energetically similar conformational states, which allow differently stable oxidation states. Indeed, below we provide evidence that shows MAO-B is able to sample a range of conformational states.

Table 7. Spectral Parameters of EPR Data Extracted by Simulation and Fitting.

		WT MAO-B	Y398F MAO-B
g-tensor	g_x	2.00444	2.00444
	g_y	2.00429	2.00429
	g_z	2.00191	2.00191
	g_{iso}	2.00355	2.00355
A(¹⁴N) (MHz)	$A_{ }$	39.3	38.4
	A_{\perp}	0	0
A(¹⁴N) (MHz)	$A_{ }$	34.1	40.1
	A_{\perp}	0	0
Linewidth (mT)		1.1	1.1
RMSD		0.0255	0.0440

Influence of the membrane environment on MAO-B turnover.

To assess how/if the membrane environment affects enzyme turnover, we measured MAO-B turnover with both BZA and kynuramine dihydrobromide (KYN) at a range of temperatures. We monitor enzyme turnover based on the absorption features of benzaldehyde product formation at 250 nm for BZA,³³⁷ and 4-hydroxyquinoline product formation at 316 nm for KYN.³⁶⁰ This assay notionally primarily reflects the rate of reduction of the flavin.³³⁷ Our temperature-dependence studies allow us to analyze not just the observed rate of enzyme turnover but also the thermodynamics of the system. The temperature-dependence of the observed rate is shown in **Figure 44**.

There has been a recent move to fitting enzyme temperature-dependence data to physical models that allow for curvature in the associated plots. Such models often provide a more realistic fit to experimental data. We fit the MAO-B temperature-dependence data to a model that incorporates the heat capacity of catalysis (ΔC_p^\ddagger) Eq 16 into the Eyring equation (Eq 15).²⁹⁰

$$k = (k_B T/h) e^{\Delta S^\ddagger/R} e^{-\Delta H^\ddagger/RT} \quad (15)$$

$$\ln k = \ln \frac{k_B T}{h} - \left[\frac{\Delta H_{T_R}^\ddagger + \Delta C_p^\ddagger (T - T_R)}{RT} \right] + \left[\frac{\Delta S_{T_R}^\ddagger + \Delta C_p^\ddagger (\ln T - \ln T_R)}{R} \right] \quad (16)$$

Where ΔH^\ddagger is the change in enthalpy and ΔS^\ddagger is the change in entropy between the ground and transition state of the reaction at an arbitrary reference temperature (T_0). This model has recently gained traction in studying enzyme temperature-dependencies.^{287,288,297,361} In the absence of other confounding factors, ΔC_p^\ddagger quantifies the temperature-dependence of ΔH^\ddagger and ΔS^\ddagger and reflects the difference in the distribution and frequency of vibrational modes between the ground state and transition state.^{290,291} We have recently suggested that ΔC_p^\ddagger can be used as a proxy for the changes in these vibrational modes during enzyme turnover and thus relates to some aspect of the protein's molecular dynamics.²⁸⁷ This is relevant in the present study where the membrane environment may not alter the tertiary structure of the enzyme, but potentially alters protein fluctuations, which have previously been proposed to affect small molecule binding to the active site.³⁴⁵

From **Figure 44** and **Table S3**, we find ΔC_p^\ddagger to be the same within error for both KYN and BZA substrates when in a reduced Triton X-100 environment. However, when in nanodiscs, the difference in ΔC_p^\ddagger for the different substrates is $\Delta \Delta C_p^\ddagger = 2.4 \pm 1.0$ kJ mol⁻¹. The ΔC_p^\ddagger increases in magnitude from reduced Triton to nanodiscs with KYN, and decreases with BZA (**Figure 44**). These data suggest that the difference in conformational fluctuations in the reactant and transition states is different in a more native membrane environment and for different substrates. Moreover, we find a significant difference in the observed rate of enzyme turnover in the nanodiscs (~5 times faster). These data therefore provide experimental evidence of the notion that the membrane environment has a role in 'tuning' the molecular dynamics of MAO-B. However, given we observe a retention of the putative anionic flavin semiquinone, we would suggest the membrane does not affect the chemical mechanism of enzyme turnover *per se*.

Computational evidence for a new entrance to the MAO-B active site mediated by the protein-membrane interaction.

To obtain detailed insight into the role of the membrane environment on protein dynamics, we performed MD simulations of the MAO-B dimer embedded in a phospholipid membrane. We explore (i) the influence of the phospholipid membrane in modulating substrate/inhibitor binding, (ii) the accessibility of small molecules to the active site of MAO-B, and (iii) the potential for half-site reactivity, as discussed above.

MD simulations of MAO-B in complex with FAD in a POPC/POPE lipid bilayer (similar to the native environment in the mitochondrial membrane)³⁶² were performed in triplicate

for 150 ns in 3 different states: no BZA present (BZA_0), one active site occupied with BZA (BZA_1), and both active sites occupied with BZA (BZA_2). Protein C α RMSD and area per lipid head groups (**Figure S10**) indicated equilibration of both the protein and membrane after 50 ns of production MD. The analysis described below is therefore from 50 to 150 ns.

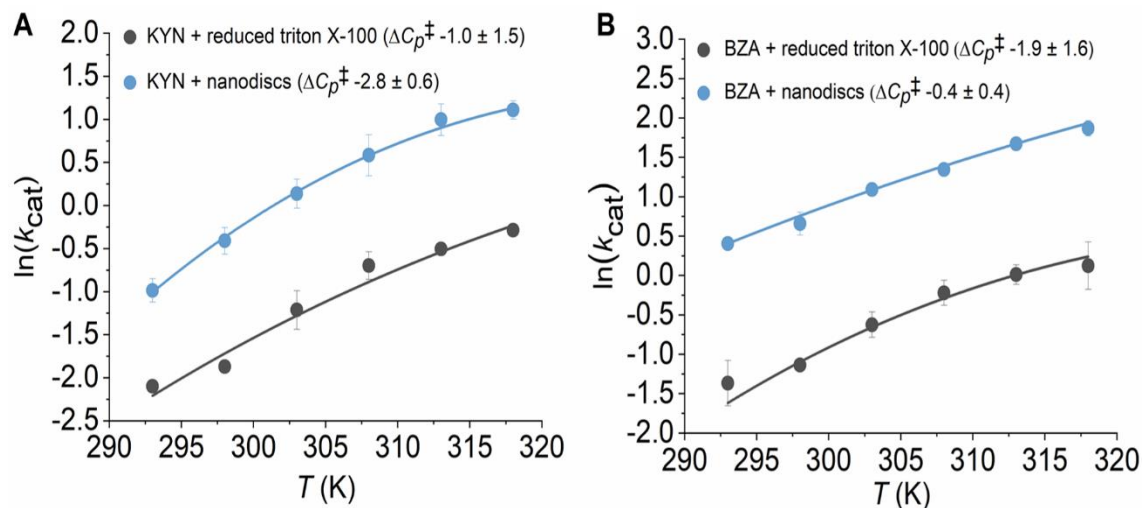


Figure 44: The temperature-dependence of MAO-B with BZA (**A**) and KYN (**B**) with reduced Triton and nanodisc environments, fit to the MMRT equation. Conditions, 50 mM HEPES 0.5 % w/v reduced Triton X-100 pH 7.5. Reduced Triton X-100: 1.5 mM BZA, 0.75 mM KYN. Nanodiscs: 3 mM BZA, 0.66 mM KYN.

Prior simulations of BZA_0 -MAO-B in a bilayer established that MAO-B ligand binding site access is modulated by the membrane.³⁴⁵ We investigate this further using longer simulations (150 ns vs. 50 ns) and with substrate (BZA) bound. To measure possible access to the substrate cavity of MAO-B via the membrane, we quantify the occurrence and features of tunnels in our simulations using Caver 3.0.³⁶³ Tunnels identified are grouped into clusters, allowing for the quantification of various characteristics, such as the frequency of occurrence and smallest width (bottleneck radius), as used here.³⁶⁴ This tool has previously been used to identify tunnels for ligand-induced protein flexibility analysis,³⁶⁵ to rationalize change in mechanism and kinetics of an enzyme upon a point mutation,³⁶⁶ and to identify a tunnel to the FAD moiety in MAO-A.³⁶⁷

Two main possible entrances for ligands into the MAO-B active site are found at either side of the gating loop residues 99 to 112 (**Figure 45**). Entrance A (**Figure 45A**) is accessed via the membrane and its opening has previously been observed.³⁴⁵ Briefly, a π - π stacking interaction between Tyr97 and Trp107 is lost as Trp107 buries into the aliphatic lipid tails of the bilayer, establishing an additional interaction of MAO-B with the phospholipid bilayer.

Entrance B (**Figure 45B**) is solvent accessible, and its opening is controlled by the conformation of three loop regions (81 to 88, 99 to 112 and 198 to 208). To the best of our knowledge, its opening has not yet been observed through protein crystallography or

simulation. To provide a qualitative description of the open and closed conformations of this entrance, we have performed clustering analysis on the entrance loop residues (see *Materials and Methods*). Opening of Entrance B (**Figure S11**) can be described by: 1) loop 81-88 separates from loop 199-206, breaking a number of transiently formed electrostatic interactions and instead forming interactions with the solvent in the open conformation; 2) the central region of the gating loop (residues P102, F103, P104) rotates down and away from the hydrophobic core of the entrance cavity. The opening of both entrance A and B involves residues directly interacting with the bilayer. This indicates that the membrane is important in modulating access to the substrate binding pocket of MAO-B.

Evidence of asymmetry in MAO-B from MD simulation.

Previous studies have suggested that the presence of an anionic semiquinone could be mechanistically significant as part of a half-site reactivity mechanism.³⁴⁴ To investigate the potential for half-site reactivity in MAO-B, tunnels from the N5 of the flavin to the solvent were first identified from all MD trajectories with BZA₀, BZA₁ and BZA₂ (with three replicates). These data are useful to determine the size of species that could access the active site and to investigate whether access to the active site is half-site specific. To avoid identifying numerous tunnels that cannot accommodate substrate/product molecules, a minimum tunnel radius was set to 1.5 Å. (This avoids identifying water tunnels, as water has a Van der Waals radius of ~1.4 Å.³⁶⁸) With this criterion, no tunnels were found, which demonstrates a closed active-site on the timescales of our simulations. Whilst larger conformational changes may be occurring on longer time-scales (or in presence of substrate in the vicinity of a bottleneck) to enable substrate access to the active site, a generally closed off active site is in agreement with previous experimental findings.³⁴⁴

Next, tunnels were identified from the substrate binding cavity towards the protein surface. The starting point was defined as the centre of mass between the alpha carbon (C α) atoms of residues 168 and 316 (**Figure S12**). This starting point was chosen as it is located within the substrate binding cavity and is common to both previously described substrate entrance tunnels (**Figure 45**). Tunnel frequency and the average bottleneck radius of all tunnels identified in both monomers were obtained for BZA₀ and BZA₁ (**Figure 46A+B**). To aid discussion, we designate each monomer as monomer 1 or monomer 2, noting that this does not imply a structural or other difference between monomers.

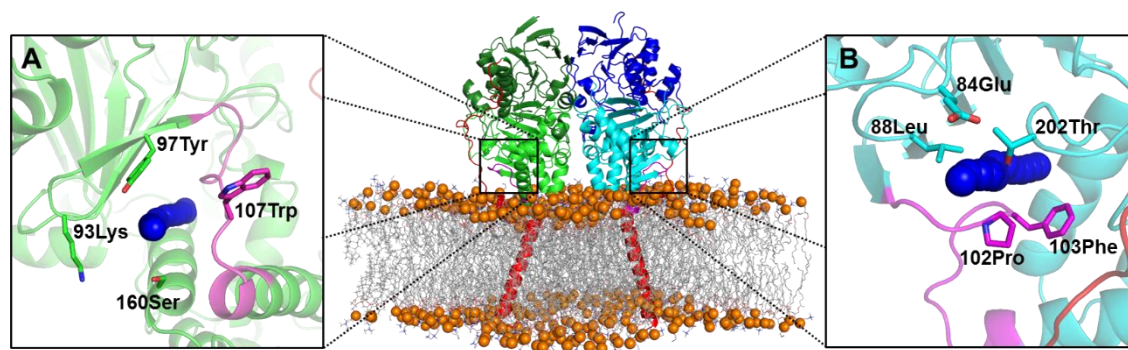


Figure 45: MAO-B in POPC/POPE membrane environment. The substrate binding area is shown in light green and light blue (residues 80-210, 286-390) for each monomer. The active site area is shown in dark green and dark blue (residues 4-79, 211-285, and 391-452). The C-terminal tail and transmembrane helix are shown in red (residues 453-520). The binding site gating loop is shown in magenta (residues 99-112). Images **A** and **B** inset denote the two main entrances (or tunnels) to the binding site (denoted Entrance A and Entrance B throughout the manuscript), with a representative tunnel (navy spheres) shown for both entrances. Key residues which describe the location of the entrance are shown as sticks and labelled. Entrance B is newly identified here, Entrance A has been previously described.³⁴⁵

Notably, both the frequency and maximum bottleneck radius of tunnels to the substrate binding cavity of MAO-B is significantly different for each monomer. In the absence of BZA, monomer 1 presents a higher frequency of tunnels, with a larger average maximum bottleneck radius. When BZA is present in monomer 1 and not monomer 2, the frequency of tunnels and maximum bottleneck measurements are higher for monomer 2 (**Figure 46A+B** and **Table S4**). In **Figure 46C+D**, the pathways are coloured according to the entrance pathway they use from the starting point to the solvent, with the majority of tunnels passing the aforementioned entrances A and B (**Figure 45**).

The observed ‘closing’ of the active site entrance in the BZA₁-bound monomer and ‘opening’ in the BZA₁-unbound monomer are particularly notable. This indicates that when BZA is bound to the active site cavity of one monomer, a subsequent BZA is more likely to enter the binding site of the opposite monomer. Such asymmetry may prevent binding of subsequent BZA into the same monomer, allowing for the release of products, whilst increasing the efficiency of binding in the opposite ‘free’ monomer. Previous experimental work indicated that binding of inhibitor to the intermediate binding site of MAO-B was increased where inhibitor was already bound into the active site (see **Figure 45** for binding/active site differentiation).³⁶⁹ Whilst it is not known whether this is within the same monomer, it could explain the asymmetry in binding site access seen here when BZA₁ is present.

BZA₂ was not considered for half-site reactivity analysis. This was due to BZA leaving the active site altogether in one simulation trajectory (**Figure S13**), meaning that the identified tunnels would not be comparable to those where BZA does not leave. This BZA

movement may indicate that BZA₂ is potentially an unstable configuration, providing further evidence towards asymmetry in MAO-B.

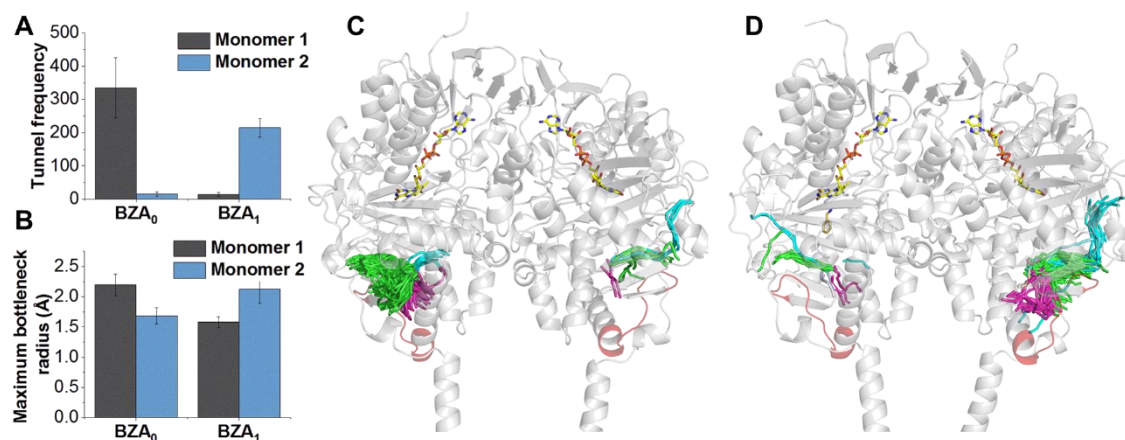


Figure 46: The summed tunnel frequency from the substrate binding site of MAO-B to solvent, identified by Caver.³⁶³ BZA₀ – No BZA is bound in either monomer. BZA₁ – BZA is bound only in monomer 1. All are from triplicate MD runs. **B**, The average maximum bottleneck (Å) from the substrate binding site of MAO-B to the solvent, as identified by Caver. **C + D**, All tunnels identified by Caver over all three MD simulation repeats for BZA₀ (**C**) and BZA₁ (**D**). FAD and BZA (when present) are shown as yellow sticks. Tunnels are colored according to entrance/exit pathway used, with magenta and green representing Entrances A and B respectively. Blue tunnels indicate pathways which go through neither of the two main entrances described.

Mutagenesis reveals long-range networks of motion are important in MAO-B.

Our findings above suggest a model where MAO-B intramolecular dynamics are intimately linked to the environment. That is, we hypothesize, that functionally important MAO-B conformational changes are at least in part controlled by the membrane environment. We expect that those conformational changes (or protein dynamics) are influenced by the global protein scaffold. To explore this hypothesis, we have turned to computationally informed mutagenesis studies.

We identify enzyme variants that are predicted to affect networks of flexible motion throughout MAO-B using the FIRST (floppy inclusion and rigid substructure topography) algorithm.³⁷⁰ FIRST uses a single conformation (e.g. a crystal structure) to define a constraint-network of movement for a protein. This constraint-network is composed of terms describing covalent and non-covalent contributions to the rigidity of the protein, allowing prediction of the relative rigidity/flexibility of each residue. These calculations can therefore be used to predict the relative rigidity (and therefore stability) of MAO-B variants, both on a local (per-residue) and global (sum of all residues) level.

FIRST was used to perform high-throughput screening to identify point variants that were predicted to significantly rigidify the network of flexible motion of *wt*MAO-B, without significantly perturbing the flexibility of residues considered

In short, we perform computational alanine scanning, by applying FIRST with the contribution of each amino-acid side chain to the constraint network discounted in turn. This allows us to identify residues with side chains that make a negligible contribution to the overall rigidity of MAO-B. All residues identified with the above approach were then subjected to *in silico* mutagenesis to all other canonical amino acids and subsequent FIRST analysis.

From these calculations, we selected three enzyme variants (W184F, F402V and E466Y) that, based on our calculations, were predicted to increase the rigidity of the protein scaffold, but are distal to the active site (between 13-21 Å from BZA) and do not significantly alter the rigidity of catalytically relevant residues (**Figure S14A-C**; see *SI Materials and Methods*). We find that these variants retain the absorption feature at ~415 nm (**Figure S8A**) and their overall structure is not significantly perturbed, at least as assessed from their far-UV circular dichroism spectra (**Figure S8B**). However, despite the variants being located a significant distance from the active site, we find that enzyme activity is ablated. Given that these variants are expected to alter the network of flexible motion, we suggest these data reflect the importance of protein conformational changes in MAO-B turnover. Moreover, the importance of such changes may also provide a rationale for the differences of our temperature-dependence studies when MAO-B is in different environments (**Figure 41**).

The FIRST calculations have the advantage that they are rapid and so enable very large *in silico* screening. However, they are thus necessarily approximate and do not reflect the realistic (and complex) protein molecular dynamics. An additional important caveat of our FIRST calculations is that they do not include the membrane environment. Therefore, in an attempt to rationalize the loss of activity for the variants, and explore our hypothesis above, we have performed 3 independent 100 ns MD simulations on each variant in the BZA₁ state (as above, see *SI Materials and Methods*).

Based on the flexibility in MD simulations (root-mean square fluctuations, RMSF), each variant is predicted to be slightly more rigid (**Figure S14D-F**) than *wt*MAO-B. Comparison of predictions of relative rigidity (by FIRST) and flexibility (by MD) for *wt*MAO-B show only moderate correlation (**Figure S15**). Poor correlation is particularly found for residues near the bilayer, which is not taken into account in our FIRST calculations; this demonstrates the importance of modelling the molecule in a native-like environment.

Figure 47A-D, (additional measurements can be found in **Figures S16** and **S17**) shows the fluctuation in the catalytic distance between BZA and FAD (BZA CH₂ and FAD N5) for each variant and *wt*MAO-B. There is clearly some fluctuation in individual simulations (with excursions to longer distances), but no significant differences are

observed between the variants and wtMAO-B. That is, we do not find any evidence (at least on these timescales) that catalytically competent conformations are disadvantaged in the variants. Therefore, based on these data, we do not expect the immediate active site environment to be compromised in the enzyme variants, which correlates with the observation of the retention of the putative anionic semiquinone peak in the absorption spectrum of each variant (**Figure S8**).

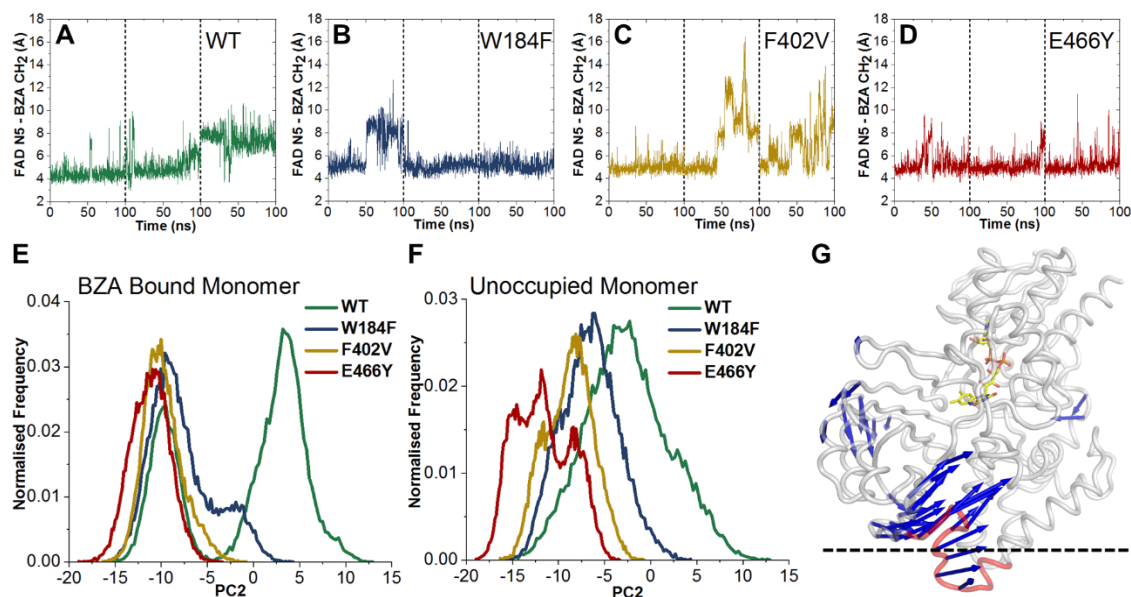


Figure 47: (A–D) Distance between the N5 (FAD) and CH2 group of BZA for wtMAO-B and all three enzyme variant simulations. A black dotted line indicates the start of each new trajectory (all runs performed in triplicate). Additional measurements for all BZA containing simulations can be found in **Figures S16** and **S17**. (E+F), Normalized histograms (bin width 0.25 Å) of principle principal component 2 (PC2) for all BZA1 simulations of the bound (E) and unoccupied monomers (F). (G) ‘Porcupine’ plot of PC2, with arrows indicating the direction of the PC2 eigenvector and arrow size indicating the magnitude of the corresponding eigenvalue, for all C α atoms with eigenvalues greater than 4 Å. The gating loop residues (99 – 112) are coloured in red, and the approximate location of the bilayer is indicated with a black dotted line. (A mobility plot of PC2 is provided in **Figure S18A**). key for catalysis (see *Supporting Information Materials and Methods* and **Table S5**).

To explore the changes in global protein dynamics for the MAO-B variants in more detail, we performed principal component analysis (PCA) on the C α carbon of residues 1-455 for all BZA₁ simulations (**Figure 47** and **S18**). The vector which describes each PC can be projected onto a static structure (to create a porcupine plot; **Figure 47G**), with the magnitude of the vector describing the relative change in each residues position over the PC. Analysis of the porcupine plot of PC1 (**Figure S18B**) shows a global motion

dependent on several flexible regions of MAO-B, whereas PC2 is dominated by the movement of the gating loop (residues 99–112) region and residues around it (**Figure 47G**). This gating loop sits between Entrance A and B (**Figure 45**) and is key for substrate binding and product release. Histograms of PC2 (**Figure 47E+F**) show reduced/differential conformational sampling of PC2 for all variants compared to *wt*MAO-B. Our MD simulations therefore suggest that the variants have lost activity because the mutations have altered the normal dynamics of the gating loop, potentially affecting substrate entrance/exit to the active site. The fact that these variants (distal from the gating loop) have such a significant and specific effect on the gating loop sampling implies the presence of a long-range network of motions in MAO-B (**Figure 47G**) through which dynamics of the gating loop are affected.

Conclusions

MAO-B is an important biomedical target, and as with many such targets, is associated with a biological membrane. Such association places enzymes in specific physiological contexts, can promote interaction with other species and enable specific structural features. Less obvious is the role of the membrane environment in mediating the conformational dynamics of enzymes, and whether this is functionally important.

Previous molecular dynamics studies have illustrated that the immediate membrane environment of MAO-B is involved in controlling substrate entrance to channels leading to the active site. Here, using enzyme kinetic and mutagenesis studies, we consider the role of the membrane environment in tuning the molecular dynamics of MAO-B more widely, including its influence on turnover and catalysis. We find that placement of MAO-B in SMA nanodiscs instead of in reduced Triton X-100 has a significant effect on the heat capacity of catalysis (ΔC_p^\ddagger). Differences in ΔC_p^\ddagger reflect differences in the distribution and frequency of vibrational modes between the ground (reactant) and transition states and implies that the membrane environment is affecting the global protein dynamics of MAO-B *and* that these dynamics influence the thermodynamics of enzyme turnover. Indeed, k_{cat} is ~ 5 times faster in the nanodisc environment versus reduced Triton X-100. These data further hint at the role of the specific lipid composition and fine structure of the membrane to tune MAO-B turnover efficiency. These findings are corroborated by studying enzyme variants that are predicted to alter the network of flexible motion in the enzyme, but do not affect the overall structure. These variants, which are distal to the active site, all lead to inactive enzyme, indicating the critical role of ‘optimized’ global protein dynamics of MAO-B.

Through atomistic molecular dynamics simulations with and without substrate bound, we find two substrate entrance/product exit channels that are mediated by membrane interaction, one of which was not previously described. The simulations further indicate an asymmetry in substrate access in the MOA-B dimer. Specifically, either one or the

other monomer may allow substrate access at any one time, with active site occupation in one monomer preferentially allowing substrate access to the other. Moreover, the gating loop dynamics appear to be highly sensitive to the global enzyme dynamics, potentially reflecting long-range networks of enzyme motion.

Taken together, our study suggests that the global protein dynamics of MAO-B are ‘tuned’ by the specific immediate membrane environment. These protein dynamics have a major effect on MAO-B function, through tuning fluctuations linked to enzyme turnover, including controlling the opening and closing of substrate/product channels. The finding that two different channels mediated by the membrane environment are present in MAO-B illustrates the potential to exploit novel small molecule binding sites with rational drug design. Therefore, our study illustrates that when searching for novel small molecule binding sites, one should consider not just the static structure of the system in isolation, but time-dependent changes in the population of conformational sub-states and in the ‘native’ environment.^{371,372} The finding that there may be long range networks of motions that can, in particular, affect the gating loop also indicates the potential for allosteric inhibitors. Further, as MAO-B resides in the outer mitochondrial membrane of cells, the finding of a solvent accessible entrance is important for rational drug design efforts. That is, inhibitors that target MAO-B may not need to enter the mitochondrial membrane in order to access the active site of MAO-B.

Materials and Methods

Unless otherwise stated, all reagents were obtained from Sigma-Aldrich.

MAO-B expression and purification.

MAO-B was expressed and purified following the purification protocol by Newton-Vinson *et al.*³⁴² Small variations from the protocol include shake flask fermentation instead of bioreactor fermentation, with BMMY (buffered methanol-complex medium) media instead of MM (minimal methanol) media for induction, storage of cell pellets in a buffer with protease inhibitor tablets instead of PMSF, suspension of pellet in 100 ml of breaking buffer instead of 1 L, and cell breakage of 30 s on 30 s off x 10 sonication in addition to bead beating. The purification was completed after MAO-B was passed over a DEAE-sepharose FF column, achieving satisfactory purity. The additional polymer partition step detailed by Newton-Vinson *et al.*³⁴² did not change the state of the semiquinone MAO-B species seen here, with additional purity achieved by size exclusion chromatography if necessary. MAO-B concentration was determined spectroscopically using $\epsilon_{415} = 11\,000\text{ M}^{-1}\text{ cm}^{-1}$.³⁷³

Nanodisc preparation.

Nanodiscs were prepared following methods by McDowall *et al.*,³⁵⁹ suspended in 50 mM HEPES, pH 7.5. and incubated with MAO-B in a 10x molar excess, for 2 hours prior to experiments.

Enzyme assays.

MAO-B was transferred from Triton X-100 containing buffer to reduced Triton X-100 or nanodisc containing buffer using detergent removal spin columns (Thermo Scientific Pierce). Steady-state MAO-B kinetic measurements were carried out using a 1 ml quartz cuvette and a UV/Vis spectro-photometer (Agilent Cary 60 UV-Vis spectrometer) in 50 mM HEPES (pH 7.5), containing 0.5 % (w/v) reduced Triton X-100 or SMA 1- α -Phosphatidylcholine nanodiscs. Enzyme activity was measured by following the formation of benzaldehyde using $\epsilon_{250} = 12,800 \text{ M}^{-1} \text{ cm}^{-1}$ for BZA,³³⁷ and 4-hydroxyquinoline using $\epsilon_{316} = 12,300 \text{ M}^{-1} \text{ cm}^{-1}$ for KYN.³⁶⁰ For each condition substrate dependences were monitored at 40 °C ; the data fitted well to Michaelis-Menten kinetics (**Figure S7**). Temperature-dependences were carried out from 20 °C - 45 °C at 5 °C intervals using initial velocity measurements at substrate concentrations maintained above 10x K_m to ensure saturation. The data were fitted to (Eq 16) as described in the manuscript using OriginPro 2017.

Redox assays.

These experiments were performed anaerobically, all buffer was purged with nitrogen and samples were prepared in an anaerobic box. Glucose and glucose oxidase were added to maintain anaerobic conditions.

EPR.

Measurements were performed using WT and Y398F variant MAO-B in 50 mM HEPES 0.5% Triton, pH 7.5, flash-frozen in liquid nitrogen in suprasil quartz sample tubes. X-band cw-EPR spectra were recorded on a Bruker eleXsys E500 spectrometer using a standard rectangular Bruker EPR cavity (ER4102T) equipped with an Oxford helium cryostat (ESR900). Experimental parameters: microwave power, 0.2 mW; field modulation amplitude, 5 G; field modulation frequency, 100 kHz; measuring time 10 s; temperature 16 K. Q-band cw-EPR spectra were performed on a Bruker eleXsys E-560 spectrometer using a ER 5106QT-W1 resonator equipped with a home-built ARS cryogen-free cryostat (data not shown). Spectral simulations were performed using the Matlab-based Easyspin package.³⁷⁴

Computational materials and methods.

The X-ray crystal structure of MAO-B in complex with 6-hydroxy-N-propargyl-1(R)-aminoindan (1S3E)³⁷⁵ was used as the starting point for all MD simulations. The missing C-terminal residues (502 – 520 Chain A and 497 – 520 Chain B) which form the remainder of the transmembrane helix were built using Avogadro³⁷⁶ (assuming the standard backbone dihedral angles of an α -helix). MAO-B was then inserted into a lipid bilayer comprised of a 4:3 ratio of palmitoyloleoylphosphatidylcholine (POPC) and

palmitoyl-oleoyl-phosphatidylethanolamine (POPE) using CHARMM-GUI.³⁷⁷ This composition has been used in prior bilayer MD simulations of MAO-B, and was chosen to represent the composition of the outer mitochondrial membrane.^{345,351,362} BZA was placed in the active site by alignment with the crystal structure of MAO-B in complex with nitrobenzylamine (2C70,³⁷⁸ C α RMSD 0.118 Å to 1S3E; the latter was used due to its higher resolution). All simulations of BZA were performed with the amino group in its neutral form, as this is widely believed to be the catalytically competent state of BZA.^{332,338,379} Titratable residues were simulated in their standard protonation states, based on calculations with PropKa 3.0¹⁵⁸. MolProbity was used to determine the optimum tautomerisation states of every His residue, and any required Asn/Gln side chain flips, based on optimizing the hydrogen bonding network⁷⁵. Histidines 91, 115 and 452 were singly protonated on their N δ 1, with all others singly protonated on their N ϵ 2. The system was then solvated such that there was no protein or lipid atom within 20 Å of the edge of the periodic box along the z-coordinate (bilayer normal). The total number of atoms for each system simulated was approximately 90,000.

Periodic boundary simulations were performed with Amber16, using the CHARMM36 force field to describe protein³⁸⁰ and lipid³⁸¹ atoms, and TIP3P for water. Parameters for FAD in its oxidized form and BZA in its neutral form were taken from Abad *et al.*³³⁸ Following minimization, heating and equilibration (see SI Materials and Methods), production MD simulations were run in the NpT ensemble at 310 K, with semi-isotropic coupling to a Monte Carlo barostat. Temperature was regulated using Langevin dynamics with a collision frequency of 1 ps⁻¹. A time step of 2 fs was applied with the covalent bonds to hydrogen constrained by the SHAKE algorithm. A 12 Å non-bonded cut-off was applied with a force switch smoothing function from 10 to 12 Å. Long range electrostatics were evaluated with the particle mesh Ewald method.⁷² For *wt*MAO-B simulations, a total of 9 simulations were performed for 150 ns each, with the first 50 ns of simulation used to equilibrate the protein and bilayer (see **Figure S10**). The BZA₂ run in which a BZA escapes the active site cavity was extended for a further 50 ns. Enzyme variant simulations (W184F, F402V, E466Y) were performed for 3 x 100 ns each in the BZA1 state (see SI Materials and Methods for further details). Coordinates were saved every 10 ps for further analysis. Routine analysis of trajectories was performed using CPPTRAJ¹⁶⁰, from the AmberTools suite. Area per lipid calculations were performed with GridMAT-MD,³⁸² using a grid resolution of 200 x 200 points for each measurement. Tunnel analysis was performed with Caver 3.0³⁶³. All settings were kept default apart from bottleneck radius (1.5 Å).

ASSOCIATED CONTENT

Supporting Information. FIRST calculations methodology, molecular dynamics simulation methodology, MAO-B mechanisms, table of caver data, MAO-B variant and environment spectra, Michealis-Menten of Y398F MAO-B, MD RMSD and area per lipid calculations, Entrance B structure, Caver calculation starting point, RMSF of gating loop,

BZA leading the active site by MD simulations, Michaelis-Mentens of *wt*MAO-B. This information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Authors

*Christopher R Pudney, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom. c.r.pudney@bath.ac.uk.

*Marc W van der Kamp, School of Biochemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom. marc.vanderkamp@bristol.ac.uk.

Author Contributions

HBLJ and AM performed experimental work. RMC and HBLJ performed computational calculations and simulations. All authors discussed and interpreted data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. HBLJ and RMC contributed equally[†].

ABBREVIATIONS

Alpha carbon ($C\alpha$), benzylamine (BZA), electron paramagnetic resonance (EPR), flavin adenine dinucleotide (FAD), heat capacity of catalysis (ΔC_p^\ddagger), kynuramine (KYN), molecular dynamics (MD), monoamine oxidase (MAO), principal component analysis (PCA), palmitoylcholine (POPC), palmitoylcholine (POPE), root mean square fluctuation (RMSF), reactive oxygen species (ROS).

ACKNOWLEDGMENTS

HBLJ's studentship is funded by the University of Bath. RMC's studentship is funded by the EPSRC. MWvdK is a BBSRC David Phillips Fellow (BB/M026280/1). SAW has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 648283 "GROWMOF"). This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath

Supporting information

Exposing the Interplay Between Enzyme Turnover, Protein Dynamics and the Membrane Environment in Monoamine Oxidase B

Hannah B. L. Jones^{†||}, Rory M. Crean^{#†||}, Anna Mullen[§], Emanuele G. Kendrick[†], Steven D Bull^Δ, Stephen A. Wells[‡], David R. Carbery^Δ, Fraser MacMillan[§], Marc W. Van der Kamp^{*%}, Christopher R. Pudney^{*†|}.

[†]Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom.

^ΔDepartment of Chemistry, University of Bath, Bath BA2 7AY, United Kingdom.

[‡]Department of Chemical Engineering, University of Bath, Bath BA2 7AY, United Kingdom.

[#] Doctoral Training Centre in Sustainable Chemical Technologies, University of Bath, Bath BA2 7AY, United Kingdom.

[|]Centre for Therapeutic Innovation, University of Bath, Bath BA2 7AY, United Kingdom.

[%]School of Biochemistry, University of Bristol, Biomedical Sciences building, University Walk, Bristol BS8 1TD, United Kingdom.

[§]School of Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom.

*Christopher R Pudney, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom. c.r.pudney@bath.ac.uk.

*Marc W van der Kamp, School of Biochemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom. marc.vanderkamp@bristol.ac.uk.

Author Contributions

^{||}H.B.L.J. and R.M.C. contributed equally. HBLJ and AM performed experimental work. RMC and HBLJ performed computational calculations and simulations. All authors discussed and interpreted data. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

FIRST calculations

FIRST³⁷⁰ (floppy inclusion and rigid substructure topography) calculations, as mentioned in the main text, were carried out on a crystal structure of Monoamine oxidase B (MAO-B) (PDB 1S3E)³⁷⁵. The aim of this procedure was to establish rigidified mutations of MAO-B, and to assess their impact on catalysis. Scwrl4³¹⁵ was used to optimize the side-chain conformations of 1S3E MAO-B, and calculations were run on both the original and optimized structures. Structures that were optimized using Scwrl4 were constrained against their FAD cofactor (FAD atoms are treated as non-interacting particles that provide steric restraints). Of mutants carried forward for experimental testing, F402V and W184F arose from Scwrl4 optimized 1S3E, and E466Y arose from the original 1S3E crystal structure. Prior to running calculations, the X-ray structure was pre-processed by removing all hetero atoms (apart from FAD), and water molecules with PyMOL, and adding hydrogens to all residues (all residues were at their standard protonation state for pH 7) using Molprobit⁷⁵.

An initial FIRST run was used to generate the constraint files for wildtype (WT) MAO-B, which describe the strength of the covalent bonds, hydrogen bonds, hydrophobic interactions and pi-pi stacking interactions throughout the protein. FIRST was then run over 20 energy cut offs (Ecuts), using rigid cluster decomposition (RCD) as previously described.³⁸³ Results are quantified as 'the fraction of residues in rigid clusters' for each Ecut. Ecuts ranged from 0-4 kcal mol⁻¹ with a step size of 0.2 kcal mol⁻¹. The fraction of residues in rigid clusters from all Ecuts were then averaged to give a single value to reflect the overall rigidity of the protein. This was used as the background value for comparison to all further calculations.

To replicate the premise of alanine scanning, a protocol 'Sidescan' (available upon request) was developed to produce the constraint files calculated by FIRST with the constraint contribution of a specified amino acid side chain past the beta carbon (C β) missing from the files. These altered constraint files were produced sequentially for each amino acid (not including alanine or glycine) of MAO-B, and the RCD analysis was carried out to determine the rigidity value of MAO-B without each amino acid. A random sample of residues whose side chains were identified as making no contribution to the rigidity of MAO-B, were taken forward for *in silico* mutagenesis and additional screening within FIRST. Scwrl4 was used to mutate these residues to each possible amino acid alternative, and the FIRST/RCD calculations were implemented, to predict the relative rigidity of each mutant compared to the WT.

Mutations that improved the overall rigidity of MAO-B were then analyzed using the protocol 'GetRCD' (available upon request). GetRCD assigned each amino acid in MAO-B an E_{cut} (kcal mol⁻¹) value, which reflects the amount of energy required to free that residue from its constraint network. This was carried out on the selected mutants to ascertain whether they would significantly affect the flexibility/rigidity of the catalytically

relevant residues of *wt*MAO-B (**Table S5**). If a mutation altered the flexibility/rigidity of a catalytically relevant residue by more than 1 kcal mol⁻¹ it was disregarded.

From these calculations we selected three single point mutations to experimentally test (W184F, F402V, E466Y). W184F is located in the FAD binding domain of MAO-B and is 13.3 Å away from BZA and 15.9 Å away from the flavin ring of FAD. F402V is located in the FAD binding domain of MAO-B and is 15.3 Å away from BZA and 15.3 Å away from the flavin ring of FAD. E466Y is located on the C-terminal tail of MAO-B and is 21.4 Å away from BZA and 16.6 Å away from the flavin ring of FAD. All measurements are based on PDB 2C70 (chosen because it is co-crystallized with nitrobenzylamine) and based on the closest atom to atom distance of the amino acid backbone and FAD or BZA. Graphical representations of the location of each mutant, alongside their predicted change in rigidity from FIRST calculations can be found in **Figure S14A-C**.

Molecular Dynamics (MD) Simulations

Equilibration Procedure

Following structure preparation (as described in the main text), all systems were subjected to the same equilibration procedure as detailed below. First, all water molecules and hydrogen atoms were minimized with 500 steps of steepest descent followed by 500 steps of conjugate gradient minimization. The system was then rapidly heated from 100 K to 310 K in the NVT ensemble over 100 ps, with 10 kcal mol⁻¹ positional restraints on all protein and lipid atoms. Whilst retaining the 10 kcal mol⁻¹ restraints on protein atoms, lipid restraints were gradually reduced to 2.5 kcal mol⁻¹ over the course of 200 ps. During NVT simulation, the timestep was set to 1 fs and a collision frequency of 1 ps⁻¹ was used with Langevin temperature control. Systems were then simulated in the NPT ensemble (1 atm, 310 K, semi-isotropic coupling to a Monte Carlo barostat, Langevin temperature control with a collision frequency of 1 ps⁻¹, 2 fs time step), with lipid restraints reduced from 2.5 to 0 kcal mol⁻¹ over the course of 200 ps. Finally, protein restraints were reduced from 10 kcal mol⁻¹ to 0 kcal mol⁻¹ over the course of 1 ns. At this point 50 ns of production MD simulations were performed to equilibrate the bilayer and protein (as described in the main text).

Mutant Simulations

The point mutants (W184F, F402V and E466Y) were each subjected to 3 x 100 ns of MD simulation in the BZA₁ bound state. The structure of *wt*MAOB-BZA₁ at 50 ns into each of the three independent simulations was used to generate three unique starting structures for the point mutant MD simulations. BZA was re-orientated to a catalytically competent state for Run 3 simulations based on the structure of nitrobenzylamine in complex with MAO-B (PDB 2C70). Mutations were made in silico using PyMOL³⁸⁴, and any overlapping water molecules with the newly introduced sidechains were removed. For the E466Y mutation a single sodium atom was removed from each simulation box to ensure

the system had an overall charge zero. Each system was then prepared as described above (see Equilibration Procedure). Trajectory analysis was performed without the first 10 ns of simulation time, which was used to equilibrate each mutant simulation.

Trajectory Analysis

Trajectory analysis was performed using CPPTRAJ¹⁶⁰, (part of the AmberTools suite of programmes). Per residue Root Mean Square Fluctuations (RMSFs) were performed on snapshots at 10 ps intervals using average structures calculated from 1 ns time windows over the course of each trajectory. Performing RMSF calculations in this way reduces the influence of larger scale conformational changes on the calculation, providing a more realistic indication of backbone mobility.³⁸⁵ RMS fitting was performed on the C α carbon of residues 1-455 of each monomer. Clustering analysis was performed using the DBSCAN algorithm³⁸⁶, using all backbone heavy atoms of the three loop regions (81–85, 100–108 and 202–207) which describe the opening of Entrance B. Principal Component Analysis was performed on all BZA₁ simulations (WT, W184F, F402V and E466Y) simultaneously using the C α carbon of residues 1-455 of each monomer (with monomers inputted separately).

PLEASE NOTE: In the interest of saving paper/space, only the supporting information Methods are included and not the figures or tables. They can instead be found online free of charge at:
<https://pubs.acs.org/doi/10.1021/acs.biochem.9b00213>

References

- (1) Goldsby, R. A.; Kuby, J. *Immunology*, 5th ed.; Freeman, W. H., Ed.; New York, 2002.
- (2) Badalà, F.; Nouri-mahdavi, K.; Raoof, D. A. Reconstructing Immune Phylogeny: New Perspectives. *Nat. Rev. Immunol.* **2005**, *5*, 866–879.
- (3) Bonilla, F. A.; Oettgen, H. C. Adaptive Immunity. *J. Allergy Clin. Immunol.* **2010**, *125*, S33–S40.
- (4) Chaplin, D. D. Overview of the Immune Response. *J. Allergy Clin. Immunol.* **2010**, *125*, S3–S23.
- (5) Rudolph, M. G.; Stanfield, R. L.; Wilson, I. A. How TCRs Bind MHCs, Peptides, and Coreceptors. *Annu. Rev. Immunol.* **2006**, *24*, 419–466.
- (6) Rossjohn, J.; Gras, S.; Miles, J. J.; Turner, S. J.; Godfrey, D. I.; McCluskey, J. T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules. *Annu. Rev. Immunol.* **2015**, *33*, 169–200.
- (7) Comber, J. D.; Philip, R. MHC Class I Antigen Presentation and Implications for Developing a New Generation of Therapeutic Vaccines. *Ther. Adv. Vaccines* **2014**, *2*, 77–89.
- (8) Hewitt, E. W. The MHC Class I Antigen Presentation Pathway: Strategies for Viral Immune Evasion. *Immunology* **2003**, *110*, 163–169.
- (9) Jones, E. Y.; Fugger, L.; Strominger, J. L.; Siebold, C. MHC Class II Proteins and Disease: A Structural Perspective. *Nat. Rev. Immunol.* **2006**, *6*, 271–282.
- (10) Trolle, T.; McMurtrey, C. P.; Sidney, J.; Bardet, W.; Osborn, S. C.; Kaever, T.; Sette, A.; Hildebrand, W. H.; Nielsen, M.; Peters, B. The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. *J. Immunol.* **2016**, *196*, 1480–1487.
- (11) Parker, K. C.; Bednarek, M. A.; Hull, L. K.; Utz, U.; Cunningham, B.; Zweerink, H. J.; Biddison, W. E.; Coligan, J. E. Sequence Motifs Important for Peptide Binding to the Human MHC Class I Molecule, HLA-A2. *J. Immunol.* **1992**, *149*, 3580–3587.
- (12) Guillaume, P.; Picaud, S.; Baumgaertner, P.; Montandon, N.; Schmidt, J.; Speiser, D. E.; Coukos, G.; Bassani-Sternberg, M.; Filippakopoulos, P.; Gfeller, D. The C-Terminal Extension Landscape of Naturally Presented HLA-I Ligands. *Proc. Natl. Acad. Sci.* **2018**, *115*, 5083–5088.
- (13) Matsumura, M.; Fremont, D.; Peterson, P.; Wilson, I. Emerging Principles for the Recognition of Peptide Antigens by MHC Class I Molecules. *Science* **1992**, *257*, 927–934.
- (14) Arstila, T. P. A Direct Estimate of the Human T Cell Receptor Diversity. *Science* **1999**, *286*, 958–961.
- (15) Mason, D. A Very High Level of Crossreactivity Is an Essential Feature of the T-Cell Receptor. *Immunol. Today* **1998**, *19*, 395–404.
- (16) Wooldridge, L.; Ekeruche-Makinde, J.; van den Berg, H. A.; Skowera, A.; Miles, J. J.; Tan, M. P.; Dolton, G.; Clement, M.; Llewellyn-Lacey, S.; Price, D. A.; Peakman, M.; Sewell, A. K. A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides. *J. Biol. Chem.* **2012**, *287*, 1168–1177.
- (17) Vriskoop, N.; Monteiro, J. P.; Mandl, J. N.; Germain, R. N. Revisiting Thymic Positive Selection and the Mature T Cell Repertoire for Antigen. *Immunity* **2014**, *41*, 181–190.
- (18) Aleksic, M.; Liddy, N.; Molloy, P. E.; Pumphrey, N.; Vuidepot, A.; Chang, K.-M.; Jakobsen, B. K.

Different Affinity Windows for Virus and Cancer-Specific T-Cell Receptors: Implications for Therapeutic Strategies. *Eur. J. Immunol.* **2012**, *42*, 3174–3179.

- (19) Garrido, F.; Cabrera, T.; Aptsiauri, N. “Hard” and “Soft” Lesions Underlying the HLA Class I Alterations in Cancer Cells: Implications for Immunotherapy. *Int. J. Cancer* **2010**, *127*, 249–256.
- (20) Dunbar, J.; Fuchs, A.; Shi, J.; Deane, C. M. ABangle: Characterising the VH-VL Orientation in Antibodies. *Protein Eng. Des. Sel.* **2013**, *26*, 611–620.
- (21) Dunbar, J.; Knapp, B.; Fuchs, A.; Shi, J.; Deane, C. M. Examining Variable Domain Orientations in Antigen Receptors Gives Insight into TCR-Like Antibody Design. *PLoS Comput. Biol.* **2014**, *10*, e1003852.
- (22) Riley, T. P.; Baker, B. M. The Intersection of Affinity and Specificity in the Development and Optimization of T Cell Receptor Based Therapeutics. *Semin. Cell Dev. Biol.* **2018**, *84*, 30–41.
- (23) Cole, D. K.; Bulek, A. M.; Dolton, G.; Schauenberg, A. J.; Szomolay, B.; Rittase, W.; Trimby, A.; Jothikumar, P.; Fuller, A.; Skowera, A.; Rossjohn, J.; Zhu, C.; Miles, J. J.; Peakman, M.; Wooldridge, L.; Rizkallah, P. J.; Sewell, A. K. Hotspot Autoimmune T Cell Receptor Binding Underlies Pathogen and Insulin Peptide Cross-Reactivity. *J. Clin. Invest.* **2016**, *126*, 2191–2204.
- (24) Riley, T. P.; Hellman, L. M.; Gee, M. H.; Mendoza, J. L.; Alonso, J. A.; Foley, K. C.; Nishimura, M. I.; Vander Kooi, C. W.; Garcia, K. C.; Baker, B. M. T Cell Receptor Cross-Reactivity Expanded by Dramatic Peptide–MHC Adaptability. *Nat. Chem. Biol.* **2018**, *14*, 934–942.
- (25) Armstrong, K. M.; Piepenbrink, K. H.; Baker, B. M. Conformational Changes and Flexibility in T-Cell Receptor Recognition of Peptide-MHC Complexes. *Biochem. J.* **2008**, *415*, 183–196.
- (26) Oates, J.; Jakobsen, B. K. ImmTACs. *Oncoimmunology* **2013**, *2*, e22891.
- (27) Rohaan, M. W.; van den Berg, J. H.; Kvistborg, P.; Haanen, J. B. A. G. Adoptive Transfer of Tumor-Infiltrating Lymphocytes in Melanoma: A Viable Treatment Option. *J. Immunother. Cancer* **2018**, *6*, 102.
- (28) Baruch, E. N.; Berg, A. L.; Besser, M. J.; Schachter, J.; Markel, G. Adoptive T Cell Therapy: An Overview of Obstacles and Opportunities. *Cancer* **2017**, *123*, 2154–2162.
- (29) Feins, S.; Kong, W.; Williams, E. F.; Milone, M. C.; Fraietta, J. A. An Introduction to Chimeric Antigen Receptor (CAR) T-cell Immunotherapy for Human Cancer. *Am. J. Hematol.* **2019**, *94*, S3–S9.
- (30) Gebauer, M.; Skerra, A. Engineered Protein Scaffolds as Next-Generation Antibody Therapeutics. *Curr. Opin. Chem. Biol.* **2009**, *13*, 245–255.
- (31) Chames, P.; Van Regenmortel, M.; Weiss, E.; Baty, D. Therapeutic Antibodies: Successes, Limitations and Hopes for the Future. *Br. J. Pharmacol.* **2009**, *157*, 220–233.
- (32) de Souza, J. E. S.; Galante, P. A. F.; de Almeida, R. V. B.; da Cunha, J. P. C.; Ohara, D. T.; Ohno-Machado, L.; Old, L. J.; de Souza, S. J. SurfaceomeDB: A Cancer-Orientated Database for Genes Encoding Cell Surface Proteins. *Cancer Immun.* **2012**, *12*, 15.
- (33) Liddy, N.; Bossi, G.; Adams, K. J.; Lissina, A.; Mahon, T. M.; Hassan, N. J.; Gavarret, J.; Bianchi, F. C.; Pumphrey, N. J.; Ladell, K.; Gostick, E.; Sewell, A. K.; Lissin, N. M.; Harwood, N. E.; Molloy, P. E.; Li, Y.; Cameron, B. J.; Sami, M.; Baston, E. E.; Todorov, P. T.; Paston, S. J.; Dennis, R. E.; Harper, J. V.; Dunn, S. M.; Ashfield, R.; Johnson, A.; McGrath, Y.; Plesa, G.; June, C. H.; Kalos, M.; Price, D. A.; Vuidepot, A.; Williams, D. D.; Sutton, D. H.; Jakobsen, B. K. Monoclonal TCR-Redirected Tumor Cell Killing. *Nat. Med.* **2012**, *18*, 980–987.
- (34) Stewart-Jones, G.; Wadle, A.; Hombach, A.; Shenderov, E.; Held, G.; Fischer, E.; Kleber, S.; Nuber, N.; Stenner-Liewen, F.; Bauer, S.; McMichael, A.; Knuth, A.; Abken, H.; Hombach, A. A.; Cerundolo, V.; Jones, E. Y.; Renner, C. Rational Development of High-Affinity T-Cell Receptor

like Antibodies. *Proc. Natl. Acad. Sci.* **2009**, *106*, 5784–5788.

- (35) Hülsmeier, M.; Chames, P.; Hillig, R. C.; Stanfield, R. L.; Held, G.; Coulie, P. G.; Alings, C.; Wille, G.; Saenger, W.; Uchanska-Ziegler, B.; Hoogenboom, H. R.; Ziegler, A. A Major Histocompatibility Complex-Peptide-Restricted Antibody and T Cell Receptor Molecules Recognize Their Target by Distinct Binding Modes. *J. Biol. Chem.* **2005**, *280*, 2972–2980.
- (36) Ataie, N.; Xiang, J.; Cheng, N.; Brea, E. J.; Lu, W.; Scheinberg, D. A.; Liu, C.; Ng, H. L. Structure of a TCR-Mimic Antibody with Target Predicts Pharmacogenetics. *J. Mol. Biol.* **2016**, *428*, 194–205.
- (37) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of “Alchemical Perturbation” in Medicinal Chemistry. *J. Med. Chem.* **2018**, *61*, 638–649.
- (38) Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; Wiley: West Sussex, 2007.
- (39) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- (40) Zhang, L.; Udaka, K.; Mamitsuka, H.; Zhu, S. Toward More Accurate Pan-Specific MHC-Peptide Binding Prediction: A Review of Current Methods and Tools. *Brief. Bioinform.* **2012**, *13*, 350–364.
- (41) Weitzner, B. D.; Jeliakov, J. R.; Lyskov, S.; Marze, N.; Kuroda, D.; Frick, R.; Adolf-Bryfogle, J.; Biswas, N.; Dunbrack, R. L.; Gray, J. J. Modeling and Docking of Antibody Structures with Rosetta. *Nat. Protoc.* **2017**, *12*, 401–416.
- (42) Maier, J. K. X.; Labute, P. Assessment of Fully Automated Antibody Homology Modeling Protocols in Molecular Operating Environment. *Proteins Struct. Funct. Bioinforma.* **2014**, *82*, 1599–1610.
- (43) 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.
- (44) Cole, D. K.; Yuan, F.; Rizkallah, P. J.; Miles, J. J.; Gostick, E.; Price, D. A.; Gao, G. F.; Jakobsen, B. K.; Sewell, A. K. Germ Line-Governed Recognition of a Cancer Epitope by an Immunodominant Human T-Cell Receptor. *J. Biol. Chem.* **2009**, *284*, 27281–27289.
- (45) Sami, M.; Rizkallah, P. J.; Dunn, S.; Molloy, P.; Moysey, R.; Vuidepot, A.; Baston, E.; Todorov, P.; Li, Y.; Gao, F.; Boulter, J. M.; Jakobsen, B. K. Crystal Structures of High Affinity Human T-Cell Receptors Bound to Peptide Major Histocompatibility Complex Reveal Native Diagonal Binding Geometry. *Protein Eng. Des. Sel.* **2007**, *20*, 397–403.
- (46) 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.
- (47) Pierce, B. G.; Hellman, L. M.; Hossain, M.; Singh, N. K.; Vander Kooi, C. W.; Weng, Z.; Baker, B. M. Computational Design of the Affinity and Specificity of a Therapeutic T Cell Receptor. *PLoS Comput. Biol.* **2014**, *10*, e1003478.
- (48) Pierce, B.; Weng, Z. ZRANK: Reranking Protein Docking Predictions with an Optimized Energy Function. *Proteins Struct. Funct. Bioinforma.* **2007**, *67*, 1078–1086.
- (49) Song, Y.; DiMaio, F.; Wang, R. Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-Resolution Comparative Modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742.

- (50) Tynan, F. E.; Reid, H. H.; Kjer-Nielsen, L.; Miles, J. J.; Wilce, M. C. J.; Kostenko, L.; Borg, N. A.; Williamson, N. a; Beddoe, T.; Purcell, A. W.; Burrows, S. R.; McCluskey, J.; Rossjohn, J. A T Cell Receptor Flattens a Bulged Antigenic Peptide Presented by a Major Histocompatibility Complex Class I Molecule. *Nat. Immunol.* **2007**, *8*, 268–276.
- (51) Fodor, J.; Riley, B. T.; Borg, N. A.; Buckle, A. M. Previously Hidden Dynamics at the TCR–Peptide–MHC Interface Revealed. *J. Immunol.* **2018**, *200*, 4134–4145.
- (52) Pierce, B. G.; Weng, Z. A Flexible Docking Approach for Prediction of T Cell Receptor–Peptide–MHC Complexes. *Protein Sci.* **2013**, *22*, 35–46.
- (53) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Jacob, I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations. *Living J. Comput. Mol. Sci.* **2018**, *1*, 1–28.
- (54) Kamerlin, S. C. L.; Warshel, A. The Empirical Valence Bond Model: Theory and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 30–45.
- (55) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.
- (56) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- (57) Mark, P.; Nilsson, L. Structure and Dynamics of Liquid Water with Different Long-Range Interaction Truncation and Temperature Control Methods in Molecular Dynamics Simulations. *J. Comput. Chem.* **2002**, *23*, 1211–1219.
- (58) Vega, C.; Sanz, E.; Abascal, J. L. F. The Melting Temperature of the Most Common Models of Water. *J. Chem. Phys.* **2005**, *122*, 114507.
- (59) Vega, C.; Abascal, J. L. F.; Nezbeda, I. Vapor-Liquid Equilibria from the Triple Point up to the Critical Point for the New Generation of TIP4P-like Models: TIP4P/Ew, TIP4P/2005, and TIP4P/Ice. *J. Chem. Phys.* **2006**, *125*, 034503.
- (60) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (61) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (62) Dickson, C. J.; Madej, B. D.; Skjerve, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879.
- (63) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.
- (64) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (65) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (66) Jensen, J. H. *Molecular Modelling Basics*, 2nd ed.; CRC Press, 2010.
- (67) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**,

23, 327–341.

- (68) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (69) Schneider, T.; Stoll, E. Molecular-Dynamics Study of a Three-Dimensional One-Component Model for Distortive Phase Transitions. *Phys. Rev. B* **1978**, *17*, 1302–1322.
- (70) Tsai, D. H. The Virial Theorem and Stress Calculation in Molecular Dynamics. *J. Chem. Phys.* **1979**, *70*, 1375–1382.
- (71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (72) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (73) Muhammed, M. T.; Aki-Yalcin, E. Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem. Biol. Drug Des.* **2019**, *93*, 12–20.
- (74) Pagadala, N. S.; Syed, K.; Tuszyński, J. Software for Molecular Docking: A Review. *Biophys. Rev.* **2017**, *9*, 91–102.
- (75) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 12–21.
- (76) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
- (77) Swaminathan, S.; Harte, W. E.; Beveridge, D. L. Investigation of Domain Structure in Proteins via Molecular Dynamics Simulation: Application to HIV-1 Protease Dimer. *J. Am. Chem. Soc.* **1991**, *113*, 2717–2721.
- (78) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22*, 2695–2696.
- (79) Girvan, M.; Newman, M. E. J. Community Structure in Social and Biological Networks. *Proc. Natl. Acad. Sci.* **2002**, *99*, 7821–7826.
- (80) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003**, *330*, 891–913.
- (81) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.
- (82) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15*, 377–392.
- (83) Lee, M. R.; Duan, Y.; Kollman, P. a. Use of MM-PB/SA in Estimating the Free Energies of Proteins: Application to Native, Intermediates, and Unfolded Villin Headpiece. *Proteins* **2000**, *39*, 309–316.
- (84) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31*, 114–122.
- (85) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding

Affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.

- (86) 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.* **2012**, *52*, 2079–2088.
- (87) Manzoni, F.; Uranga, J.; Genheden, S.; Ryde, U. Can System Truncation Speed up Ligand-Binding Calculations with Periodic Free-Energy Simulations? *J. Chem. Inf. Model.* **2017**, *57*, 2865–2873.
- (88) Cornell, W.; Abseher, R.; Nilges, M.; Case, D. A. Continuum Solvent Molecular Dynamics Study of Flexibility in Interleukin-8. *J. Mol. Graph. Model.* **2001**, *19*, 136–145.
- (89) Fenley, A. T.; Muddana, H. S.; Gilson, M. K. Entropy-Enthalpy Transduction Caused by Conformational Shifts Can Obscure the Forces Driving Protein-Ligand Binding. *Proc. Natl. Acad. Sci.* **2012**, *109*, 20006–20011.
- (90) Duan, L.; Liu, X.; Zhang, J. Z. H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138*, 5722–5728.
- (91) 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*, 1228–1241.
- (92) Wan, S.; Knapp, B.; Wright, D. W.; Deane, C. M.; 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*, 3346–3356.
- (93) Genheden, S.; Ryde, U. How to Obtain Statistically Converged MM/GBSA Results. *J. Comput. Chem.* **2010**, *31*, 837–846.
- (94) Godschalk, F.; Genheden, S.; Söderhjelm, P.; Ryde, U. Comparison of MM/GBSA Calculations Based on Explicit and Implicit Solvent Simulations. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7731.
- (95) Maffucci, I.; Contini, A. Explicit Ligand Hydration Shells Improve the Correlation between MM-PB/GBSA Binding Energies and Experimental Activities. *J. Chem. Theory Comput.* **2013**, *9*, 2706–2717.
- (96) Zhu, Y.-L.; Beroza, P.; Artis, D. R. Including Explicit Water Molecules as Part of the Protein Structure in MM/PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 462–469.
- (97) Coppieters, K. T.; Dotta, F.; Amirian, N.; Campbell, P. D.; Kay, T. W. H.; Atkinson, M. A.; Roep, B. O.; von Herrath, M. G. Demonstration of Islet-Autoreactive CD8 T Cells in Insulinitic Lesions from Recent Onset and Long-Term Type 1 Diabetes Patients. *J. Exp. Med.* **2012**, *209*, 51–60.
- (98) Bulek, A. M.; Cole, D. K.; Skowera, A.; Dolton, G.; Gras, S.; Madura, F.; Fuller, A.; Miles, J. J.; Gostick, E.; Price, D. A.; Drijfhout, J. W.; Knight, R. R.; Huang, G. C.; Lissin, N.; Molloy, P. E.; Wooldridge, L.; Jakobsen, B. K.; Rossjohn, J.; Peakman, M.; Rizkallah, P. J.; Sewell, A. K. Structural Basis for the Killing of Human Beta Cells by CD8 + T Cells in Type 1 Diabetes. *Nat. Immunol.* **2012**, *13*, 283–289.
- (99) Cole, D. K.; Fuller, A.; Dolton, G.; Zervoudi, E.; Legut, M.; Miles, K.; Blanchfield, L.; Madura, F.; Holland, C. J.; Bulek, A. M.; Bridgeman, J. S.; Miles, J. J.; Schauenburg, A. J. A.; Beck, K.; Evavold, B. D.; Rizkallah, P. J.; Sewell, A. K. Dual Molecular Mechanisms Govern Escape at Immunodominant HLA A2-Restricted HIV Epitope. *Front. Immunol.* **2017**, *8*, 1–18.
- (100) Archbold, J. K.; Macdonald, W. A.; Miles, J. J.; Brennan, R. M.; Kjer-Nielsen, L.; McCluskey, J.; Burrows, S. R.; Rossjohn, J. Alloreactivity between Disparate Cognate and Allogeneic PMHC-I Complexes Is the Result of Highly Focused, Peptide-Dependent Structural Mimicry. *J. Biol. Chem.* **2006**, *281*, 34324–34332.

- (101) Madden, D. R.; Garboczi, D. N.; Wiley, D. C. The Antigenic Identity of Peptide-MHC Complexes: A Comparison of the Conformations of Five Viral Peptides Presented by HLA-A2. *Cell* **1993**, *75*, 693–708.
- (102) Collins, E. J.; Garboczi, D. N.; Wiley, D. C. Three-Dimensional Structure of a Peptide Extending from One End of a Class I MHC Binding Site. *Nature* **1994**, *371*, 626–629.
- (103) Saini, S. K.; Tamhane, T.; Anjanappa, R.; Saikia, A.; Ramskov, S.; Donia, M.; Svane, I. M.; Jakobsen, S. N.; Garcia-Alai, M.; Zacharias, M.; Meijers, R.; Springer, S.; Hadrup, S. R. Empty Peptide-Receptive MHC Class I Molecules for Efficient Detection of Antigen-Specific T Cells. *Sci. Immunol.* **2019**, *4*, eaau9039.
- (104) Momburg, F.; Tan, P. Tapasin - The Keystone of the Loading Complex Optimizing Peptide Binding by MHC Class I Molecules in the Endoplasmic Reticulum. *Mol. Immunol.* **2002**, *39*, 217–233.
- (105) 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.
- (106) Garcia, K. C.; Degano, M.; Stanfield, R. L.; Brunmark, A.; Jackson, M. R.; Peterson, P. a; Teyton, L.; Wilson, I. a. An Alphabeta T Cell Receptor Structure at 2.5 Å and Its Orientation in the TCR-MHC Complex. *Science* **1996**, *274*, 209–219.
- (107) Gao, G. F.; Tormo, J.; Gerth, U. C.; Wyer, J. R.; McMichael, a J.; Stuart, D. I.; Bell, J. I.; Jones, E. Y.; Jakobsen, B. K. Crystal Structure of the Complex between Human CD8alpha(Alpha) and HLA-A2. *Nature* **1997**, *387*, 630–634.
- (108) Wang, J. H.; Meijers, R.; Xiong, Y.; Liu, J. H.; Sakihama, T.; Zhang, R.; Joachimiak, A.; Reinherz, E. L. Crystal Structure of the Human CD4 N-Terminal Two-Domain Fragment Complexed to a Class II MHC Molecule. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10799–10804.
- (109) Yin, Y.; Wang, X. X.; Mariuzza, R. A. Crystal Structure of a Complete Ternary Complex of T-Cell Receptor, Peptide-MHC, and CD4. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5405–5410.
- (110) Cole, D. K.; Dunn, S. M.; Sami, M.; Boulter, J. M.; Jakobsen, B. K.; Sewell, A. K. T Cell Receptor Engagement of Peptide-Major Histocompatibility Complex Class I Does Not Modify CD8 Binding. *Mol. Immunol.* **2008**, *45*, 2700–2709.
- (111) Van Laethem, F.; Sarafova, S. D.; Park, J.-H.; Tai, X.; Pobeziński, L.; Guinter, T. I.; Adoro, S.; Adams, A.; Sharrow, S. O.; Feigenbaum, L.; Singer, A. Deletion of CD4 and CD8 Coreceptors Permits Generation of Alphabeta T Cells That Recognize Antigens Independently of the MHC. *Immunity* **2007**, *27*, 735–750.
- (112) Wooldridge, L.; Laugel, B.; Ekeruche, J.; Clement, M.; van den Berg, H. A.; Price, D. A.; Sewell, A. K. CD8 Controls T Cell Cross-Reactivity. *J. Immunol.* **2010**, *185*, 4625–4632.
- (113) Hutchinson, S. L.; Wooldridge, L.; Tafuro, S.; Laugel, B.; Glick, M.; Boulter, J. M.; Jakobsen, B. K.; Price, D. A.; Sewell, A. K. The CD8 T Cell Coreceptor Exhibits Disproportionate Biological Activity at Extremely Low Binding Affinities. *J. Biol. Chem.* **2003**, *278*, 24285–24293.
- (114) Bridgeman, J. S.; Sewell, A. K.; Miles, J. J.; Price, D. A.; Cole, D. K. Structural and Biophysical Determinants of Aβ T-Cell Antigen Recognition. *Immunology* **2012**, *135*, 9–18.
- (115) van der Merwe, P. A.; Davis, S. J. Molecular Interactions Mediating T Cell Antigen Recognition. *Annu. Rev. Immunol.* **2003**, *21*, 659–684.
- (116) Schmid, D. a.; Irving, M. B.; Posevitz, V.; Hebeisen, M.; Posevitz-Fejfar, A.; Sarria, J.-C. F.; Gomez-Eerland, R.; Thome, M.; Schumacher, T. N. M.; Romero, P.; Speiser, D. E.; Zoete, V.; Michielin, O.; Rufer, N. Evidence for a TCR Affinity Threshold Delimiting Maximal CD8 T Cell

- Function. *J. Immunol.* **2010**, *184*, 4936–4946.
- (117) van der Merwe, P. A.; Dushek, O. Mechanisms for T Cell Receptor Triggering. *Nat. Rev. Immunol.* **2011**, *11*, 47–55.
- (118) Holland, C. J.; MacLachlan, B. J.; Bianchi, V.; Hesketh, S. J.; Morgan, R.; Vickery, O.; Bulek, A. M.; Fuller, A.; Godkin, A.; Sewell, A. K.; Rizkallah, P. J.; Wells, S.; Cole, D. K. In Silico and Structural Analyses Demonstrate That Intrinsic Protein Motions Guide T Cell Receptor Complementarity Determining Region Loop Flexibility. *Front. Immunol.* **2018**, *9*, 674.
- (119) Willcox, B. E.; Gao, G. F.; Wyer, J. R.; Ladbury, J. E.; Bell, J. I.; Jakobsen, B. K.; van der Merwe, P. A. TCR Binding to Peptide-MHC Stabilizes a Flexible Recognition Interface. *Immunity* **1999**, *10*, 357–365.
- (120) Krosgaard, M.; Prado, N.; Adams, E. J.; He, X. L.; Chow, D. C.; Wilson, D. B.; Garcia, K. C.; Davis, M. M. Evidence That Structural Rearrangements and/or Flexibility during TCR Binding Can Contribute to T Cell Activation. *Mol. Cell* **2003**, *12*, 1367–1378.
- (121) Hawse, W. F.; De, S.; Greenwood, A. I.; Nicholson, L. K.; Zajicek, J.; Kovrigin, E. L.; Kranz, D. M.; Garcia, K. C.; Baker, B. M. TCR Scanning of Peptide/MHC through Complementary Matching of Receptor and Ligand Molecular Flexibility. *J. Immunol.* **2014**, *192*, 2885–2891.
- (122) Scott, D. R.; Borbulevych, O. Y.; Piepenbrink, K. H.; Corcelli, S. A.; Baker, B. M. Disparate Degrees of Hypervariable Loop Flexibility Control T-Cell Receptor Cross-Reactivity, Specificity, and Binding Mechanism. *J. Mol. Biol.* **2011**, *414*, 385–400.
- (123) Ayres, C. M.; Scott, D. R.; Corcelli, S. A.; Baker, B. M. Differential Utilization of Binding Loop Flexibility in T Cell Receptor Ligand Selection and Cross-Reactivity. *Sci. Rep.* **2016**, *6*, 1–14.
- (124) Scott, D. R.; Vardeman, C. F.; Corcelli, S. A.; Baker, B. M. Limitations of Time-Resolved Fluorescence Suggested by Molecular Simulations: Assessing the Dynamics of T Cell Receptor Binding Loops. *Biophys. J.* **2012**, *103*, 2532–2540.
- (125) 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*, e1003748.
- (126) Reboul, C. F.; Meyer, G. R.; Porebski, B. T.; Borg, N. A.; Buckle, A. M. Epitope Flexibility and Dynamic Footprint Revealed by Molecular Dynamics of a PMHC-TCR Complex. *PLoS Comput. Biol.* **2012**, *8*, e1002404.
- (127) Sewell, A. K. Why Must T Cells Be Cross-Reactive? *Nat. Rev. Immunol.* **2012**, *12*, 669–677.
- (128) Mason, D. A Very High Level of Crossreactivity Is an Essential Feature of the T- Cell Receptor. *Immunol. Today* **1998**, *19*, 395–404.
- (129) Cole, D. K.; van den Berg, H. A.; Lloyd, A.; Crowther, M. D.; Beck, K.; Ekeruche-Makinde, J.; Miles, J. J.; Bulek, A. M.; Dolton, G.; Schauenburg, A. J.; Wall, A.; Fuller, A.; Clement, M.; Laugel, B.; Rizkallah, P. J.; Wooldridge, L.; Sewell, A. K. Structural Mechanism Underpinning Cross-Reactivity of a CD8⁺ T-Cell Clone That Recognizes a Peptide Derived from Human Telomerase Reverse Transcriptase. *J. Biol. Chem.* **2017**, *292*, 802–813.
- (130) Birnbaum, M. E.; Mendoza, J. L.; Sethi, D. K.; Dong, S.; Glanville, J.; Dobbins, J.; Ozkan, E.; Davis, M. M.; Wucherpfennig, K. W.; Garcia, K. C. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell* **2014**, *157*, 1073–1087.
- (131) Borbulevych, O. Y.; Piepenbrink, K. H.; Baker, B. M. Conformational Melding Permits a Conserved Binding Geometry in TCR Recognition of Foreign and Self Molecular Mimics. *J. Immunol.* **2011**, *186*, 2950–2958.
- (132) Insaiddo, F. K.; Zajicek, J.; Baker, B. M. A General and Efficient Approach for NMR Studies of

Peptide Dynamics in Class I MHC Peptide Binding Grooves. *Biochemistry* **2009**, *48*, 9708–9710.

- (133) Ayres, C. M.; Corcelli, S. A.; Baker, B. M. Peptide and Peptide-Dependent Motions in MHC Proteins: Immunological Implications and Biophysical Underpinnings. *Front. Immunol.* **2017**, *8*, 1–9.
- (134) Madura, F.; Rizkallah, P. J.; Holland, C. J.; Fuller, A.; Bulek, A.; Godkin, A. J.; Schauenburg, A. J.; Cole, D. K.; Sewell, A. K. Structural Basis for Ineffective T-Cell Responses to MHC Anchor Residue-Improved “Heteroclitic” Peptides. *Eur. J. Immunol.* **2015**, *45*, 584–591.
- (135) Madura, F.; Rizkallah, P. J.; Legut, M.; Holland, C. J.; Fuller, A.; Bulek, A.; Schauenburg, A. J.; Trimby, A.; Hopkins, J. R.; Wells, S. A.; Godkin, A.; Miles, J. J.; Sami, M.; Li, Y.; Liddy, N.; Jakobsen, B. K.; Loveridge, E. J.; Cole, D. K.; Sewell, A. K. TCR-Induced Alteration of Primary MHC Peptide Anchor Residue. *Eur. J. Immunol.* **2019**, *49*, 1052–1066.
- (136) Borbulevych, O. Y.; Piepenbrink, K. H.; Gloor, B. E.; Scott, D. R.; Sommese, R. F.; Cole, D. K.; Sewell, A. K.; Baker, B. M. T Cell Receptor Cross-Reactivity Directed by Antigen-Dependent Tuning of Peptide-MHC Molecular Flexibility. *Immunity* **2009**, *31*, 885–896.
- (137) Reiser, J.-B.; Darnault, C.; Grégoire, C.; Mosser, T.; Mazza, G.; Kearney, A.; van der Merwe, P. A.; Fontecilla-Camps, J. C.; Housset, D.; Malissen, B. CDR3 Loop Flexibility Contributes to the Degeneracy of TCR Recognition. *Nat. Immunol.* **2003**, *4*, 241–247.
- (138) Auphan-Anezin, N.; Mazza, C.; Guimezanes, A.; Barrett-Wilt, G. A.; Montero-Julian, F.; Roussel, A.; Hunt, D. F.; Malissen, B.; Schmitt-Verhulst, A.-M. Distinct Orientation of the Alloreactive Monoclonal CD8 T Cell Activation Program by Three Different Peptide/MHC Complexes. *Eur. J. Immunol.* **2006**, *36*, 1856–1866.
- (139) Hawse, W. F.; Gloor, B. E.; Ayres, C. M.; Kho, K.; Nuter, E.; Baker, B. M. Peptide Modulation of Class I Major Histocompatibility Complex Protein Molecular Flexibility and the Implications for Immune Recognition. *J. Biol. Chem.* **2013**, *288*, 24372–24381.
- (140) Ayres, C. M.; Abualrous, E. T.; Bailey, A.; Abraham, C.; Hellman, L. M.; Corcelli, S. A.; Noé, F.; Elliott, T.; Baker, B. M. Dynamically Driven Allostery in MHC Proteins: Peptide-Dependent Tuning of Class I MHC Global Flexibility. *Front. Immunol.* **2019**, *10*, 1–13.
- (141) Skowera, A.; Ladell, K.; McLaren, J. E.; Dolton, G.; Matthews, K. K.; Gostick, E.; Kronenberg-Versteeg, D.; Eichmann, M.; Knight, R. R.; Heck, S.; Powrie, J.; Bingley, P. J.; Dayan, C. M.; Miles, J. J.; Sewell, A. K.; Price, D. A.; Peakman, M. β -Cell-Specific CD8 T Cell Phenotype in Type 1 Diabetes Reflects Chronic Autoantigen Exposure. *Diabetes* **2015**, *64*, 916–925.
- (142) Skowera, A.; Ellis, R. J.; Varela-Calviño, R.; Arif, S.; Huang, G. C.; Van-Krinks, C.; Zaremba, A.; Rackham, C.; Allen, J. S.; Tree, T. I. M.; Zhao, M.; Dayan, C. M.; Sewell, A. K.; Unger, W. W.; Unger, W.; Drijfhout, J. W.; Ossendorp, F.; Roep, B. O.; Peakman, M. CTLs Are Targeted to Kill Beta Cells in Patients with Type 1 Diabetes through Recognition of a Glucose-Regulated Preproinsulin Epitope. *J. Clin. Invest.* **2008**, *118*, 3390–3402.
- (143) Knight, R. R.; Kronenberg, D.; Zhao, M.; Huang, G. C.; Eichmann, M.; Bulek, A.; Wooldridge, L.; Cole, D. K.; Sewell, A. K.; Peakman, M.; Skowera, A. Human β -Cell Killing by Autoreactive Preproinsulin-Specific CD8 T Cells Is Predominantly Granule-Mediated With the Potency Dependent Upon T-Cell Receptor Avidity. *Diabetes* **2013**, *62*, 205–213.
- (144) Wiedersich, J.; Kohler, S.; Skerra, A.; Friedrich, J. Temperature and Pressure Dependence of Protein Stability: The Engineered Fluorescein-Binding Lipocalin FluA Shows an Elliptic Phase Diagram. *Proc. Natl. Acad. Sci.* **2008**, *105*, 5756–5761.
- (145) Akasaka, K. Probing Conformational Fluctuation of Proteins by Pressure Perturbation. *Chem. Rev.* **2006**, *106*, 1814–1835.

- (146) Catici, D. A. M.; Amos, H. E.; Yang, Y.; van den Elsen, J. M. H.; Pudney, C. R. The Red Edge Excitation Shift Phenomenon Can Be Used to Unmask Protein Structural Ensembles: Implications for NEMO-Ubiquitin Interactions. *FEBS J.* **2016**, *283*, 2272–2284.
- (147) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.
- (148) Thomas, C.; Tampé, R. Structure of the TAPBPR-MHC I Complex Defines the Mechanism of Peptide Loading and Editing. *Science* **2017**, *358*, 1060–1064.
- (149) Jiang, J.; Natarajan, K.; Boyd, L. F.; Morozov, G. I.; Mage, M. G.; Margulies, D. H. Crystal Structure of a TAPBPR–MHC I Complex Reveals the Mechanism of Peptide Editing in Antigen Presentation. *Science* **2017**, *358*, 1064–1068.
- (150) Cole, D. K.; Rizkallah, P. J.; Boulter, J. M.; Sami, M.; Vuidepot, A.; Glick, M.; Gao, F.; Bell, J. I.; Jakobsen, B. K.; Gao, G. F. Computational Design and Crystal Structure of an Enhanced Affinity Mutant Human CD8 A α Coreceptor. *Proteins* **2007**, *67*, 65–74.
- (151) Laugel, B.; van den Berg, H. A.; Gostick, E.; Cole, D. K.; Wooldridge, L.; Boulter, J.; Milicic, A.; Price, D. A.; Sewell, A. K. Different T Cell Receptor Affinity Thresholds and CD8 Coreceptor Dependence Govern Cytotoxic T Lymphocyte Activation and Tetramer Binding Properties. *J. Biol. Chem.* **2007**, *282*, 23799–23810.
- (152) Cole, D. K.; Laugel, B.; Clement, M.; Price, D. A.; Wooldridge, L.; Sewell, A. K. The Molecular Determinants of CD8 Co-Receptor Function. *Immunology* **2012**, *137*, 139–148.
- (153) Valmori, D.; Fonteneau, J. F.; Lizana, C. M.; Gervois, N.; Liénard, D.; Rimoldi, D.; Jongeneel, V.; Jotereau, F.; Cerottini, J. C.; Romero, P. Enhanced Generation of Specific Tumor-Reactive CTL in Vitro by Selected Melan-A/MART-1 Immunodominant Peptide Analogues. *J. Immunol.* **1998**, *160*, 1750–1758.
- (154) Cole, D. K.; Edwards, E. S. J.; Wynn, K. K.; Clement, M.; Miles, J. J.; Ladell, K.; Ekeruche, J.; Gostick, E.; Adams, K. J.; Skowera, A.; Peakman, M.; Wooldridge, L.; Price, D. A.; Sewell, A. K. Modification of MHC Anchor Residues Generates Heteroclitic Peptides That Alter TCR Binding and T Cell Recognition. *J. Immunol.* **2010**, *185*, 2600–2610.
- (155) Garboczi, D. N.; Hung, D. T.; Wiley, D. C. HLA-A2-Peptide Complexes: Refolding and Crystallization of Molecules Expressed in Escherichia Coli and Complexed with Single Antigenic Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 3429–3433.
- (156) Cole, D. K.; Rizkallah, P. J.; Gao, F.; Watson, N. I.; Boulter, J. M.; Bell, J. I.; Sami, M.; Gao, G. F.; Jakobsen, B. K. Crystal Structure of HLA-A*2402 Complexed with a Telomerase Peptide. *Eur. J. Immunol.* **2006**, *36*, 170–179.
- (157) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **2014**, *47*, 5.6.1-5.6.32.
- (158) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- (159) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (160) 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.
- (161) Skjærven, L.; Yao, X.-Q.; Scarabelli, G.; Grant, B. J. Integrating Protein Structural Dynamics and Evolutionary Analysis with Bio3D. *BMC Bioinformatics* **2014**, *15*, 399.
- (162) Csardi, G., & Nepusz, T. The Igraph Software Package for Complex Network Research.

InterJournal **2006**, *Complex Sy*, 1695.

- (163) Yao, X.-Q.; Malik, R. U.; Griggs, N. W.; Skjærven, L.; Traynor, J. R.; Sivaramakrishnan, S.; Grant, B. J. Dynamic Coupling and Allosteric Networks in the α Subunit of Heterotrimeric G Proteins. *J. Biol. Chem.* **2016**, *291*, 4742–4753.
- (164) Heemskerk, B.; Kvistborg, P.; Schumacher, T. N. M. The Cancer Antigenome. *EMBO J.* **2012**, *32*, 194–203.
- (165) Williams, K. A.; Hart, D. N.; Fabre, J. W.; Morris, P. J. Distribution and Quantitation of HLA-ABC and DR (Ia) Antigens on Human Kidney and Other Tissues. *Transplantation* **1980**, *29*, 274–279.
- (166) Berah, M.; Hors, J.; Dausset, J. A Study of HL-A Antigens in Human Organs. *Transplantation* **1970**, *9*, 185–192.
- (167) Irvine, D. J.; Purbhoo, M. A.; Krogsgaard, M.; Davis, M. M. Direct Observation of Ligand Recognition by T Cells. *Nature* **2002**, *419*, 845–849.
- (168) Purbhoo, M. a; Irvine, D. J.; Huppa, J. B.; Davis, M. M. T Cell Killing Does Not Require the Formation of a Stable Mature Immunological Synapse. *Nat. Immunol.* **2004**, *5*, 524–530.
- (169) Brameshuber, M.; Kellner, F.; Rosboth, B. K.; Ta, H.; Alge, K.; Sevcsik, E.; Göhring, J.; Axmann, M.; Baumgart, F.; Gascoigne, N. R. J.; Davis, S. J.; Stockinger, H.; Schütz, G. J.; Huppa, J. B. Monomeric TCRs Drive T Cell Antigen Recognition. *Nat. Immunol.* **2018**, *19*, 487–496.
- (170) Cole, D. K.; Pumphrey, N. J.; Boulter, J. M.; Sami, M.; Bell, J. I.; Gostick, E.; Price, D. A.; Gao, G. F.; Sewell, A. K.; Jakobsen, B. K. Human TCR-Binding Affinity Is Governed by MHC Class Restriction. *J. Immunol.* **2007**, *178*, 5727–5734.
- (171) Rapoport, A. P.; Stadtmauer, E. A.; Binder-Scholl, G. K.; Goloubeva, O.; Vogl, D. T.; Lacey, S. F.; Badros, A. Z.; Garfall, A.; Weiss, B.; Finklestein, J.; Kulikovskaya, I.; Sinha, S. K.; Kronsberg, S.; Gupta, M.; Bond, S.; Melchiori, L.; Brewer, J. E.; Bennett, A. D.; Gerry, A. B.; Pumphrey, N. J.; Williams, D.; Tayton- Martin, H. K.; Ribeiro, L.; Holdich, T.; Yanovich, S.; Hardy, N.; Yared, J.; Kerr, N.; Philip, S.; Westphal, S.; Siegel, D. L.; Levine, B. L.; Jakobsen, B. K.; Kalos, M.; June, C. H. NY-ESO-1-Specific TCR-Engineered T Cells Mediate Sustained Antigen-Specific Antitumor Effects in Myeloma. *Nat. Med.* **2015**, *21*, 914–921.
- (172) Rohaan, M. W.; Wilgenhof, S.; Haanen, J. B. A. G. Adoptive Cellular Therapies: The Current Landscape. *Virchows Arch.* **2019**, *474*, 449–461.
- (173) Husain, B.; Ellerman, D. Expanding the Boundaries of Biotherapeutics with Bispecific Antibodies. *BioDrugs* **2018**, *32*, 441–464.
- (174) Boulter, J. M.; Glick, M.; Todorov, P. T.; Baston, E.; Sami, M.; Rizkallah, P.; Jakobsen, B. K. Stable, Soluble T-Cell Receptor Molecules for Crystallization and Therapeutics. *Protein Eng.* **2003**, *16*, 707–711.
- (175) 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.
- (176) Chang, A. Y.; Dao, T.; Gejman, R. S.; Jarvis, C. A.; Scott, A.; Dubrovsky, L.; Mathias, M. D.; Korontsvit, T.; Zakhaleva, V.; Curcio, M.; Hendrickson, R. C.; Liu, C.; Scheinberg, D. A. A Therapeutic T Cell Receptor Mimic Antibody Targets Tumor-Associated PRAME Peptide/HLA-I Antigens. *J. Clin. Invest.* **2017**, *127*, 2705–2718.
- (177) Maus, M. V; Plotkin, J.; Jakka, G.; Stewart-Jones, G.; Rivière, I.; Merghoub, T.; Wolchok, J.; Renner, C.; Sadelain, M. An MHC-Restricted Antibody-Based Chimeric Antigen Receptor Requires TCR-like Affinity to Maintain Antigen Specificity. *Mol. Ther. - Oncolytics* **2016**, *3*,

- (178) Dao, T.; Pankov, D.; Scott, A.; Korontsvit, T.; Zakhaleva, V.; Xu, Y.; Xiang, J.; Yan, S.; de Moraes Guerreiro, M. D.; Veomett, N.; Dubrovsky, L.; Curcio, M.; Dubrovina, E.; Ponomarev, V.; Liu, C.; O'Reilly, R. J.; Scheinberg, D. A. Therapeutic Bispecific T-Cell Engager Antibody Targeting the Intracellular Oncoprotein WT1. *Nat. Biotechnol.* **2015**, *33*, 1079–1086.
- (179) Zhao, Q.; Ahmed, M.; Tassev, D. V.; Hasan, A.; Kuo, T.-Y.; Guo, H.-F.; O'Reilly, R. J.; Cheung, N.-K. V. Affinity Maturation of T-Cell Receptor-like Antibodies for Wilms Tumor 1 Peptide Greatly Enhances Therapeutic Potential. *Leukemia* **2015**, *29*, 2238–2247.
- (180) Oren, R.; Hod-Marco, M.; Haus-Cohen, M.; Thomas, S.; Blat, D.; Duvshani, N.; Denkberg, G.; Elbaz, Y.; Benchetrit, F.; Eshhar, Z.; Stauss, H.; Reiter, Y. Functional Comparison of Engineered T Cells Carrying a Native TCR versus TCR-like Antibody-Based Chimeric Antigen Receptors Indicates Affinity/Avidity Thresholds. *J. Immunol.* **2014**, *193*, 5733–5743.
- (181) Chames, P.; Willemsen, R. A.; Rojas, G.; Dieckmann, D.; Rem, L.; Schuler, G.; Bolhuis, R. L.; Hoogenboom, H. R. TCR-like Human Antibodies Expressed on Human CTLs Mediate Antibody Affinity-Dependent Cytolytic Activity. *J. Immunol.* **2002**, *169*, 1110–1118.
- (182) Michaeli, Y.; Denkberg, G.; Sinik, K.; Lantzy, L.; Chih-Sheng, C.; Beauverd, C.; Ziv, T.; Romero, P.; Reiter, Y. Expression Hierarchy of T Cell Epitopes from Melanoma Differentiation Antigens: Unexpected High Level Presentation of Tyrosinase-HLA-A2 Complexes Revealed by Peptide-Specific, MHC-Restricted, TCR-like Antibodies. *J. Immunol.* **2009**, *182*, 6328–6341.
- (183) Held, G.; Wadle, A.; Dauth, N.; Stewart-Jones, G.; Sturm, C.; Thiel, M.; Zwick, C.; Dieckmann, D.; Schuler, G.; Hoogenboom, H. R.; Lévy, F.; Cerundolo, V.; Pfreundschuh, M.; Renner, C. MHC-Peptide-Specific Antibodies Reveal Inefficient Presentation of an HLA-A*0201-Restricted, Melanoma-Derived Peptide after Active Intracellular Processing. *Eur. J. Immunol.* **2007**, *37*, 2008–2017.
- (184) Biddison, W. E.; Turner, R. V.; Gagnon, S. J.; Lev, A.; Cohen, C. J.; Reiter, Y. Tax and M1 Peptide/HLA-A2-Specific Fabs and T Cell Receptors Recognize Nonidentical Structural Features on Peptide/HLA-A2 Complexes. *J. Immunol.* **2003**, *171*, 3064–3074.
- (185) Yamano, Y.; Cohen, C. J.; Takenouchi, N.; Yao, K.; Tomaru, U.; Li, H.-C.; Reiter, Y.; Jacobson, S. Increased Expression of Human T Lymphocyte Virus Type I (HTLV-I) Tax11-19 Peptide-Human Histocompatibility Leukocyte Antigen A*201 Complexes on CD4⁺ CD25⁺ T Cells Detected by Peptide-Specific, Major Histocompatibility Complex-Restricted Antibodies in Pati. *J. Exp. Med.* **2004**, *199*, 1367–1377.
- (186) Sastry, K. S. R.; Too, C. T.; Kaur, K.; Gehring, A. J.; Low, L.; Javiad, A.; Pollicino, T.; Li, L.; Kennedy, P. T. F.; Lopatin, U.; Macary, P. A.; Bertoletti, A. Targeting Hepatitis B Virus-Infected Cells with a T-Cell Receptor-Like Antibody. *J. Virol.* **2011**, *85*, 1935–1942.
- (187) Hassan, C.; Kester, M. G. D.; de Ru, A. H.; Hombrink, P.; Drijfhout, J. W.; Nijveen, H.; Leunissen, J. A. M.; Heemskerk, M. H. M.; Falkenburg, J. H. F.; van Veelen, P. A. The Human Leukocyte Antigen-Presented Ligandome of B Lymphocytes. *Mol. Cell. Proteomics* **2013**, *12*, 1829–1843.
- (188) Abelin, J. G.; Keskin, D. B.; Sarkizova, S.; Hartigan, C. R.; Zhang, W.; Sidney, J.; Stevens, J.; Lane, W.; Zhang, G. L.; Eisenhaure, T. M.; Clauser, K. R.; Hacohen, N.; Rooney, M. S.; Carr, S. A.; Wu, C. J. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-Allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **2017**, *46*, 315–326.
- (189) Ritz, D.; Gloger, A.; Weide, B.; Garbe, C.; Neri, D.; Fugmann, T. High-Sensitivity HLA Class I Peptidome Analysis Enables a Precise Definition of Peptide Motifs and the Identification of Peptides from Cell Lines and Patients' Sera. *Proteomics* **2016**, *16*, 1570–1580.
- (190) Shao, W.; Pedrioli, P. G. A.; Wolski, W.; Scurtescu, C.; Schmid, E.; Vizcaíno, J. A.; Courcelles, M.; Schuster, H.; Kowalewski, D.; Marino, F.; Arlehamn, C. S. L.; Vaughan, K.; Peters, B.; Sette, A.; Ottenhoff, T. H. M.; Meijgaarden, K. E.; Nieuwenhuizen, N.; Kaufmann, S. H. E.; Schlapbach,

- R.; Castle, J. C.; Nesvizhskii, A. I.; Nielsen, M.; Deutsch, E. W.; Campbell, D. S.; Moritz, R. L.; Zubarev, R. A.; Ytterberg, A. J.; Purcell, A. W.; Marcilla, M.; Paradela, A.; Wang, Q.; Costello, C. E.; Ternette, N.; van Veelen, P. A.; van Els, C. A. C. M.; Heck, A. J. R.; de Souza, G. A.; Sollid, L. M.; Admon, A.; Stevanovic, S.; Rammensee, H.-G.; Thibault, P.; Perreault, C.; Bassani-Sternberg, M.; Aebersold, R.; Caron, E. The SystemMHC Atlas Project. *Nucleic Acids Res.* **2018**, *46*, D1237–D1247.
- (191) Dhanik, A.; Kirshner, J. R.; MacDonald, D.; Thurston, G.; Lin, H. C.; Murphy, A. J.; Zhang, W. In-Silico Discovery of Cancer-Specific Peptide-HLA Complexes for Targeted Therapy. *BMC Bioinformatics* **2016**, *17*, 1–14.
- (192) Kumar, P.; Vahedi-Faridi, A.; Saenger, W.; Ziegler, A.; Uchanska-Ziegler, B. Conformational Changes within the HLA-A1:MAGE-A1 Complex Induced by Binding of a Recombinant Antibody Fragment with TCR-like Specificity. *Protein Sci.* **2009**, *18*, 37–49.
- (193) Raman, M. C. C.; Rizkallah, P. J.; Simmons, R.; Donnellan, Z.; Dukes, J.; Bossi, G.; Le Provost, G. S.; Todorov, P.; Baston, E.; Hickman, E.; Mahon, T.; Hassan, N.; Vuidepot, A.; Sami, M.; Cole, D. K.; Jakobsen, B. K. Direct Molecular Mimicry Enables Off-Target Cardiovascular Toxicity by an Enhanced Affinity TCR Designed for Cancer Immunotherapy. *Sci. Rep.* **2016**, *6*, 18851.
- (194) Adams, J. J.; Narayanan, S.; Birnbaum, M. E.; Sidhu, S. S.; Blevins, S. J.; Gee, M. H.; Sibener, L. V.; Baker, B. M.; Kranz, D. M.; Garcia, K. C. Structural Interplay between Germline Interactions and Adaptive Recognition Determines the Bandwidth of TCR-Peptide-MHC Cross-Reactivity. *Nat. Immunol.* **2015**, *17*, 87–94.
- (195) Wang, C.; Greene, D.; Xiao, L.; Qi, R.; Luo, R. Recent Developments and Applications of the MMPBSA Method. *Front. Mol. Biosci.* **2018**, *4*, 1–18.
- (196) Lees, W. D.; Stejskal, L.; Moss, D. S.; Shepherd, A. J. Investigating Substitutions in Antibody-Antigen Complexes Using Molecular Dynamics: A Case Study with Broad-Spectrum, Influenza A Antibodies. *Front. Immunol.* **2017**, *8*, 1–11.
- (197) Wong, S.; Amaro, R. E.; McCammon, J. A.; Andrew McCammon, J. MM-PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* **2009**, *5*, 422–429.
- (198) Jaigirdar, A.; Rosenberg, S. A.; Parkhurst, M. A High-Avidity WT1-Reactive T-Cell Receptor Mediates Recognition of Peptide and Processed Antigen but Not Naturally Occurring WT1-Positive Tumor Cells. *J. Immunother.* **2016**, *39*, 105–116.
- (199) Park, J. H.; Geyer, M. B.; Brentjens, R. J. CD19-Targeted CAR T-Cell Therapeutics for Hematologic Malignancies: Interpreting Clinical Outcomes to Date. *Blood* **2016**, *127*, 3312–3320.
- (200) Tikhonova, A. N.; Van Laethem, F.; Hanada, K.; Lu, J.; Pobezinsky, L. A.; Hong, C.; Guintier, T. I.; Jeurling, S. K.; Bernhardt, G.; Park, J.-H.; Yang, J. C.; Sun, P. D.; Singer, A. $\alpha\beta$ T Cell Receptors That Do Not Undergo Major Histocompatibility Complex-Specific Thymic Selection Possess Antibody-like Recognition Specificities. *Immunity* **2012**, *36*, 79–91.
- (201) Scott-Browne, J. P.; White, J.; Kappler, J. W.; Gapin, L.; Marrack, P. Germline-Encoded Amino Acids in the Alphabeta T-Cell Receptor Control Thymic Selection. *Nature* **2009**, *458*, 1043–1046.
- (202) Garcia, K. C.; Adams, J. J.; Feng, D.; Ely, L. K. The Molecular Basis of TCR Germline Bias for MHC Is Surprisingly Simple. *Nat. Immunol.* **2009**, *10*, 143–147.
- (203) Feng, D.; Bond, C. J.; Ely, L. K.; Maynard, J.; Garcia, K. C. Structural Evidence for a Germline-Encoded T Cell Receptor-Major Histocompatibility Complex Interaction “Codon”. *Nat. Immunol.* **2007**, *8*, 975–983.
- (204) 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.
- (205) Cole, D. K.; Miles, K. M.; Madura, F.; Holland, C. J.; Schauenburg, A. J. A. A.; Godkin, A. J.; Bulek, A. M.; Fuller, A.; Akpovwa, H. J. E. E.; Pymm, P. G.; Liddy, N.; Sami, M.; Li, Y.; Rizkallah, P. J.; Jakobsen, B. K.; Sewell, A. K. T-Cell Receptor (TCR)-Peptide Specificity Overrides Affinity-Enhancing TCR-Major Histocompatibility Complex Interactions. *J. Biol. Chem.* **2014**, *289*, 628–638.
- (206) Madura, F.; Rizkallah, P. J.; Miles, K. M.; Holland, C. J.; Bulek, A. M.; Fuller, A.; Schauenburg, A. J. A.; Miles, J. J.; Liddy, N.; Sami, M.; Li, Y.; Hossain, M.; Baker, B. M.; Jakobsen, B. K.; Sewell, A. K.; Cole, D. K. T-Cell Receptor Specificity Maintained by Altered Thermodynamics. *J. Biol. Chem.* **2013**, *288*, 18766–18775.
- (207) Tan, M. P.; Gerry, A. B.; Brewer, J. E.; Melchiori, L.; Bridgeman, J. S.; Bennett, A. D.; Pumphrey, N. J.; Jakobsen, B. K.; Price, D. A.; Ladell, K.; Sewell, A. K. T Cell Receptor Binding Affinity Governs the Functional Profile of Cancer-Specific CD8⁺ T Cells. *Clin. Exp. Immunol.* **2015**, *180*, 255–270.
- (208) van den Berg, H. A.; Ladell, K.; Miners, K.; Laugel, B.; Llewellyn-Lacey, S.; Clement, M.; Cole, D. K.; Gostick, E.; Wooldridge, L.; Sewell, A. K.; Bridgeman, J. S.; Price, D. A. Cellular-Level versus Receptor-Level Response Threshold Hierarchies in T-Cell Activation. *Front. Immunol.* **2013**, *4*, 250.
- (209) McKeithan, T. W. Kinetic Proofreading in T-Cell Receptor Signal Transduction. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5042–5046.
- (210) Aleksic, M.; Dushek, O.; Zhang, H.; Shenderov, E.; Chen, J. L.; Cerundolo, V.; Coombs, D.; van der Merwe, P. A. Dependence of T Cell Antigen Recognition on T Cell Receptor-Peptide MHC Confinement Time. *Immunity* **2010**, *32*, 163–174.
- (211) Inaguma, Y.; Akahori, Y.; Murayama, Y.; Shiraishi, K.; Tsuzuki-Iba, S.; Endoh, A.; Tsujikawa, J.; Demachi-Okamura, A.; Hiramatsu, K.; Saji, H.; Yamamoto, Y.; Yamamoto, N.; Nishimura, Y.; Takahashi, T.; Kuzushima, K.; Emi, N.; Akatsuka, Y. Construction and Molecular Characterization of a T-Cell Receptor-like Antibody and CAR-T Cells Specific for Minor Histocompatibility Antigen HA-1H. *Gene Ther.* **2014**, *21*, 575–584.
- (212) Stadinski, B. D.; Obst, R.; Huseby, E. S. A “Hotspot” for Autoimmune T Cells in Type 1 Diabetes. *J. Clin. Invest.* **2016**, *126*, 2040–2042.
- (213) Trenevskaja, I.; Li, D.; Banham, A. H. Therapeutic Antibodies against Intracellular Tumor Antigens. *Front. Immunol.* **2017**, *8*, 1001.
- (214) Ahmed, M.; Lopez-Albaitero, A.; Pankov, D.; Santich, B. H.; Liu, H.; Yan, S.; Xiang, J.; Wang, P.; Hasan, A. N.; Selvakumar, A.; O’Reilly, R. J.; Liu, C.; Cheung, N.-K. V. TCR-Mimic Bispecific Antibodies Targeting LMP2A Show Potent Activity against EBV Malignancies. *JCI Insight* **2018**, *3*, 1–8.
- (215) Liu, H.; Xu, Y.; Xiang, J.; Long, L.; Green, S.; Yang, Z.; Zimdahl, B.; Lu, J.; Cheng, N.; Horan, L. H.; Liu, B.; Yan, S.; Wang, P.; Diaz, J.; Jin, L.; Nakano, Y.; Morales, J. F.; Zhang, P.; Liu, L.-X.; Staley, B. K.; Priceman, S. J.; Brown, C. E.; Forman, S. J.; Chan, V. W.; Liu, C. Targeting Alpha-Fetoprotein (AFP)-MHC Complex with CAR T-Cell Therapy for Liver Cancer. *Clin. Cancer Res.* **2017**, *23*, 478–488.
- (216) Harper, J.; Adams, K. J.; Bossi, G.; Wright, D. E.; Stacey, A. R.; Bedke, N.; Martinez-Hague, R.; Blat, D.; Humbert, L.; Buchanan, H.; Le Provost, G. S.; Donnellan, Z.; Carreira, R. J.; Paston, S. J.; Weigand, L. U.; Canestraro, M.; Sanderson, J. P.; Botta Gordon-Smith, S.; Lowe, K. L.; Rygiel, K. A.; Powlesland, A. S.; Vuidepot, A.; Hassan, N. J.; Cameron, B. J.; Jakobsen, B. K.; Dukes, J. An Approved in Vitro Approach to Preclinical Safety and Efficacy Evaluation of Engineered T Cell

Receptor Anti-CD3 Bispecific (ImmTAC) Molecules. *PLoS One* **2018**, *13*, e0205491.

- (217) Barbas, C. F. *Phage Display : A Laboratory Manual*; Cold Spring Harbor Laboratory Press, 2001.
- (218) Karlsson, R.; Katsamba, P. S.; Nordin, H.; Pol, E.; Myszka, D. G. Analyzing a Kinetic Titration Series Using Affinity Biosensors. *Anal. Biochem.* **2006**, *349*, 136–147.
- (219) Turchaninova, M. A.; Davydov, A.; Britanova, O. V.; Shugay, M.; Bikos, V.; Egorov, E. S.; Kirgizova, V. I.; Merzlyak, E. M.; Staroverov, D. B.; Bolotin, D. A.; Mamedov, I. Z.; Izraelson, M.; Logacheva, M. D.; Kladova, O.; Plevova, K.; Pospisilova, S.; Chudakov, D. M. High-Quality Full-Length Immunoglobulin Profiling with Unique Molecular Barcoding. *Nat. Protoc.* **2016**, *11*, 1599–1616.
- (220) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved Visualization of Protein Consensus Sequences by IceLogo. *Nat. Methods* **2009**, *6*, 786–787.
- (221) Andreatta, M.; Lund, O.; Nielsen, M. Simultaneous Alignment and Clustering of Peptide Data Using a Gibbs Sampling Approach. *Bioinformatics* **2013**, *29*, 8–14.
- (222) 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, D. M. Y. and P. A. K. AMBER 2016. University of Calafornia: San Francisco 2017.
- (223) Maffucci, I.; Contini, A. Improved Computation of Protein–Protein Relative Binding Energies with the Nwat-MMGBSA Method. *J. Chem. Inf. Model.* **2016**, *56*, 1692–1704.
- (224) Held, G.; Matsuo, M.; Epel, M.; Gnjatic, S.; Ritter, G.; Lee, S. Y.; Tai, T. Y.; Cohen, C. J.; Old, L. J.; Pfreundschuh, M.; Reiter, Y.; Hoogenboom, H. R.; Renner, C. Dissecting Cytotoxic T Cell Responses towards the NY-ESO-1 Protein by Peptide/MHC-Specific Antibody Fragments. *Eur. J. Immunol.* **2004**, *34*, 2919–2929.
- (225) Dao, T.; Yan, S.; Veomett, N.; Pankov, D.; Zhou, L.; Korontsvit, T.; Scott, A.; Whitten, J.; Maslak, P.; Casey, E.; Tan, T.; Liu, H.; Zakhaleva, V.; Curcio, M.; Doubrovina, E.; O'Reilly, R. J.; Liu, C.; Scheinberg, D. A. Targeting the Intracellular WT1 Oncogene Product with a Therapeutic Human Antibody. *Sci. Transl. Med.* **2013**, *5*, 176ra33.
- (226) Chames, P.; Hufton, S. E.; Coulie, P. G.; Uchanska-Ziegler, B.; Hoogenboom, H. R. Direct Selection of a Human Antibody Fragment Directed against the Tumor T-Cell Epitope HLA-A1-MAGE-A1 from a Nonimmunized Phage-Fab Library. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 7969–7974.
- (227) Denkberg, G.; Klechevsky, E.; Reiter, Y. Modification of a Tumor-Derived Peptide at an HLA-A2 Anchor Residue Can Alter the Conformation of the MHC-Peptide Complex: Probing with TCR-like Recombinant Antibodies. *J. Immunol.* **2002**, *169*, 4399–4407.
- (228) Zhang, G.; Wang, L.; Cui, H.; Wang, X.; Zhang, G.; Ma, J.; Han, H.; He, W.; Wang, W.; Zhao, Y.; Liu, C.; Sun, M.; Gao, B. Anti-Melanoma Activity of T Cells Redirected with a TCR-like Chimeric Antigen Receptor. *Sci. Rep.* **2014**, *4*, 3571.
- (229) Lev, A.; Denkberg, G.; Cohen, C. J.; Tzukerman, M.; Skorecki, K. L.; Chames, P.; Hoogenboom, H. R.; Reiter, Y. Isolation and Characterization of Human Recombinant Antibodies Endowed with the Antigen-Specific, Major Histocompatibility Complex-Restricted Specificity of T Cells Directed toward the Widely Expressed Tumor T-Cell Epitopes of the Telomerase Catalytic Sub. *Cancer Res.* **2002**, *62*, 3184–3194.
- (230) Cohen, C. J.; Hoffmann, N.; Farago, M.; Hoogenboom, H. R.; Eisenbach, L.; Reiter, Y. Direct Detection and Quantitation of a Distinct T-Cell Epitope Derived from Tumor-Specific Epithelial Cell-Associated Mucin Using Human Recombinant Antibodies Endowed with the Antigen-

- Specific, Major Histocompatibility Complex-Restricted Specificity of T . *Cancer Res.* **2002**, *62*, 5835–5844.
- (231) Bernardeau, K.; Gouard, S.; David, G.; Ruellan, A.-L.; Devys, A.; Barbet, J.; Bonneville, M.; Chérel, M.; Davodeau, F. Assessment of CD8 Involvement in T Cell Clone Avidity by Direct Measurement of HLA-A2/Mage3 Complex Density Using a High-Affinity TCR like Monoclonal Antibody. *Eur. J. Immunol.* **2005**, *35*, 2864–2875.
- (232) Wittman, V. P.; Woodburn, D.; Nguyen, T.; Neethling, F. A.; Wright, S.; Weidanz, J. A. Antibody Targeting to a Class I MHC-Peptide Epitope Promotes Tumor Cell Death. *J. Immunol.* **2006**, *177*, 4187–4195.
- (233) Neethling, F. A.; Ramakrishna, V.; Keler, T.; Buchli, R.; Woodburn, T.; Weidanz, J. A. Assessing Vaccine Potency Using TCRmimic Antibodies. *Vaccine* **2008**, *26*, 3092–3102.
- (234) Weidanz, J. A.; Nguyen, T.; Woodburn, T.; Neethling, F. A.; Chiriva-Internati, M.; Hildebrand, W. H.; Lustgarten, J. Levels of Specific Peptide-HLA Class I Complex Predicts Tumor Cell Susceptibility to CTL Killing. *J. Immunol.* **2006**, *177*, 5088–5097.
- (235) Klechevsky, E.; Gallegos, M.; Denkberg, G.; Palucka, K.; Banchereau, J.; Cohen, C.; Reiter, Y. Antitumor Activity of Immunotoxins with T-Cell Receptor-like Specificity against Human Melanoma Xenografts. *Cancer Res.* **2008**, *68*, 6360–6367.
- (236) Epel, M.; Carmi, I.; Soueid-Baumgarten, S.; Oh, S.; Bera, T.; Pastan, I.; Berzofsky, J.; Reiter, Y. Targeting TARP, a Novel Breast and Prostate Tumor-Associated Antigen, with T Cell Receptor-like Human Recombinant Antibodies. *Eur. J. Immunol.* **2008**, *38*, 1706–1720.
- (237) Li, D.; Bentley, C.; Anderson, A.; Wiblin, S.; Cleary, K. L. S.; Koustoulidou, S.; Hassanali, T.; Yates, J.; Greig, J.; Nordkamp, M. O.; Trenevskaja, I.; Ternette, N.; Kessler, B. M.; Cornelissen, B.; Cragg, M. S.; Banham, A. H. Development of a T-Cell Receptor Mimic Antibody against Wild-Type P53 for Cancer Immunotherapy. *Cancer Res.* **2017**, *77*, 2699–2711.
- (238) Li, D.; Bentley, C.; Yates, J.; Salimi, M.; Greig, J.; Wiblin, S.; Hassanali, T.; Banham, A. H. Engineering Chimeric Human and Mouse Major Histocompatibility Complex (MHC) Class I Tetramers for the Production of T-Cell Receptor (TCR) Mimic Antibodies. *PLoS One* **2017**, *12*, e0176642.
- (239) Verma, B.; Hawkins, O. E.; Neethling, F. A.; Caseltine, S. L.; Largo, S. R.; Hildebrand, W. H.; Weidanz, J. A. Direct Discovery and Validation of a Peptide/MHC Epitope Expressed in Primary Human Breast Cancer Cells Using a TCRm Monoclonal Antibody with Profound Antitumor Properties. *Cancer Immunol. Immunother.* **2010**, *59*, 563–573.
- (240) Hawkins, O.; Verma, B.; Lightfoot, S.; Jain, R.; Rawat, A.; McNair, S.; Caseltine, S.; Mojsilovic, A.; Gupta, P.; Neethling, F.; Almanza, O.; Dooley, W.; Hildebrand, W.; Weidanz, J. An HLA-Presented Fragment of Macrophage Migration Inhibitory Factor Is a Therapeutic Target for Invasive Breast Cancer. *J. Immunol.* **2011**, *186*, 6607–6616.
- (241) Sergeeva, A.; Alatrash, G.; He, H.; Ruisaard, K.; Lu, S.; Wygant, J.; McIntyre, B. W.; Ma, Q.; Li, D.; St John, L.; Clise-Dwyer, K.; Mollidrem, J. J. An Anti-PR1/HLA-A2 T-Cell Receptor-like Antibody Mediates Complement-Dependent Cytotoxicity against Acute Myeloid Leukemia Progenitor Cells. *Blood* **2011**, *117*, 4262–4272.
- (242) Harris, D. T.; Wang, N.; Riley, T. P.; Anderson, S. D.; Singh, N. K.; Procko, E.; Baker, B. M.; Kranz, D. M. Deep Mutational Scans as a Guide to Engineering High Affinity T Cell Receptor Interactions with Peptide-Bound Major Histocompatibility Complex. *J. Biol. Chem.* **2016**, *291*, 24566–24578.
- (243) Chervin, A. S.; Aggen, D. H.; Raseman, J. M.; Kranz, D. M. Engineering Higher Affinity T Cell Receptors Using a T Cell Display System. *J. Immunol. Methods* **2008**, *339*, 175–184.

- (244) Richman, S. A.; Healan, S. J.; Weber, K. S.; Donermeyer, D. L.; Dossett, M. L.; Greenberg, P. D.; Allen, P. M.; Kranz, D. M. Development of a Novel Strategy for Engineering High-Affinity Proteins by Yeast Display. *Protein Eng. Des. Sel.* **2006**, *19*, 255–264.
- (245) 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.
- (246) Borbulevych, O. Y.; Insaïdoo, F. K.; Baxter, T. K.; Powell, D. J.; Johnson, L. A.; Restifo, N. P.; Baker, B. M. Structures of MART-126/27–35 Peptide/HLA-A2 Complexes Reveal a Remarkable Disconnect between Antigen Structural Homology and T Cell Recognition. *J. Mol. Biol.* **2007**, *372*, 1123–1136.
- (247) Jeliaskov, J. R.; Sljoka, A.; Kuroda, D.; Tsuchimura, N.; Katoh, N.; Tsumoto, K.; Gray, J. J. Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification. *Front. Immunol.* **2018**, *9*, 413.
- (248) Ma, B.; Nussinov, R. Trp/Met/Phe Hot Spots in Protein-Protein Interactions: Potential Targets in Drug Design. *Curr. Top. Med. Chem.* **2007**, *7*, 999–1005.
- (249) Borbulevych, O. Y.; Santhanagopalan, S. M.; Hossain, M.; Baker, B. M. TCRs Used in Cancer Gene Therapy Cross-React with MART-1/Melan-A Tumor Antigens via Distinct Mechanisms. *J. Immunol.* **2011**, *187*, 2453–2463.
- (250) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230.
- (251) Leem, J.; de Oliveira, S. H. P.; Krawczyk, K.; Deane, C. M. STCRDab: The Structural T-Cell Receptor Database. *Nucleic Acids Res.* **2018**, *46*, D406–D412.
- (252) Chen, F.; Liu, H.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. 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.* **2016**, *18*, 22129–22139.
- (253) Sun, H.; Duan, L.; Chen, F.; Liu, H.; Wang, Z.; Pan, P.; Zhu, F.; Zhang, J. Z. H.; Hou, T. 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.* **2018**, *20*, 14450–14460.
- (254) Hoppes, R.; Oostvogels, R.; Luimstra, J. J.; Wals, K.; Toebes, M.; Bies, L.; Ekkebus, R.; Rijal, P.; Celie, P. H. N.; Huang, J. H.; Emmelot, M. E.; Spaapen, R. M.; Lokhorst, H.; Schumacher, T. N. M.; Mutis, T.; Rodenko, B.; Ovaa, H. Altered Peptide Ligands Revisited: Vaccine Design through Chemically Modified HLA-A2-Restricted T Cell Epitopes. *J. Immunol.* **2014**, *193*, 4803–4813.
- (255) 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*, 1221–1230.
- (256) Perilla, J. R.; Schulten, K. Physical Properties of the HIV-1 Capsid from All-Atom Molecular Dynamics Simulations. *Nat. Commun.* **2017**, *8*, 15959.
- (257) Javanainen, M.; Martinez-Seara, H.; Vattulainen, I. Excessive Aggregation of Membrane Proteins in the Martini Model. *PLoS One* **2017**, *12*, e0187936.
- (258) Boyle, L. H.; Hermann, C.; Boname, J. M.; Porter, K. M.; Patel, P. A.; Burr, M. L.; Duncan, L. M.; Harbour, M. E.; Rhodes, D. A.; Skjødt, K.; Lehner, P. J.; Trowsdale, J. Tapasin-Related Protein TAPBPR Is an Additional Component of the MHC Class I Presentation Pathway. *Proc. Natl. Acad.*

- Sci.* **2013**, *110*, 3465–3470.
- (259) Thomas, C.; Tampé, R. Structure of the TAPBPR–MHC I Complex Defines the Mechanism of Peptide Loading and Editing. *Science* **2017**, *358*, 1060–1064.
- (260) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.
- (261) Meli, M.; Colombo, G. A Hamiltonian Replica Exchange Molecular Dynamics (MD) Method for the Study of Folding, Based on the Analysis of the Stabilization Determinants of Proteins. *Int. J. Mol. Sci.* **2013**, *14*, 12157–12169.
- (262) Kiyota, Y.; Yoshida, N.; Hirata, F. A New Approach for Investigating the Molecular Recognition of Protein: Toward Structure-Based Drug Design Based on the 3D-RISM Theory. *J. Chem. Theory Comput.* **2011**, *7*, 3803–3815.
- (263) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10*, 2769–2780.
- (264) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137*, 14930–14943.
- (265) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A2Aantagonists as a Case Study. *J. Chem. Inf. Model.* **2013**, *53*, 1700–1713.
- (266) Robinson, D.; Bertrand, T.; Carry, J. C.; Halley, F.; Karlsson, A.; Mathieu, M.; Minoux, H.; Perrin, M. A.; Robert, B.; Schio, L.; Sherman, W. Differential Water Thermodynamics Determine PI3K-Beta/Delta Selectivity for Solvent-Exposed Ligand Modifications. *J. Chem. Inf. Model.* **2016**, *56*, 886–894.
- (267) Haider, K.; Huggins, D. J. Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 2571–2586.
- (268) Riley, T. P.; Ayres, C. M.; Hellman, L. M.; Singh, N. K.; Cosiano, M.; Cimons, J. M.; Anderson, M. J.; Piepenbrink, K. H.; Pierce, B. G.; Weng, Z.; Baker, B. M. A Generalized Framework for Computational Design and Mutational Scanning of T-Cell Receptor Binding Interfaces. *Protein Eng. Des. Sel.* **2016**, *29*, 595–606.
- (269) 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.
- (270) Cook, P. F. Mechanism from Isotope Effects. *Isotopes Environ. Health Stud.* **1998**, *34*, 3–17.
- (271) Schwartz, S. D.; Schramm, V. L. Enzymatic Transition States and Dynamic Motion in Barrier Crossing. *Nat. Chem. Biol.* **2009**, *5*, 551–558.
- (272) Allemann, R. K.; Evans, R. M.; Tey, L.; Maglia, G.; Pang, J.; Rodriguez, R.; Shrimpton, P. J.; Swanwick, R. S. Protein Motions during Catalysis by Dihydrofolate Reductases. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2006**, *361*, 1317–1321.
- (273) Luk, L. Y. P.; Javier Ruiz-Pernia, J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tunon, I.; Moliner, V.; Allemann, R. K. Unraveling the Role of Protein Dynamics in Dihydrofolate Reductase Catalysis. *Proc. Natl. Acad. Sci.* **2013**, *110*, 16344–16349.
- (274) Thornton, E. R. Physical Organic Chemistry. *Annu. Rev. Phys. Chem.* **1966**, *17*, 349–372.

- (275) Knapp, M. J.; Klinman, J. P. Environmentally Coupled Hydrogen Tunneling. *Eur. J. Biochem.* **2002**, *269*, 3113–3121.
- (276) Knapp, M. J.; Rickert, K.; Klinman, J. P. Temperature-Dependent Isotope Effects in Soybean Lipoxygenase-1: Correlating Hydrogen Tunneling with Protein Dynamics. *J. Am. Chem. Soc.* **2002**, *124*, 3865–3874.
- (277) Roy, S.; Schopf, P.; Warshel, A. Origin of the Non-Arrhenius Behavior of the Rates of Enzymatic Reactions. *J. Phys. Chem. B* **2017**, *121*, 6520–6526.
- (278) Francis, K.; Sapienza, P. J.; Lee, A. L.; Kohen, A. The Effect of Protein Mass Modulation on Human Dihydrofolate Reductase. *Biochemistry* **2016**, *55*, 1100–1106.
- (279) Wang, Z.; Singh, P.; Czekster, C. M.; Kohen, A.; Schramm, V. L. Protein Mass-Modulated Effects in the Catalytic Mechanism of Dihydrofolate Reductase: Beyond Promoting Vibrations. *J. Am. Chem. Soc.* **2014**, *136*, 8333–8341.
- (280) Kipp, D. R.; Silva, R. G.; Schramm, V. L. Mass-Dependent Bond Vibrational Dynamics Influence Catalysis by HIV-1 Protease. *J. Am. Chem. Soc.* **2011**, *133*, 19358–19361.
- (281) Soriano, A.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V.; Bertrán, J. Electrostatic Effects in Enzyme Catalysis: A Quantum Mechanics/Molecular Mechanics Study of the Nucleophilic Substitution Reaction in Haloalkane Dehalogenase. *Theor. Chem. Acc.* **2004**, *112*, 327–334.
- (282) García-Meseguer, R.; Martí, S.; Ruiz-Pernía, J. J.; Moliner, V.; Tuñón, I. Studying the Role of Protein Dynamics in an SN2 Enzyme Reaction Using Free-Energy Surfaces and Solvent Coordinates. *Nat. Chem.* **2013**, *5*, 566–571.
- (283) Olsson, M. H. M.; Parson, W. W.; Warshel, A. Dynamical Contributions to Enzyme Catalysis: Critical Tests of A Popular Hypothesis. *Chem. Rev.* **2006**, *106*, 1737–1756.
- (284) Åqvist, J.; Kazemi, M.; Isaksen, G. V.; Brandsdal, B. O. Entropy and Enzyme Catalysis. *Acc. Chem. Res.* **2017**, *50*, 199–207.
- (285) Schramm, V. L.; Schwartz, S. D. Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* **2018**, *57*, 3299–3308.
- (286) Warshel, A.; Bora, R. P. Perspective: Defining and Quantifying the Role of Dynamics in Enzyme Catalysis. *J. Chem. Phys.* **2016**, *144*, 180901.
- (287) van der Kamp, M. W.; Prentice, E. J.; Kraakman, K. L.; Connolly, M.; Mulholland, A. J.; Arcus, V. L. Dynamical Origins of Heat Capacity Changes in Enzyme-Catalysed Reactions. *Nat. Commun.* **2018**, *9*, 1177.
- (288) Jones, H. B.; Wells, S. A.; Prentice, E. J.; Kwok, A.; Liang, L. L.; Arcus, V. L.; Pudney, C. R. A Complete Thermodynamic Analysis of Enzyme Turnover Links the Free Energy Landscape to Enzyme Catalysis. *FEBS J.* **2017**, *284*, 2829–2842.
- (289) Arcus, V. L.; Pudney, C. R. Change in Heat Capacity Accurately Predicts Vibrational Coupling in Enzyme Catalyzed Reactions. *FEBS Lett.* **2015**, *589*, 2200–2206.
- (290) Hobbs, J. K.; Jiao, W.; Easter, A. D.; Parker, E. J.; Schipper, L. A.; Arcus, V. L. Change in Heat Capacity for Enzyme Catalysis Determines Temperature Dependence of Enzyme Catalyzed Rates. *ACS Chem. Biol.* **2013**, *8*, 2388–2393.
- (291) Arcus, V. L.; Prentice, E. J.; Hobbs, J. K.; Mulholland, A. J.; Van der Kamp, M. W.; Pudney, C. R.; Parker, E. J.; Schipper, L. A. On the Temperature Dependence of Enzyme-Catalyzed Rates. *Biochemistry* **2016**, *55*, 1681–1688.
- (292) Daniel, R. M.; Danson, M. J.; Eisenthal, R.; Lee, C. K.; Peterson, M. E. The Effect of Temperature

- on Enzyme Activity: New Insights and Their Implications. *Extremophiles* **2008**, *12*, 51–59.
- (293) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. Taking Ockham's Razor to Enzyme Dynamics and Catalysis. *Nat. Chem.* **2012**, *4*, 169–176.
- (294) Bigeleisen, J. Effect of Isotopic Substitution on the Entropy, Enthalpy, and Heat Capacity of Ideal Gases. I. Systems in Thermal Equilibrium. II. Chemically Reacting Systems. *J. Chem. Phys.* **1953**, *21*, 1333–1339.
- (295) Tjahjono, M.; Garland, M. Deuterium Isotope Effect on Molar Heat Capacities and Apparent Molar Heat Capacities in Dilute Aqueous Solutions: A Multi-Channel Heat-Flow Microcalorimeter Study. *J. Chem. Thermodyn.* **2008**, *40*, 1600–1606.
- (296) Ibberson, R. M.; David, W. I. F.; Yamamuro, O.; Miyoshi, Y.; Matsuo, T.; Suga, H. Calorimetric, Dielectric, and Neutron Diffraction Studies on Phase Transitions in Ordinary and Deuterated Acetone Crystals. *J. Phys. Chem.* **1995**, *99*, 14167–14173.
- (297) Longbotham, J. E.; Hardman, S. J. O.; Gö, S.; Scrutton, N. S.; Hay, S.; Görlich, S.; Scrutton, N. S.; Hay, S. Untangling Heavy Protein and Cofactor Isotope Effects on Enzyme-Catalyzed Hydride Transfer. **2016**, *138*, 13693–13699.
- (298) Lamble, H. J.; Heyer, N. I.; Bull, S. D.; Hough, D. W.; Danson, M. J. Metabolic Pathway Promiscuity in the Archaeon *Sulfolobus Solfataricus* Revealed by Studies on Glucose Dehydrogenase and 2-Keto-3-Deoxygluconate Aldolase. *J. Biol. Chem.* **2003**, *278*, 34066–34072.
- (299) Milburn, C. C.; Lamble, H. J.; Theodossis, A.; Bull, S. D.; Hough, D. W.; Danson, M. J.; Taylor, G. L. The Structural Basis of Substrate Promiscuity in Glucose Dehydrogenase from the Hyperthermophilic Archaeon *Sulfolobus Solfataricus*. *J. Biol. Chem.* **2006**, *281*, 14796–14804.
- (300) Blomberg, M. R. a; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601–3658.
- (301) Giardina, P.; De Biasi, M.-G.; De Rosa, M.; Gambacortat, A.; Buonocore, V. Glucose Dehydrogenase from the Thermoacidophilic Archaeobacterium *Sulfolobus Solfataricus*. *Biochem. J* **1986**, *239*, 517–522.
- (302) Pudney, C. R.; Johannissen, L. O.; Sutcliffe, M. J.; Hay, S.; Scrutton, N. S. Direct Analysis of Donor–Acceptor Distance and Relationship to Isotope Effects and the Force Constant for Barrier Compression in Enzymatic H-Tunneling Reactions. *J. Am. Chem. Soc.* **2010**, *132*, 11329–11335.
- (303) Moser, C. C.; Dutton, P. L. Engineering Protein Structure for Electron Transfer Function in Photosynthetic Reaction Centers. *Biochimica Biophys. Acta* **1992**, *01*, 171–176.
- (304) Huskey, W. P.; Schowen, R. L. Reaction-Coordinate Tunneling in Hydride-Transfer Reactions. *J. Am. Chem. Soc.* **1983**, *105*, 5704–5706.
- (305) Hay, S.; Pudney, C. R.; Sutcliffe, M. J.; Scrutton, N. S. Probing Active Site Geometry Using High Pressure and Secondary Isotope Effects in an Enzyme-Catalysed 'Deep' H-Tunnelling Reaction. *J. Phys. Org. Chem.* **2010**, *23*, 696–701.
- (306) Pudney, C. R.; Guerriero, A.; Baxter, N. J.; Johannissen, L. O.; Waltho, J. P.; Hay, S.; Scrutton, N. S. Fast Protein Motions Are Coupled to Enzyme H-Transfer Reactions. *J. Am. Chem. Soc.* **2013**, *135*, 2512–2517.
- (307) Antoniou, D.; Ge, X.; Schramm, V. L.; Schwartz, S. D. Mass Modulation of Protein Dynamics Associated with Barrier Crossing in Purine Nucleoside Phosphorylase. *J. Phys. Chem. Lett.* **2012**, *3*, 3538–3544.
- (308) Ranasinghe, C.; Guo, Q.; Sapienza, P. J.; Lee, A. L.; Quinn, D. M.; Cheatum, C. M.; Kohen, A. Protein Mass Effects on Formate Dehydrogenase. *J. Am. Chem. Soc.* **2017**, *139*, 17405–17413.

- (309) Jordi, V.; Warshel, A. Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **2001**, *105*, 7887–7907.
- (310) Pislakov, A. V.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Enzyme Millisecond Conformational Dynamics Do Not Catalyze the Chemical Step. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 17359–17364.
- (311) Boekelheide, N.; Salomón-Ferrer, R.; Miller, T. F. Dynamics and Dissipation in Enzyme Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16159–16163.
- (312) Loveridge, E. J.; Behiry, E. M.; Guo, J.; Allemann, R. K. Evidence That a ‘Dynamic Knockout’ in Escherichia Coli Dihydrofolate Reductase Does Not Affect the Chemical Step of Catalysis. *Nat. Chem.* **2012**, *4*, 292–297.
- (313) Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. Protein Identification and Analysis Tools in the ExPASy Server. *Methods Mol. Biol.* **1999**, *112*, 531–552.
- (314) Mikkelsen, K.; Nielsen, S. O. ACIDITY MEASUREMENTS WITH THE GLASS ELECTRODE IN H₂O-D₂O MIXTURES. *J. Phys. Chem.* **1960**, *64*, 632–637.
- (315) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 778–795.
- (316) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **2014**, *47*, 5.6.1-5.6.32.
- (317) Holmberg, N.; Ryde, U.; Bülow, L. Redesign of the Coenzyme Specificity in L-Lactate Dehydrogenase from Bacillus Stearothermophilus Using Site-Directed Mutagenesis and Media Engineering. *Protein Eng.* **1999**, *12*, 851–856.
- (318) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947.
- (319) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D. J.; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Fox, D. J.; et al. Gaussian 16 Revision 16.A.03. 2016.
- (320) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (321) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (322) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887.
- (323) Wigner, E. P. Crossing of Potential Thresholds in Chemical Reactions. *Z. Phys. Chem.* **1932**, *B19*, 203–216.
- (324) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. In *Current Protocols in*

Bioinformatics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, NJ, USA, 2006; Vol. 24, pp 1–15.

- (325) Binda, C.; Newton-Vinson, P.; Hubálek, F.; Edmondson, D. E.; Mattevi, A. Structure of Human Monoamine Oxidase B, a Drug Target for the Treatment of Neurological Disorders. *Nat. Struct. Biol.* **2002**, *9*, 22–26.
- (326) Schnaitman, C.; Erwin, V. G.; Greenawalt, J. W. The Submitochondrial Localization of Monoamine Oxidase. An Enzymatic Marker for the Outer Membrane of Rat Liver Mitochondria. *J. Cell Biol.* **1967**, *32*, 719–735.
- (327) Knoll, J.; Ecseri, Z.; Kelemen, K.; Nievel, J.; Knoll, B. Phenylisopropylmethylpropinylamine (E-250), a New Spectrum Psychic Energizer. *Arch. Int. Pharmacodyn. Ther.* **1965**, *155*, 154–164.
- (328) Riederer, P.; Laux, G. MAO-Inhibitors in Parkinson's Disease. *Exp. Neurobiol.* **2011**, *20*, 1–17.
- (329) Jonsson, T.; Edmondson, D. E.; Klinman, J. P. Hydrogen Tunneling in the Flavoenzyme Monoamine Oxidase B. *Biochemistry* **1994**, *33*, 14871–14878.
- (330) Nagel, Z. D.; Klinman, J. P. Tunneling and Dynamics in Enzymatic Hydride Transfer. *Chem. Rev.* **2006**, *106*, 3095–3118.
- (331) MacMillar, S.; Edmondson, D. E.; Matsson, O. Nitrogen Kinetic Isotope Effects for the Monoamine Oxidase B-Catalyzed Oxidation of Benzylamine and (1,1-2H₂) Benzylamine: Nitrogen Rehybridization and CH Bond Cleavage Are Not Concerted. *J. Am. Chem. Soc.* **2011**, *133*, 12319–12321.
- (332) Silverman, R. B. Radical Ideas about Monoamine Oxidase. *Acc. Chem. Res.* **1995**, *28*, 335–342.
- (333) Akyüz, M. A.; Erdem, S. S. Computational Modeling of the Direct Hydride Transfer Mechanism for the MAO Catalyzed Oxidation of Phenethylamine and Benzylamine: ONIOM (QM/QM) Calculations. *J. Neural Transm.* **2013**, *120*, 937–945.
- (334) Kurtz, K. A. K. A.; Rishavy, M. A. M. A.; Cleland, W. W. W.; Fitzpatrick, P. F. Nitrogen Isotope Effects As Probes of the Mechanism of D-Amino Acid Oxidase. *J. Am. Chem. Soc.* **2000**, *122*, 12896–12897.
- (335) Fitzpatrick, P. F. Oxidation of Amines by Flavoproteins. *Arch. Biochem. Biophys.* **2010**, *493*, 13–25.
- (336) Vianello, R.; Repič, M.; Mavri, J. How Are Biogenic Amines Metabolized by Monoamine Oxidases? *European J. Org. Chem.* **2012**, *2012*, 7057–7065.
- (337) Walker, M. C.; Edmondson, D. E. Structure-Activity Relationships in the Oxidation of Benzylamine Analogs by Bovine Liver Mitochondrial Monoamine Oxidase B. *Biochemistry* **1994**, *33*, 7088–7098.
- (338) Abad, E.; Zenn, R. K.; Kästner, J. Reaction Mechanism of Monoamine Oxidase from QM/MM Calculations. *J. Phys. Chem. B* **2013**, *117*, 14238–14246.
- (339) Murray, A. T.; Dowley, M. J. H.; Pradaux-Caggiano, F.; Baldansuren, A.; Fielding, A. J.; Tuna, F.; Hendon, C. H.; Walsh, A.; Lloyd-Jones, G. C.; John, M. P.; Carbery, D. R. Catalytic Amine Oxidation under Ambient Aerobic Conditions: Mimicry of Monoamine Oxidase B. *Angew. Chemie - Int. Ed.* **2015**, *54*, 8997–9000.
- (340) Rigby, S. E. J.; Hynson, R. M. G.; Ramsay, R. R.; Munro, A. W.; Scrutton, N. S. A Stable Tyrosyl Radical in Monoamine Oxidase A. *J. Biol. Chem.* **2005**, *280*, 4627–4631.
- (341) Dunn, R. V.; Munro, A. W.; Turner, N. J.; Rigby, S. E. J.; Scrutton, N. S. Tyrosyl Radical Formation and Propagation in Flavin Dependent Monoamine Oxidases. *ChemBioChem* **2010**, *11*, 1228–1231.
- (342) Newton-Vinson, P.; Hubalek, F.; Edmondson, D. E. High-Level Expression of Human Liver

Monoamine Oxidase B in *Pichia Pastoris*. *Protein Expr. Purif.* **2000**, *20*, 334–345.

- (343) Yue, K. T.; Bhattacharyya, A. K.; Zhelyaskov, V. R.; Edmondson, D. E. Resonance Raman Spectroscopic Evidence for an Anionic Flavin Semiquinone in Bovine Liver Monoamine Oxidase. *Arch. Biochem. Biophys.* **1993**, *300*, 178–185.
- (344) DeRose, V. J.; Woo, J. C. G.; Hawe, W. P.; Hoffman, B. M.; Silverman, R. B.; Yelekci, K. Observation of a Flavin Semiquinone in the Resting State of Monoamine Oxidase B by Electron Paramagnetic Resonance and Electron Nuclear Double Resonance Spectroscopy. *Biochemistry* **1996**, *35*, 11085–11091.
- (345) Allen, W. J.; Bevan, D. R. Steered Molecular Dynamics Simulations Reveal Important Mechanisms in Reversible Monoamine Oxidase B Inhibition. *Biochemistry* **2011**, *50*, 6441–6454.
- (346) Is, Y. S.; Durdagi, S.; Aksoydan, B.; Yurtsever, M. Proposing Novel MAO-B Hit Inhibitors Using Multidimensional Molecular Modeling Approaches and Application of Binary QSAR Models for Prediction of Their Therapeutic Activity, Pharmacokinetic and Toxicity Properties. *ACS Chem. Neurosci.* **2018**, *9*, 1768–1782.
- (347) Braun, G. H.; M Jorge, D. M.; Ramos, H. P.; Alves, R. M.; da Silva, V. B.; Giuliatti, S.; Vilela Sampaio, S.; Taft, C. A.; T P Silva, C. H. Molecular Dynamics, Flexible Docking, Virtual Screening, ADMET Predictions, and Molecular Interaction Field Studies to Design Novel Potential MAO-B Inhibitors. *J. Biomol. Struct. Dyn.* **2008**, *254*, 347–355.
- (348) La Regina, G.; Silvestri, R.; Artico, M.; Lavecchia, A.; Novellino, E.; Befani, O.; Turini, P.; Agostinelli, E. New Pyrrole Inhibitors of Monoamine Oxidase: Synthesis, Biological Evaluation, and Structural Determinants of MAO-A and MAO-B Selectivity. *J. Med. Chem.* **2007**, *50*, 922–931.
- (349) Maccioni, E.; Alcaro, S.; Cirilli, R.; Vigo, S.; Cardia, M. C.; Sanna, M. L.; Meleddu, R.; Yanez, M.; Costa, G.; Casu, L.; Matyus, P.; Distinto, S. 3-Acetyl-2,5-Diaryl-2,3-Dihydro-1,3,4-Oxadiazoles: A New Scaffold for the Selective Inhibition of Monoamine Oxidase B. *J. Med. Chem.* **2011**, *54*, 6394–6398.
- (350) Chimenti, F.; Maccioni, E.; Secci, D.; Bolasco, A.; Chimenti, P.; Granese, A.; Befani, O.; Turini, P.; Alcaro, S.; Ortuso, F.; Cirilli, R.; Torre, F. La; Cardia, M. C.; Distinto, S. Synthesis, Molecular Modeling Studies, and Selective Inhibitory Activity against Monoamine Oxidase of 1-Thiocarbamoyl-3,5-Diaryl-4,5-Dihydro-(1H)- Pyrazole Derivatives. *J. Med. Chem.* **2005**, *48*, 7113–7122.
- (351) Fowler, P. W.; Balali-Mood, K.; Deol, S.; Coveney, P. V.; Sansom, M. S. P. Monotopic Enzymes and Lipid Bilayers: A Comparative Study. *Biochemistry* **2007**, *46*, 3108–3115.
- (352) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567.
- (353) Kao, Y. T.; Saxena, C.; He, T. F.; Guo, L.; Wang, L.; Sancar, A.; Zhong, D. Ultrafast Dynamics of Flavins in Five Redox States. *J. Am. Chem. Soc.* **2008**, *130*, 13132–13139.
- (354) Woo, J. C. G.; Silverman, R. B. Observation of Two Different Chromophores in the Resting State of Monoamine Oxidase B by Fluorescence Spectroscopy. *Biochem. Biophys. Res. Commun.* **1994**, *202*, 1574–1578.
- (355) Candeias, L. P.; Turconi, S.; Nugent, J. H. A. Tyrosine Y(Z) and Y(D) of Photosystem II Comparison of Optical Spectra to Those of Tyrosine Oxidised by Pulsed Radiolysis. *Biochim. Biophys. Acta - Bioenerg.* **1998**, *1363*, 1–5.
- (356) Geha, R. M.; Chen, K.; Wouters, J.; Ooms, F.; Shih, J. C. Analysis of Conserved Active Site

- Residues in Monoamine Oxidase A and B and Their Three-Dimensional Molecular Modeling. *J. Biol. Chem.* **2002**, *277*, 17209–17216.
- (357) Aubert, C.; Brettel, K.; Mathis, P.; Eker, A. P. M.; Boussac, A. EPR Detection of the Transient Tyrosyl Radical in DNA Photolyase from *Anacystis Nidulans*. *J. Am. Chem. Soc.* **1999**, *121*, 8659–8660.
- (358) Okafuji, A.; Schnegg, A.; Schleicher, E.; Möbius, K.; Weber, S. G-Tensors of the Flavin Adenine Dinucleotide Radicals in Glucose Oxidase: A Comparative Multifrequency Electron Paramagnetic Resonance and Electron–Nuclear Double Resonance Study. *J. Phys. Chem. B* **2008**, *112*, 3568–3574.
- (359) McDowall, J. S.; Ntai, I.; Hake, J.; Whitley, P. R.; Mason, J. M.; Pudney, C. R.; Brown, D. R. Steady-State Kinetics of α -Synuclein Ferrireductase Activity Identifies the Catalytically Competent Species. *Biochemistry* **2017**, *56*, 2497–2505.
- (360) Weyler, W.; Salach, J. I. Purification and Properties of Mitochondrial Monoamine Oxidase Type A from Human Placenta. *J. Biol. Chem.* **1985**, *260*, 13199–13207.
- (361) Jones, H. B. L.; Crean, R. M.; Matthews, C.; Troya, A. B.; Danson, M. J.; Bull, S. D.; Arcus, V. L.; van der Kamp, M. W.; Pudney, C. R. Uncovering the Relationship between the Change in Heat Capacity for Enzyme Catalysis and Vibrational Frequency through Isotope Effect Studies. *ACS Catal.* **2018**, *8*, 5340–5349.
- (362) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are and How They Behave. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 112–124.
- (363) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708.
- (364) Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **2016**, *13*, 505–517.
- (365) Kingsley, L. J.; Lill, M. A. Including Ligand-Induced Protein Flexibility into Protein Tunnel Prediction. *J. Comput. Chem.* **2014**, *35*, 1748–1756.
- (366) Biedermannová, L.; Prokop, Z.; Gora, A.; Chovancová, E.; Kovács, M.; Damborský, J.; Wade, R. C. A Single Mutation in a Tunnel to the Active Site Changes the Mechanism and Kinetics of Product Release in Haloalkane Dehalogenase LinB. *J. Biol. Chem.* **2012**, *287*, 29062–29074.
- (367) Zapata-Torres, G.; Fierro, A.; Miranda-Rojas, S.; Guajardo, C.; Saez-Briones, P.; Salgado, J. C.; Celis-Barros, C. Influence of Protonation on Substrate and Inhibitor Interactions at the Active Site of Human Monoamine Oxidase-A. *J. Chem. Inf. Model.* **2012**, *52*, 1213–1221.
- (368) Tabbutt, F. Water: A Matrix of Life, 2nd Edition (Franks, Felix). *J. Chem. Educ.* **2001**, *78*, 593.
- (369) Bonivento, D.; Milczek, E. M.; McDonald, G. R.; Binda, C.; Holt, A.; Edmondson, D. E.; Mattevi, A. Potentiation of Ligand Binding through Cooperative Effects in Monoamine Oxidase B. *J. Biol. Chem.* **2010**, *285*, 36849–36856.
- (370) Wells, S.; Menor, S.; Hesperheide, B.; Thorpe, M. F. Constrained Geometric Simulation of Diffusive Motion in Proteins. *Phys. Biol.* **2005**, *2*, S127–S136.
- (371) Oleinikovas, V.; Saladino, G.; Cossins, B. P.; Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 14257–14263.
- (372) Beglov, D.; Hall, D. R.; Wakefield, A. E.; Luo, L.; Allen, K. N.; Kozakov, D.; Whitty, A.; Vajda,

- S. Exploring the Structural Origins of Cryptic Sites on Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 3416–3425.
- (373) Evans, E. W.; Dodson, C. A.; Maeda, K.; Biskup, T.; Wedge, C. J.; Timme, C. R. Magnetic Field Effects in Flavoproteins and Related Systems. *Interface Focus* **2013**, *3*.
- (374) Stoll, S.; Schweiger, A. EasySpin, a Comprehensive Software Package for Spectral Simulation and Analysis in EPR. *J. Magn. Reson.* **2006**, *178*, 42–55.
- (375) Binda, C.; Hubálek, F.; Li, M.; Herzig, Y.; Sterling, J.; Edmondson, D. E.; Mattevi, A. Crystal Structures of Monoamine Oxidase B in Complex with Four Inhibitors of the N - Propargylaminoindan Class. *J. Med. Chem.* **2004**, *47*, 1767–1774.
- (376) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminform.* **2012**, *4*, 1–17.
- (377) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI *Membrane Builder* toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35*, 1997–2004.
- (378) Li, M.; Binda, C.; Mattevi, A.; Edmondson, D. E. Functional Role of the “Aromatic Cage” in Human Monoamine Oxidase B: Structures and Catalytic Properties of Tyr435 Mutant Proteins. *Biochemistry* **2006**, *45*, 4775–4784.
- (379) Edmondson, D. E.; Binda, C.; Wang, J.; Upadhyay, A. K.; Mattevi, A. Molecular and Mechanistic Properties of the Membrane-Bound Mitochondrial Monoamine Oxidases. *Biochemistry* **2009**, *48*, 4220–4230.
- (380) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain X1 and X2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (381) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O’Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
- (382) Allen, W. J.; Lemkul, J. A.; Bevan, D. R. GridMAT-MD: A Grid-Based Membrane Analysis Tool for Use with Molecular Dynamics. *J. Comput. Chem.* **2009**, *30*, 1952–1958.
- (383) Wells, S. A.; Crennell, S. J.; Danson, M. J. Structures of Mesophilic and Extremophilic Citrate Synthases Reveal Rigidity and Flexibility for Function. *Proteins Struct. Funct. Bioinforma.* **2014**, *82*, 2657–2670.
- (384) Schrödinger. PyMol Molecular Graphics System. Schrödinger, LLC.
- (385) van der Kamp, M. W.; Daggett, V. Influence of PH on the Human Prion Protein: Insights into the Early Steps of Misfolding. *Biophys. J.* **2010**, *99*, 2289–2298.
- (386) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Data Min Knowl Discov*; Elsevier, 1996; Vol. 90, pp 226–231.