Molecular Immunology 94 (2018) 91–97

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring

Esteban Lanzarotti[a], Paolo Marcatili[b], Morten Nielsen[a,b,*]

[a] Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina
[b] Department of Bio and Health Informatics, Technical University of Denmark, Building 208, Kemitorvet, 2800 Lyngby, Denmark

ARTICLE INFO

Keywords:
MHC
T cell receptor
Antigens/Peptides/Epitopes
Modelling
Pipeline

ABSTRACT

Interactions of T cell receptors (TCR) to peptides in complex with MHC (p:MHC) are key features that mediate cellular immune responses. While MHC binding is required for a peptide to be presented to T cells, not all MHC binders are immunogenic. The interaction of a TCR to the p:MHC complex holds a key, but currently poorly comprehended, component for our understanding of this variation in the immunogenicity of MHC binding peptides. Here, we demonstrate that identification of the cognate target of a TCR from a set of p:MHC complexes to a high degree is achievable using simple force-field energy terms. Building a benchmark of TCR:p:MHC complexes where epitopes and non-epitopes are modelled using state-of-the-art molecular modelling tools, scoring p:MHC to a given TCR using force-fields, optimized in a cross-validation setup to evaluate TCR inter atomic interactions involved with each p:MHC, we demonstrate that this approach can successfully be used to distinguish between epitopes and non-epitopes. A detailed analysis of the performance of this force-field-based approach demonstrate that its predictive performance depend on the ability to both accurately predict the binding of the peptide to the MHC and model the TCR:p:MHC complex structure. In summary, we conclude that it is possible to identify the TCR cognate target among different candidate peptides by using a force-field based model, and believe this works could lay the foundation for future work within prediction of TCR:p:MHC interactions.

1. Introduction

Binding to MHC (Major Histocompatibility Complex) is a prerequisite for peptide T cell immunogenicity. Given this, large efforts have been dedicated to the development of methods capable of accurately predict this event (some of the most accurate and publicly available at the IEDB are described in: Andreatta and Nielsen 2016; Nielsen and Andreatta 2016; Andreatta et al., 2015; Karosiene et al., 2013; Kim et al., 2009). The accuracy of the state-of-the-art methods has proven to be very high (in particularly for MHC class I), and most projects will in one way or another apply such prediction tools to guide the process of rational T cell epitope discovery (a few examples include Braendstrup et al., 2014; Pérez et al., 2008; Paul et al., 2015). However, not all peptides processed along the MHC pathways and bound by MHC turn out immunogenic. The main reason for this is the unavailability of T cells reactive to the given peptide-MHC (p:MHC) complex due to tolerance. The general rules underlying tolerance are well defined and deal with negative selection of T cells expressing a T cell receptor (TCR) with binding specificity towards p:MHC complexes of self-peptides. However, the details of these rules remain poorly described, and our understanding of the rules that define which p:MHCs are the targets of a given TCR remains highly limited.

In the last years, many efforts have been made modelling TCR:p:MHC systems. These efforts include simulation methods that have evolved from simulating the peptide in the MHC binding pocket for 1 nanosecond to simulating the entire TCR:p:MHC complex for more than 1 microsecond (Kass et al., 2014). Also, as more TCR:p:MHC complexes have been resolved by crystallography, template-based modelling techniques have achieved considerable accuracy, either using a single template or multiple templates (Liu et al., 2011). In other studies, force fields have been adapted in order to estimate changes in binding affinities, proving that structure-based methods are useful tools to design and engineer TCR and pMHC (mainly class I) interactions modulating both affinity and specificity (Pierce et al., 2014; Laugel et al., 2005). Also, docking approaches have shown that interactions between TCRs and pMHC complexes can be modelled when “good”

Abbreviations: TCR, T cel receptors; MHC, Major Histocompatibility Complex; IEDB, immune epitope database; HLA, human leukocyte antigen; CDR, complementarity determining region; RMSD, root mean square deviation; PDB, Protein Data Bank; MCM, Monte Carlo Method; AUC, area under the curve; LOCO, leave one cluster out

* Corresponding author at: Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Av. 25 de Mayo y Francia CP, San Martín, Buenos Aires, Argentina.
E-mail address: mniel@cbs.dtu.dk (M. Nielsen).

https://doi.org/10.1016/j.molimm.2017.12.019
Received 5 September 2017; Received in revised form 27 October 2017; Accepted 20 December 2017

Table 1
Epitopes in TCR:p:MHCII complexes. First column corresponds to the PDB code, underlined entries are excluded from the benchmark (see Section 2.1). Second column is for MHC family (DR, DP, DQ are human alleles and IE, IA are mouse alleles). TCR is specified in third column and the peptide or protein name is in fourth column. Fifth column corresponds to the Epitope sequence (cores in bold) and sixth column is for the UniprotID that was used to extract other peptides as Non-Epitopes (* is for cases that have no sequence, so random peptides were chosen from UniRef50).

PDB ID	MHCII	TCR	Peptide Name	Epitope sequence	Pept Full Seq
1D9 K	IA^k	D10	ConAlb	**HRGAIEWEG**IESG	P02789
1FYT	DR1	HA1.7	HA	PK**YVKQNTLKL**AT	Q38SR9
1J8H	DR4	HA1.7	HA	PK**YVKQNTLKL**AT	Q38SR9
1U3H	IA^u	172.1	MBP1-11	SR**GGASQYRPS**Q	P04370
1YMM	DR2	OB.1A12	MBP85-99	ENPV**VHFFKNIVT**PR	P02686
1ZGL	DR2a	3A6	MBP89-101	VHF**FKNIVTPRT**P	P02686
2IAM	DR1	E8	mutTPI	GEL**IGTLNAAKV**PAD	P60174
2IAN	DR1	E8	TPI	GEL**IGTLNAAKV**PAD	P60174
2PXY	IA^u	1934	MBP1-11	SR**GGASQYRPS**Q	P04370
2WBJ	DR2	OB.1A12	ENGA	FAR**VHFISALHG**	A7ZPV4
2Z31	IA^u	Cl19	MBP1-11	SR**GGASQYRPS**Q	P04370
3C5Z	IA^b	B3K506	p3 K	FE**AQKAKANKA**	*
3C60	IA^b	YAE62	p3 K	FE**AQKAKANKA**	*
3C6L	IA^b	2W20	P3 K	FE**AQKAKANKA**	*
3MBE	IA^g	21.3	HEL	AMK**RHGLDNYRG**YSLGN	P00698
3O6F	DR4	MS2-3C8	MBP114-126	FS**WGAEGQRPG**FG	P02686
3PL6	DQ1	Hy.1B11	MBP85-99	ENPV**VHFFKNIVT**PR	P02686
3QIB	IE^k	2B4	MCC88-104	ADL**IAYLKQATK**	P00039
3QIU	IE^k	226	MCC88-104	ADL**IAYLKQATK**	P00039
3QIW	IE^k	226	MCC88-104p5E	ADL**IAYLKQATK**	P00039
3RDT	IA^b	J809.B5	P3 K	FE**AQKAKANKA**	*
3T0E	DR4	MS2-3C8	MBP114-126	FS**WGAEGQRPG**FG	P02686
4C56	DR1	AV22/BV19	HA	PK**YVKQNTLKL**AT	Q38SR9
4E41	DRA1	G4	mutTPI	GEL**IGTLNAAKV**PAD	P60174
4GG6	DQ8	SP3.4	Glia-alpha1	SG**EGSFQPSQE**NP	X2KVI4
4GRL	DQ1	Hy.1B11	pMM	DRLL**MLFAKDVVS**RN	P26276
4H1L	DR52c	Ani2.3	pHIR(Ni2 +)	H**IRCNIPKRI**	*
4MAY	DQ1	Hy.1B11	UL15	FRQL**VHFVRDFAQ**LL	P04295
4OZF	DQ2	JR5.1	Glia-alpha2	PF**PQPELPYPQ**PQ	X2KWL1
4OZG	DQ2	D2	Glia-alpha2	PF**PQPELPYPQ**PQ	X2KWL1
4OZH	DQ2	S16	Glia-alpha2	PF**PQPELPYPQ**PQ	X2KWL1
4OZI	DQ2.5	S2	Glia-alpha1a	LQ**PFPQPELPY**PQ	X2KWL1
4P23	IA^b	J809.B5	p3 K	FE**AQKAKANKA**	*
4P2Q	IE^k	5cc7	5c2	ADG**LAYFRSSFK**	*
4P2R	IE^k	5cc7	5c1	ANG**VAFFLTPFK**A	*
4P46	IA^b	J809.B5 Y31A	p3 K	FE**AQKAKANKA**	*
4P4 K	DP2	AV22	M2(Be2 +)	**FWIDLFETI**G	*
4P5T	IA^b	14.C6	p3 K	FE**AQKAKANKA**	*
4Y19	DR4	FS18	Insulin	GSLQP**LALEGSLQK**RGIV	P01308
4Y1A	DR4	FS19	Insulin	GSLQP**LALEGSLQK**RGIV	P01308
4Z7U	DQ8	S13	Glia-alpha1	SG**EGSFQPSQE**NP	X2KVI4
4Z7V	DQ8	L3	Glia-alpha1	SG**EGSFQPSQE**NP	X2KVI4
4Z7W	DQ8	T316	Glia-alpha1	SG**EGSFQPSQE**NP	X2KVI4

scoring functions are used (Riley et al., 2016; Pierce and Weng, 2013). Focusing on peptide immunogenicity, TCR interactions with pMHC class I complexes, in particular for HLA-A*02:01, have shown that CDR loops interactions to unknown epitopes can be predicted using a very simple rule-based model learning from known complexes of the same allele (Roomp and Domingues 2011). Also, a particular case (LC13 TCR and HLA-B*08:01) was characterized using 100 ns molecular dynamics simulations. Here, however no strong difference was found regarding the binding behaviour between more and less immunogenic peptides (Knapp et al., 2014).

Given this background, we seek to answer, given TCR:p:MHC modelled complexes of different peptides interacting with the same MHC and TCR molecules, which structural properties can be used in order to predict the cognate target (i.e. the p:MHC complex) of the TCR. To address this question, first we built a benchmark set based on solved TCR:p:MHC of class II and generated homology models for both the bound epitope and a set of natural MHC-binding non-epitopes. Next, we used two well-known force fields, FoldX (Guerois et al., 2002) and Rosetta's Talaris2013 (Leaver-Fay et al., 2013; O'Meara et al., 2015), to mimic the molecular interactions and chemical properties between the TCR and the p:MHC complex. FoldX (Schymkowitz et al., 2005) has in earlier studies demonstrated high performance predicting the impact of a mutation in the context of a given biological assembly. Rosetta (Bradley et al., 2005) has been extensively used in a large range of applications, from *de novo* protein design to understanding the folding process. These two force fields are described as weighted sums of terms modelling interactions in a given molecular assembly. Here, we investigate how these force fields could be used to identify the target of a given TCR. The weights of the two force fields were adjusted in a cross-validation setup in order to detect correlations between each force field term and the peptide immunogenicity. This approach allowed us to define a robust model, that given the sequences of the MHC alpha and beta subunits, TCR alpha and beta subunits and a set of peptides, could discriminate between epitopes and non-epitopes, and thus correctly predict the cognate target for the given TCR.

2. Material and methods

2.1. The TCR:p:MHCII data set

A data set of 43 TCR:p:MHCII was downloaded from the PDB (Berman et al., 2003). Entries presenting extreme TCR orientations compared with all other entries in the dataset were excluded (4Y1A, 4Y19, 4C56, 3PL6, 2WBJ and 1YMM) (see Supplementary Fig. 1).