

## Role of in silico structural modeling in predicting immunogenic neoepitopes for cancer vaccine development

Neeha Zaidi, ... , Shozeb Haider, Elizabeth M. Jaffee

*JCI Insight*. 2020;5(17):e136991. <https://doi.org/10.1172/jci.insight.136991>.

Technical Advance

Immunology

Oncology

In prior studies, we delineated the landscape of neoantigens arising from nonsynonymous point mutations in a murine pancreatic cancer model, Panc02. We developed a peptide vaccine by targeting neoantigens predicted using a pipeline that incorporates the MHC binding algorithm NetMHC. The vaccine, when combined with immune checkpoint modulators, elicited a robust neoepitope-specific antitumor immune response and led to tumor clearance. However, only a small fraction of the predicted neoepitopes induced T cell immunity, similarly to that reported for neoantigen vaccines tested in clinical studies. While these studies have used binding affinities to MHC I as surrogates for T cell immunity, this approach does not include spatial information on the mutated residue that is crucial for TCR activation. Here, we investigate conformational alterations in and around the MHC binding groove induced by selected minimal neoepitopes, and we examine the influence of a given mutated residue as a function of its spatial position. We found that structural parameters, including the solvent-accessible surface area (SASA) of the neoepitope and the position and spatial configuration of the mutated residue within the sequence, can be used to improve the prediction of immunogenic neoepitopes for inclusion in cancer vaccines.

Find the latest version:

<https://jci.me/136991/pdf>



# Role of in silico structural modeling in predicting immunogenic neoepitopes for cancer vaccine development

Neeha Zaidi,<sup>1</sup> Mariya Soban,<sup>2,3</sup> Fangluo Chen,<sup>1</sup> Heather Kinkead,<sup>1</sup> Jocelyn Mathew,<sup>1</sup> Mark Yarchoan,<sup>1</sup> Todd D. Armstrong,<sup>1</sup> Shozeb Haider,<sup>2</sup> and Elizabeth M. Jaffee<sup>1</sup>

<sup>1</sup>The Sidney Kimmel Comprehensive Cancer Center, The Skip Viragh Center for Pancreatic Cancer, The Bloomberg–Kimmel Institute for Cancer Immunotherapy, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

<sup>2</sup>Department of Pharmaceutical and Biological Chemistry, University College London School of Pharmacy, London, United Kingdom. <sup>3</sup>Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, Aligarh, India.

In prior studies, we delineated the landscape of neoantigens arising from nonsynonymous point mutations in a murine pancreatic cancer model, Panc02. We developed a peptide vaccine by targeting neoantigens predicted using a pipeline that incorporates the MHC binding algorithm NetMHC. The vaccine, when combined with immune checkpoint modulators, elicited a robust neoepitope-specific antitumor immune response and led to tumor clearance. However, only a small fraction of the predicted neoepitopes induced T cell immunity, similarly to that reported for neoantigen vaccines tested in clinical studies. While these studies have used binding affinities to MHC I as surrogates for T cell immunity, this approach does not include spatial information on the mutated residue that is crucial for TCR activation. Here, we investigate conformational alterations in and around the MHC binding groove induced by selected minimal neoepitopes, and we examine the influence of a given mutated residue as a function of its spatial position. We found that structural parameters, including the solvent-accessible surface area (SASA) of the neoepitope and the position and spatial configuration of the mutated residue within the sequence, can be used to improve the prediction of immunogenic neoepitopes for inclusion in cancer vaccines.

**Authorship note:** SH and EMJ are co-senior authors.

**Conflict of interest:** EMJ is a paid consultant for Adaptive Biotech, CTSONE, Achilles, DrangonFly, and Genocoe. She receives funding from Lustgarten Foundation and AduroBiotech and through a licensing agreement between AduroBiotech and Johns Hopkins University (JHU), and JHU has the potential to receive royalties on GVAX. She is the Chief Medical Advisor for Lustgarten and serves on the National Cancer Advisory Board and as an Advisor to the Parker Institute for Cancer Immunotherapy (PICI). MY serves on the Advisory Board for AstraZeneca, Eisai, Exelixis, and Geneos. He also receives grant/research support from Bristol-Myer Squibb, Exelixis, Genentech/Roche, and Incyte.

**Copyright:** © 2020, Zaidi et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

**Submitted:** February 6, 2020

**Accepted:** July 24, 2020

**Published:** September 3, 2020.

**Reference information:** *JCI Insight*. 2020;5(17):e136991.  
<https://doi.org/10.1172/jci.insight.136991>

## Introduction

The past decade has proven the therapeutic benefit of instructing the immune system to identify and kill cancer cells (1, 2). Immune checkpoint inhibitors (ICIs), such as anti-programmed cell death-1 (anti-PD-1), anti-programmed cell death ligand (anti-PD-L1), and anti-cytotoxic T lymphocyte-associated protein 4 (anti-CTLA-4) antibodies, are now a standard of care for some previously incurable cancers (3). These agents unleash the killing power of cancer-antigen-targeted effector T cells that infiltrate these immune responsive cancers. These T cells often recognize tumor-specific proteins called neoantigens that differentiate the tumor cell from its normal cell of origin. Neoantigens arise from nonsynonymous mutations, chromosomal rearrangements that result in fusion proteins, and/or insertions and deletions of bases into the genome (indels) during tumorigenesis (4, 5). They are presented to T cells by human leukocyte antigen (HLA) molecules on the tumor cell surface and are expressed only on tumors and not on normal cells. Certain tumors, such as melanomas and cancers with mismatch repair deficiencies (MMR-d) that lead to microsatellite instability (MSI), express a high burden of neoantigens (6, 7), whereas others, such as pancreatic cancers, glioblastomas, and MMR proficient (MMR-p) colon cancers, have modest to low neoantigen burdens (8–10). While the former group of tumors respond well to ICIs, the latter fail to induce and attract quality effector T cells within the tumor and, thus, respond poorly to single or dual agent ICIs (11).

One approach being developed to improve ICI responses against tumors with low mutational burden is to sequence the neoantigens in a given patient's tumor and formulate a neoantigen-targeted vaccine to evoke high-quality, robust, neoepitope-specific T cell responses that can then be enhanced by ICIs. Such highly “personalized” vaccines have shown considerable promise in patients with hypermutated tumors such as melanoma. Patients with melanoma vaccinated with either synthetic long peptides (SLPs) (15–30 amino acids) or poly-neoepitope RNA (10 selected mutations within 27 amino acid peptides) displayed neoepitope-specific T cells in

peripheral blood, which were shown to kill tumor cells expressing these neoantigens in vitro (12, 13). Two recent studies in patients with glioblastoma, a tumor type with a low neoantigen burden, reported the successful use of a personalized vaccine approach to stimulate neoantigen-specific T cell responses in peripheral blood. One patient was also noted to have T cells that trafficked into the tumor (9, 14). In a preclinical murine model of pancreatic cancer, we used a similar SLP-neoantigen-based vaccine approach to induce quality T cells that trafficked into a lower neoantigen burden tumor. We found that combination ICIs were also required to fully activate these tumor infiltrating T cells (TILs) to eradicate tumors (15).

The current pipeline for personalized vaccines in both preclinical and clinical settings uses whole exome sequencing (WES) and RNA sequencing (RNA-seq) to identify mutated proteins expressed in tumor tissue. An in silico algorithm, NetMHC, is then used to predict the binding affinity of each neoepitope to its MHC molecules. Those neoantigens with MHC binding affinities greater than 500 nM are predicted as being immunogenic and, hence, have been selected for inclusion in vaccines. However, in all 4 clinical studies (9, 12–14), only a small fraction of the neoantigens selected using MHC binding algorithms elicited robust neoantigen-specific T cell responses in the peripheral blood. Likewise, in our pancreatic cancer model, there was a discordance between MHC binding affinities as reported by NetMHC (versions 3.2, 3.4, and pan 2.8) and immunogenicity, as assessed by IFN- $\gamma$  production by CD8<sup>+</sup> T cells (2). Furthermore, there are instances where NetMHC-predicted neoepitopes have been noted to be functionally inhibitory (16).

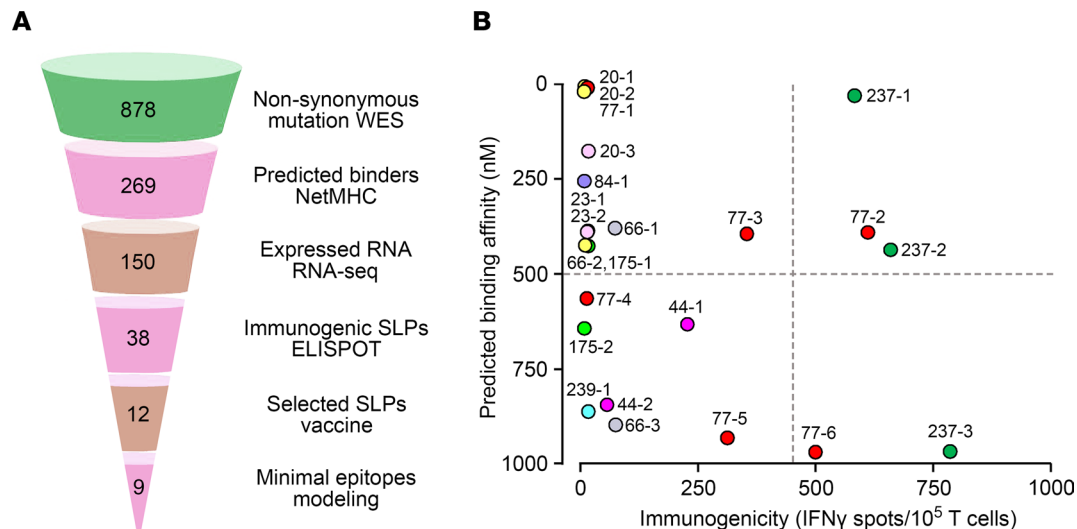
A major challenge in further developing these personalized vaccines is the need to create algorithms that accurately select neoantigens expressed by a given patient's tumor that are more likely to elicit the highest quality and most durable T cell responses. A recent study posits that the quality of neoantigens, noted for their similarity to pathogen-derived antigens known to elicit T cell responses, better predicts immunogenicity than T cell quantity alone (17, 18). Furthermore, structural modeling using existing crystal structures of HLA molecules suggest that the neoepitope, when presented at the MHC groove, must be solvent exposed to be recognized by the T cell receptor (TCR) (19). Of 170 neoepitopes predicted by NetMHC 3.4, 3.2, and pan 2.8 in a mouse colon cancer cell line (MC38), only 7 neoantigens were identified on mass spectrometry, of which 2 were immunogenic by qualitatively examining solvent exposure (19). We have defined immunogenicity narrowly as generating a positive response on our ELISPOT, with our readout showing IFN- $\gamma$ -producing CD8<sup>+</sup> T cells isolated from splenocytes vaccinated with respective neoepitope-corresponding peptides. More recently, there have been attempts to develop neural networks using the a general immune epitope database (IEDB), consisting mainly of viral antigens, by incorporating total binding energy scores from crystallographic and modeled structures (20). Here, we present a comprehensive structural evaluation in atom-level detail of how the solvent exposure and the configuration of the mutated residue within the MHC I groove facilitates its recognition by T cells.

## Results

*NetMHC predicts neoepitope immunogenicity inconsistently.* We subjected the murine pancreatic Panc02 tumor to WES and identified 878 nonsynonymous mutations (Figure 1A). Two hundred and sixty-nine of these were predicted by NetMHC to bind either H-2Kb or H-2Db with affinities  $\leq 1000$  nM. NetMHC predicted 8- to 9-amino acid-long minimal epitope peptides (MEPs) for the neoantigens as potential vaccine candidates (Table 1). We directly tested immunogenicity in vitro using an ELISPOT assay for IFN- $\gamma$ -producing CD8<sup>+</sup> T cells. Figure 1B shows an imperfect correlation between the immunogenicity of individual MEPs and their predicted binding affinity derived from NetMHC. Certain MEPs, such as 20-3, 66-2, and 84-1, 175-1, and 175-2, were predicted to have high binding affinities stronger than 500 nM, a cut-off used in clinical studies (9, 12), but failed to elicit T cell responses. In contrast, MEP 237-3 was predicted to be a poor binder based on a lower binding affinity ( $<500$  nM) but displayed the most potent immunogenic response. NetMHC correctly predicted the immunogenicity of certain variants, such as MEP 77-2 and 77-3. Thus, despite being widely used, the NetMHC-based algorithm does not consistently predict the immunogenicity of HLA/MHC binding MEPs.

*Structural templates influence neoepitope exposure.* The  $\alpha 1$  and  $\alpha 2$  subdomains of MHC fold to form a groove, with the curved base as a  $\beta$ -sheet and 2  $\alpha$ -helices separated at the top to accommodate 8- to 10-amino acid-long peptides. Therefore, we evaluated conformational alterations in and around the binding groove induced by selected minimal neoepitopes to understand the influence of a given mutated residue as a function of its spatial position.

We constructed multiple structural models of 9 MEP-HLA complexes (Figure 2A). Template-based homology modeling is based on the selection of templates with the highest sequence similarity to the modeled sequence.



**Figure 1. Identification of immunogenic tumor neoantigens in a murine pancreatic cancer model. (A)** Pipeline for identifying nonsynonymous mutations in murine Panc02 cells by whole exome sequencing (WES); examining the transcriptome by RNA-seq; predicting binding affinity and minimal epitopes by NetMHC; generating synthetic long peptides (SLPs) for ELISPOT assays; and performing structural modeling on selected neoepitopes. For further details, please refer to ref. 15. **(B)** Inadequate correlation between predicted binding affinity (NetMHC 3.2, 3.4, and pan 2.8) of MEPs and their immunogenicity, as assessed by ELISPOT for IFN- $\gamma$ -producing T cells in vitro ( $n = 3$  mice per group) (for details on immunogenicity data, please refer to ref. 15). Note the clustering of high-affinity MEPs with poor immunogenicity (upper left quadrant).

Alignment of our MEPs to the 40 H-2Db and 45 H-2Kb MHC I crystal structures available in PDB revealed sequence similarities between 33% and 87% (Table 2). We also found multiple templates in PDB that exhibited near-sequence similarities with a given MEP.

Structural modeling using existing crystal structures of HLA molecules suggest that neoepitopes must be exposed to solvent to be recognized by the TCR (19, 20). As we noted several high-sequence similarity–exhibiting structural templates for each MEP, each identified structural template needed to be inspected carefully for side chains that were either buried or exposed to solvent. For example, for MEP 44-1, which was immunogenic, we identified 5JWD as a high-sequence similarity template (at 55%) in which the mutated cysteine (Cys) residue side chain was solvent exposed (Table 2 and Figure 2A). Likewise, when modeling the nonimmunogenic 44-2 variant, 2ZOL was chosen at a 33% homology because the Cys side chain was in a buried state. For MEP 66-1, of the 5 identified templates, namely 1JPF, 3CCH, 3CH1, 3TBV, and 3TBW, we selected 1JPF with a sequence similarity of 55% (Table 2). For MEP 84-1, 4PG9 was selected with a sequence similarity of 75% from 4 templates, namely 1KPV, 2VAB, 4PG9, and 4PGC. For MEP 175-1, 7 templates (1KPV, 2VAB, 4PG9, 4PGC, 4PGD, 4PGB, and 4PGE) were identified, and 4PG9 was chosen with a sequence similarity of 62% (Table 2). For MEP 237-1, 2FO4 was selected with a sequence similarity of 87%. Three variants of MEP 77 were modeled (77-3, 77-5, and 77-6), in which the mutant Ser residue was solvent exposed, with 1BQH, ILK2, and 1S7R being chosen for the 3 variants (62%, 62%, and 50% homology, respectively). Overall, multiple templates based on epitope sequence homology were identified. The final template was chosen after taking into account whether the mutant residue was buried or exposed to solvent.

For initial testing, 8- to 9-amino acid–long epitopes containing the mutated residue, which had been previously tested for immunogenicity, were modeled into the MEP-HLA complexes using existing MHC I–peptide complex structures extracted from PDB as starting conformations (Figure 2A). To compare the modeled MEP-HLA complexes with the existing MEP-HLA complexes, we carried out an assessment of root square mean deviations ( $C\alpha$  RMSD), a measure of the average distance between atoms of docked proteins, for each optimized peptide with its corresponding template. The generally low  $C\alpha$  RMSD (0.75–1.33 Å) indicated that optimized complexes resembled the native complex conformation (Table 2).

*Mutated residue orientation and SASA predict immunogenicity.* Each modeled MEP-HLA complex was used to investigate the effect of the spatial position of the mutated residue, both qualitatively by examining buried (in) versus exposed (out) configurations (Figure 2A, top and side views), and quantitatively,

**Table 1. Sequences of MEPs that bind H-2Db or H-2Kb**

MEP	Minimal epitope
20-1 (Db)	VAFGHS <b>N</b> L
20-2	VAFGHS <b>N</b> L <b>F</b>
20-3	VAFGHS <b>N</b> L <b>F</b> I
23-1 (Db)	KIAFLPF <b>A</b> YL
23-2	IAFLPF <b>A</b> YLV
44-1 (Db)	LSCLNWSTL
44-2	SCLNWSTLV
66-1 (Db)	CGMENQVLV
66-2	LALGPFC <b>G</b> M
66-3	CGMENQVLVI
77-1 (Kb)	RIYSLFN <b>L</b> S
77-2	RIYSLFNLS <b>M</b>
77-3	LSMGKLE <b>K</b> M
77-4	SLFNLSMG <b>K</b> L
77-5	FNLSMG <b>K</b> L
77-6	IYSLFNLS <b>M</b>
84-1 (Kb)	VVVGNHNQV
175-1 (Kb)	LPLAFTLL
175-2	LLPLAFTLL
237-1 (Kb)	SAYQFRVC
237-2	SAYQFRVCA
237-3	SAYQFRVCAE
239-1 (Kb)	ISISKCAAC

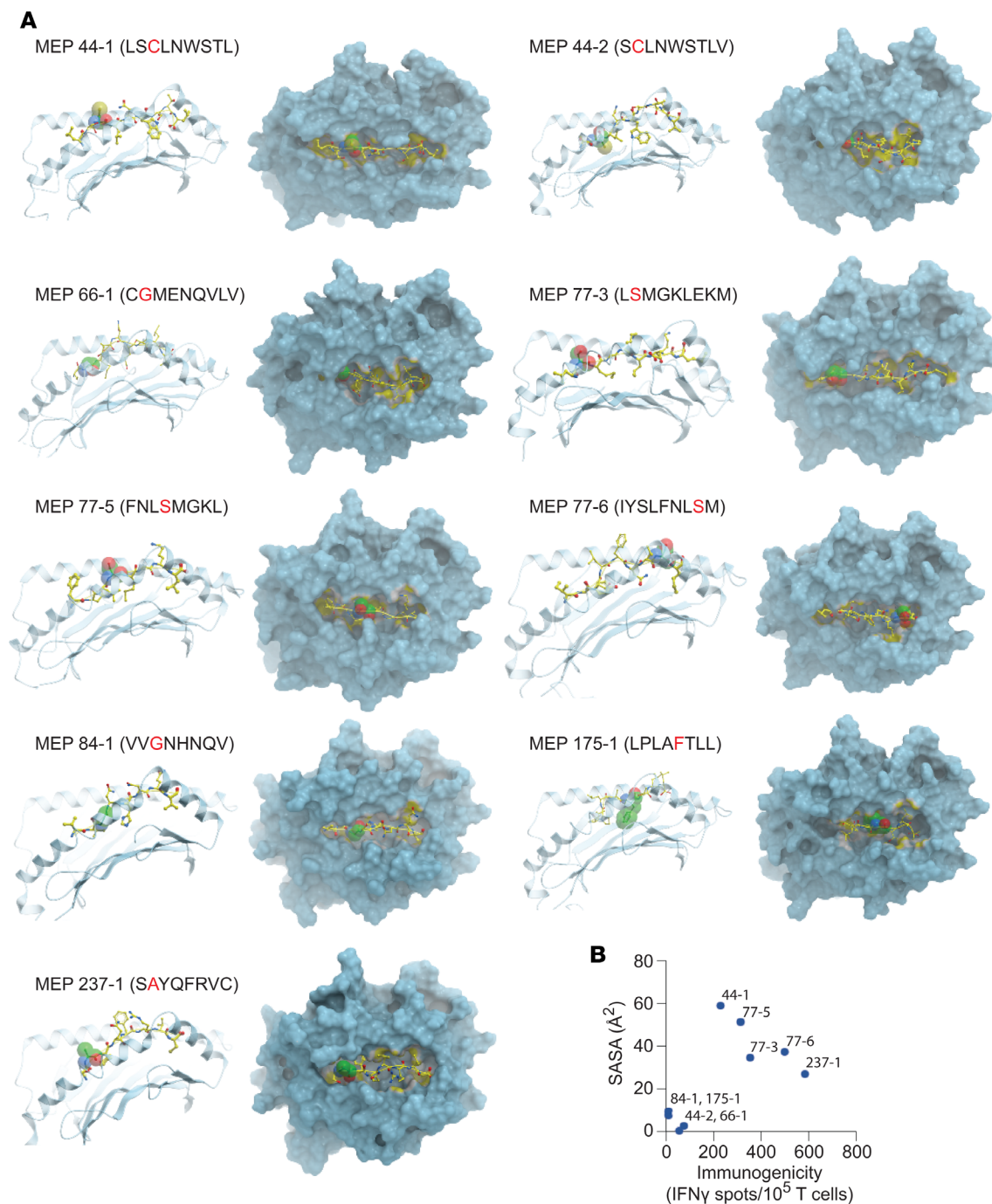
Mutated residues are noted in bold.

by calculating solvent-accessible surface area (SASA) (Table 3, Table 4, and Table 5). Prior modeling of mutant MEPs in murine colon cancer suggested that mutated residues oriented toward the solvent interface were more likely to be immunogenic and that residue numbers 3–7 of any presented peptide were involved in TCR recognition (19). However, this may not always be the case, as an Ala residue at position 2 in MEP 237-1, which is solvent exposed, triggers immunogenicity (Table 5 and Figure 2A). Table 3, Table 4, and Table 5 list the contribution of different parameters calculated by the GETAREA module within FANTOM 4.0; this includes an automated output on whether the mutated residue is buried (in) or solvent-exposed (out). Of note, high SASA values (>20) noted with MEP 44-1, 77-3, 77-5, 77-6, and 237-1 showed an outward configuration of the mutated residue, whereas MEPs with low SASAs (<15), namely 44-2, 66-1, 175-1, and 84-1, displayed buried residues. Figure 2B shows that the high-SASA MEP-HLA complexes predicted good MEP immunogenicity, whereas those with low SASA values were less immunogenic.

*Shuffling mutated residues alters SASA, orientation, and immunogenicity.* Table 3, Table 4, and Table 5 show that, across the board, SASA values for each amino acid varied within MEP-HLA complexes. For example, the Leu residue next to the Ser in the MEP 77 models displayed SASAs of 3.32, 10.49, and 40.13 (Table 4). Hence, we posit that the spatial position of residues would determine SASA. A contribution of the spatial position to immunogenicity is best appreciated with MEP 44 (Table 3). The MEP 44-1–HLA complex with the mutated Cys residue at position 3, namely LSCLNWSTL, was constructed such that the side chain of Cys was solvent exposed (Figure 2A). The complex was confirmed by GETAREA to have a high SASA value (58.9) and an outwardly facing Cys side chain (Table 3). MEP 44-2 peptide also elicited a robust CD8<sup>+</sup> T cell response on ELISPOT (Figure 1B). However, a shift of Cys by just 1 amino acid to position 2 (SCLNWSTLV) in MEP 44-2 lowered SASA by ~100-fold (to 0.51), buried the mutant Cys, and rendered the MEP less immunogenic on ELISPOT (Figure 2B).

To further test the hypothesis that the position of the mutated residue within a MEP can alter the orientation and SASA of that residue within the MHC I groove, we created 2 variants of MEP 77-5 by shifting the Ser residue by 1 amino acid. Notably, MEP 77-5 showed a low binding affinity on NetMHC





**Figure 2. Modeling neoepitopes in selected MHC I templates. (A)** Side (left panels) and top views (right panels) of modeled complexes between MEPs and appropriately selected MHC I template (also see Table 3, Table 4, and Table 5). MEP annotation and sequence are shown with the mutated residue in red. The ball represents the mutated residue and displays either an outward or inward orientation. This qualitative data are consistent with corresponding data from GETAREA that annotate the mutant residue and calculate the solvent-accessible surface area (SASA) (Table 3, Table 4, and Table 5). **(B)** Relationship between SASA (Å²) of the mutated residue of a given MEP and its immunogenicity, as assessed by ELISPOT for IFN-γ (for details on immunogenicity data, please refer to ref. 15).

(Figure 1B and Figure 2A). However, since it was immunogenic on ELISPOT and displayed an outwardly pointing Ser residue with a high SASA (51.27), we predicted that shuffling the position of the Ser residue within the MHC groove would alter SASA and residue orientation — and, therefore, immunogenicity. In the first sequence, the mutated Ser residue was shuffled from fourth to fifth position (FNLSMGKL → LFNLSMGK) (termed MEP 77-5v1) (Table 6). In the second sequence, the Ser residue was shuffled from the fourth to the sixth position (FNLSMGKL → SLFNLSMGK) (termed MEP 77-5v2).

**Table 2. MHC I templates selected from Protein Data Base (PDB) for each neoepitope**

Name	Mutation	MEP	Minimal epitope	Template homolog PDB ID	Sequence similarity %	Ca RMSD with template
<i>Glb1l2</i>	R36C	44-1	LSCLNWSTL	5JWD	55	1.20
		44-2 (Db)	SCLNWSTLV	2ZOL	33	1.33
<i>Map2k5</i>	A11G	66-1 (Db)	CGMENQVLV	1JPF	55	0.95
		77-3 (Kb)	LSMGKLEKM	1BQH	62	0.84
<i>Rasa3</i>	Y740S	77-5	FNLSMGKL	1LK2	62	0.77
		77-6	IYSLFNLSM	1S7R	50	1.02
		84-1 (Kb)	VVVGNNHNV	4PG9	75	0.86
<i>Clcn7</i>	D771G	84-1 (Kb)	VVVGNNHNV	4PG9	75	0.86
<i>Bsg</i>	A8P	175-1 (Kb)	LPLAFTLL	4PG9	62	1.01
<i>Ttn</i>	E19018A	237-1 (Kb)	SAYQFRVC	2FO4	87	0.75

Gene name, point mutation, minimal epitope peptide (MEP), and its annotation, selected MHC I template (PDB ID), percent sequence similarity, and Ca RMSD values are shown.

For MEP 77-5v1, 3 templates were identified (1G7P, 1VAD, 1G7Q), of which 1G7P was selected exhibiting 50% similarity. MEP 77-5v2 identified 4 templates (1S7Q, 1S7R, 1S7S, 1S7T), from which 1S7T was selected with 50% similarity (Figure 3A and Table 6). GETAREA analysis of the variant MEP-HLA models revealed low SASAs and buried Ser side chains. Both variant peptides also displayed poor immunogenicity on ELISPOT assays (Figure 3B). Thus, shuffling the mutated residue by 1 and 2 positions in the original MEP 77-5 significantly attenuated immunogenicity. In all, both post hoc and a priori modeling testify to the critical roles of SASA and mutant residue orientation and position in predicting the immunogenicity of minimal neoepitopes.

## Discussion

We studied the presentation of 9 predicted neoepitopes used in our pancreatic cancer vaccine PancVAX (15) by modeling these onto MHC I templates derived from crystal structures from the Protein Data Base (PDB). Much more complex than previously envisioned (19), we show that the prediction of cancer neoantigen immunogenicity by in silico modeling (a) depends upon the selected template, (b) correlates with the SASA of the neoepitope, and (c) can be modified by shuffling the mutated residue within the sequence, which, in turn, exposes or buries the single mutated residue. This atomistic detail provides the framework for incorporating 3D structural parameters into automated machine learning algorithms as a high-throughput neoepitope prediction tool of the future.

For effective personalized cancer vaccines, it is imperative that nonsynonymous mutations arising during tumorigenesis not only yield expressed antigens that are processed and presented, but that the neoantigens that are included in a vaccine are able to trigger a robust and durable T cell response. In our Panc02 model, we have effectively demonstrated a biologically meaningful response after vaccination with PancVAX by demonstrating both reduced tumor growth and induction of neoantigen-specific T cells within the tumor (15). Following WES, which identifies DNA variants, RNA-seq and, in some instances, mass spectrometry, have been used to identify the DNA variants that are tumor-expressed proteins and neoepitopes presented on MHC (2, 9, 12–14). However, it has been much more challenging to determine which expressed neoantigens evoke a neoepitope-specific T cell response within tumor tissue and in peripheral blood.

Recent studies have used NetMHC to determine binding affinities of expressed neoantigens to HLA and MHC I as surrogate readouts for T cell immunity. An algorithm based on artificial neural networks, NetMHC, is trained for 81 different human HLA alleles and 41 animal MHC I alleles, which include murine MHCs (21, 22). The algorithm has been used widely to predict MHC-peptide binding for 8-, 10-, or 11-amino acid peptides (using 9-amino acid trained predictors) in several pathogenic viral proteomes, including SARS, influenza, HIV, and hepatitis C (23, 24). This training has resulted in 75%–80% confirmed HLA/MHC I binders, with further validation and benchmarking using large sets of nonredundant affinity data (22–24). However, in the case of tumor neoantigens, NetMHC binding has only inconsistently predicted an immune response (2, 9, 12–14) (Figure 1B). In both glioblastoma and melanoma trials, patients were vaccinated with up to 20 SLPs based on NetMHC predictions. In the melanoma studies, only 2–4

**Table 3. Solvent-accessible surface area (SASA) and annotations of neoepitope orientation of MEP 44 and 66**

MEP	Residue	Total	Apolar	Backbone	Sidechain	Ratio (%)	In/out
44-1 (LSCLNWSTL)	Leu 1	23.62	23.62	0.03	23.59	16.10	i
	Ser 2	12.78	4.96	8.99	3.78	4.90	i
	<b>Cys 3</b>	<b>58.90</b>	<b>5.36</b>	<b>7.50</b>	<b>51.41</b>	<b>50.30</b>	<b>o</b>
	Leu 4	15.38	9.39	7.53	7.85	5.40	i
	Asn 5	49.60	12.93	0.21	49.39	43.20	o
	Trp 6	2.97	0.85	2.13	0.85	0.40	i
	Ser 7	17.58	17.44	11.15	6.44	8.30	i
	Thr 8	56.51	52.68	1.20	55.31	52.10	o
	Leu 9	10.18	9.33	0.85	9.33	6.40	i
44-2 (SCLNWSTLV)	Ser 1	0.18	0.18	0.00	0.18	0.20	i
	<b>Cys 2</b>	<b>0.51</b>	<b>0.13</b>	<b>0.35</b>	<b>0.16</b>	<b>0.20</b>	<b>i</b>
	Leu 3	2.51	2.49	0.13	2.38	1.60	i
	Asn 4	82.78	24.18	9.67	73.11	64.00	o
	Trp 5	12.59	0.07	12.59	0.00	0.00	i
	Ser 6	39.43	28.96	16.57	23.86	29.50	o
	Thr 7	31.79	23.98	10.71	21.08	19.80	i
	Leu 8	59.29	59.29	0.00	59.29	40.60	o
	Val 9	0.01	0.01	0.00	0.01	0.00	i
66-1 (CGMENQVLV)	Cys 1	0.00	0.00	0.00	0.00	0.00	i
	<b>Gly 2</b>	<b>2.46</b>	<b>2.46</b>	<b>2.46</b>	<b>0.00</b>	<b>2.80</b>	<b>i</b>
	Met 3	7.34	4.48	4.12	3.22	2.00	i
	Glu 4	61.76	60.54	11.80	49.97	35.40	o
	Asn 5	118.28	36.30	11.73	106.55	93.20	o
	Gln 6	0.01	1.22	7.93	0.08	0.10	i
	Val 7	36.27	36.27	6.40	29.87	24.40	o
	Leu 8	85.65	83.63	2.02	83.63	57.20	o
	Val 9	0.82	0.00	0.00	0.82	0.70	i

Readouts from GETAREA on modeled MEP-MHC I template complexes. Shown are the MEP numbers, sequences, amino acids in the 9-mers (with mutated residue shown in bold), total SASA values (total), apolarity, and contribution of the backbone and side chains in each post-processed PDB file. The 'in' or 'out' configuration is also annotated for each residue.

SLPs evoked a neoepitope-specific CD8<sup>+</sup> T cell response in the peripheral blood, whereas in the glioblastoma study, only 1 patient generated a CD8<sup>+</sup> T cell response to 2 neoepitopes. Likewise, in our study with murine pancreatic cancer, Figure 1B illustrates a cluster of high-binding MEPS, such as 20-1, 20-2, and 77-1, among others, that display minimal or no immunogenicity.

The inconsistent prediction of neoepitope immunogenicity may indeed arise from the fact that NetMHC is trained to enact binding affinities of foreign pathogenic antigens, not tumor neoantigens. Specifically, it may not be trained well enough to distinguish as precisely between tumor neoantigens arising from point mutations with just a single amino acid difference. This apparent discrepancy will then be consistent with the “quality model,” in which high-quality neoantigens are considered to rank closer to pathogen-derived peptides (17, 18). In this instance, it would also be biologically meaningful to hypothesize that the mutated residue must be prominently exposed to solvent at the MHC I groove to be detected as foreign by the TCR (19).

We tested this hypothesis by structural modeling to calculate specifically a well-characterized parameter for solvent exposure, SASA (25). SASA has been used widely to evaluate the structure, function, and stability of proteins. Its uses have ranged from drug design and discovery, to studies on the protein-DNA binding interfaces, to understanding phylogenetic coevolution of proteins (26, 27). Here, we present a comprehensive SASA-based structural analysis of MEP-HLA complexes to further inform the immunogenicity of a given neoantigen before its inclusion in a vaccine. Importantly, the automated calculation of SASA by GETAREA is also accompanied by “in” or “out” annotation on orientation of the mutant residue — overall providing a comprehensive readout for solvent exposure.

Furthermore, and to a higher level of atomistic detail, our prespecified modeling and  $\beta$ -testing both show that minor changes in the positioning of mutated residue within the MEP can have a profound



**Table 4. Solvent-accessible surface area (SASA) and annotations of neopeptide orientation of the variant peptides**

MEP	Residue	Total	Apolar	Backbone	Sidechain	Ratio (%)	In/out
77-3 (LSMGKLEKM)	Leu 1	3.32	3.32	0.00	3.32	2.30	i
	<b>Ser 2</b>	<b>34.55</b>	<b>25.03</b>	<b>2.63</b>	<b>31.92</b>	<b>41.20</b>	<b>o</b>
	Met 3	15.88	9.36	8.25	7.23	4.60	i
	Gly 4	22.61	19.45	22.61	0.00	25.90	o
	Lys 5	78.34	70.83	2.27	76.06	46.20	o
	Leu 6	0.99	0.37	0.62	0.37	0.30	i
	Glu 7	8.83	8.39	6.09	2.74	1.90	i
	Lys 8	81.26	52.66	5.00	76.26	46.40	o
	Met 9	31.72	27.19	1.52	30.21	19.10	o
77-5 (FNLSMGKL)	Phe 1	41.83	41.83	0.00	41.83	23.20	o
	Asn 2	0.12	0.12	0.12	0.00	0.00	i
	Leu 3	10.49	10.49	5.06	5.43	3.70	i
	<b>Ser 4</b>	<b>51.27</b>	<b>22.94</b>	<b>3.21</b>	<b>48.06</b>	<b>62.10</b>	<b>o</b>
	Met 5	15.32	1.36	14.48	0.84	0.50	i
	Gly 6	25.52	24.10	25.52	0.00	29.30	o
	Lys 7	91.94	56.31	4.11	87.83	53.40	o
	Leu 8	1.95	1.37	0.59	1.37	0.90	i
77-5v1 (LFNLSMGK)	Leu 1	5.43	1.94	3.49	1.94	1.30	i
	Phe 2	4.28	2.62	2.54	1.74	1.00	i
	Asn 3	10.02	5.14	6.28	3.74	3.30	i
	Leu 4	98.95	95.69	3.26	95.69	65.40	o
	<b>Ser 5</b>	<b>20.60</b>	<b>0.71</b>	<b>12.61</b>	<b>7.99</b>	<b>10.30</b>	<b>i</b>
	Met 6	53.10	52.78	3.89	49.21	31.10	o
	Gly 7	12.86	11.03	12.86	0.00	14.80	i
	Lys 8	4.48	0.00	4.48	0.00	0.00	i
77-5v2 (SLFNLSMGK)	Ser 1	4.35	0.36	0.29	4.07	5.30	i
	Leu 2	6.74	5.65	1.09	5.65	3.90	i
	Phe 3	28.44	25.24	3.45	25.00	13.90	i
	Asn 4	40.83	5.90	5.78	35.05	30.70	o
	Leu 5	170.96	156.02	29.72	141.23	96.60	o
	<b>Ser 6</b>	<b>37.89</b>	<b>16.53</b>	<b>14.74</b>	<b>23.16</b>	<b>29.90</b>	<b>i</b>
	Met 7	85.28	80.96	10.38	74.90	47.30	o
	Gly 8	27.36	23.65	27.36	0.00	31.40	i
	Lys 9	6.68	1.65	5.03	1.65	1.00	i
77-6 (IYSLFNLSM)	Ile 1	16.47	16.47	0.00	16.47	11.20	i
	Tyr 2	4.23	2.27	0.00	4.23	2.20	i
	<b>Ser 3</b>	<b>12.63</b>	<b>5.10</b>	<b>0.34</b>	<b>12.29</b>	<b>15.90</b>	<b>i</b>
	Leu 4	104.87	95.67	9.27	95.60	65.40	o
	Phe 5	134.09	130.86	15.81	118.27	65.70	o
	Asn 6	17.72	13.91	10.92	6.80	5.90	i
	Leu 7	40.13	39.07	3.73	36.39	24.90	o
	Ser 8	37.26	16.64	1.19	36.07	46.60	o
	Met 9	0.28	0.00	0.28	0.00	0.00	i

Readouts from GETAREA on modeled MEP-MHC I template complexes. Shown are the MEP numbers, sequences, amino acids in the 9-mers (with mutated residue shown in bold), total SASA values (total), apolarity, and contribution of the backbone and side chains in each postprocessed PDB file. The 'in' or 'out' configuration is also annotated for each residue.

effect on T cell recognition and immunogenicity. While for MEP 44-2 (LSCLNWSTL), the complex displayed a high SASA, a solvent-exposed Cys side chain, and a robust CD8<sup>+</sup> T cell response, a shift of the Cys residue by just 1 amino acid (SCLNWSTLV) in MEP 44-2 buried the Cys, reduced SASA by ~100-fold, and rendered the MEP less immunogenic. In our  $\beta$ -testing protocol, we created 2 variants of immunogenically responsive MEP 77-5, FNLSMGKL  $\rightarrow$  LFNLSMGK and FNLSMGKL  $\rightarrow$  SLFNLSMGK, by shifting the mutated Ser from position 4 to position 5 or 6 in the 8-mer. Both a single and 2-amino-acid shuffle reduced SASA, buried the Ser residue, and diminished immunogenicity.

**Table 5. Solvent-accessible surface area (SASA) and annotations of neopeptide orientation of MEP 84, 175 and 237**

MEP	Residue	Total	Apolar	Backbone	Sidechain	Ratio (%)	In/out
84-1 (VVGNNHNVQV)	Val 1	7.54	7.54	2.68	4.85	4.00	i
	Val 2	1.74	0.98	0.76	0.98	0.80	i
	<b>Gly 3</b>	<b>7.29</b>	<b>7.29</b>	<b>7.29</b>	<b>0.00</b>	<b>8.40</b>	<b>i</b>
	Asn 4	68.99	8.22	2.72	66.27	58.00	o
	His 5	9.15	6.34	5.56	3.59	2.30	i
	Asn 6	63.65	17.32	4.79	58.86	51.50	o
	Gln 7	76.60	22.98	7.96	68.64	47.80	o
175-1 (LPLAFTLL)	Val 8	6.29	4.30	2.00	4.30	3.50	i
	Leu 1	4.55	0.00	4.55	0.00	0.00	i
	Pro 2	2.51	2.51	0.00	2.51	2.40	i
	Leu 3	116.97	107.69	9.35	107.62	73.60	o
	Ala 4	76.86	56.00	20.86	56.00	86.30	o
	<b>Phe 5</b>	<b>9.13</b>	<b>3.68</b>	<b>8.18</b>	<b>0.95</b>	<b>0.50</b>	<b>i</b>
	Thr 6	31.91	19.23	8.36	23.55	22.20	i
237-1 (SAYQFRVC)	Leu 7	84.83	80.13	4.70	80.13	54.80	o
	Leu 8	92.80	84.68	7.26	85.53	58.50	o
	Ser 1	9.84	4.76	8.22	1.62	2.10	i
	<b>Ala 2</b>	<b>26.99</b>	<b>24.84</b>	<b>2.16</b>	<b>24.83</b>	<b>38.30</b>	<b>o</b>
	Tyr 3	20.71	16.75	8.26	12.46	6.50	i
	Gln 4	41.30	5.74	18.11	23.19	16.10	i
	Phe 5	83.07	83.07	2.37	80.70	44.80	o
	Arg 6	85.38	37.21	6.40	78.98	40.40	o
	Val 7	52.06	50.46	1.87	50.19	41.00	o
	Cys 8	0.43	0.00	0.27	0.16	0.20	i

Readouts from GETAREA on modeled MEP-MHC I template complexes. Shown are the MEP numbers, sequences, amino acids in the 9-mers (with mutated residue shown in bold), total SASA values (total), apolarity, and contribution of the backbone and side chains in each postprocessed PDB file. The 'in' or 'out' configuration is also annotated for each residue.

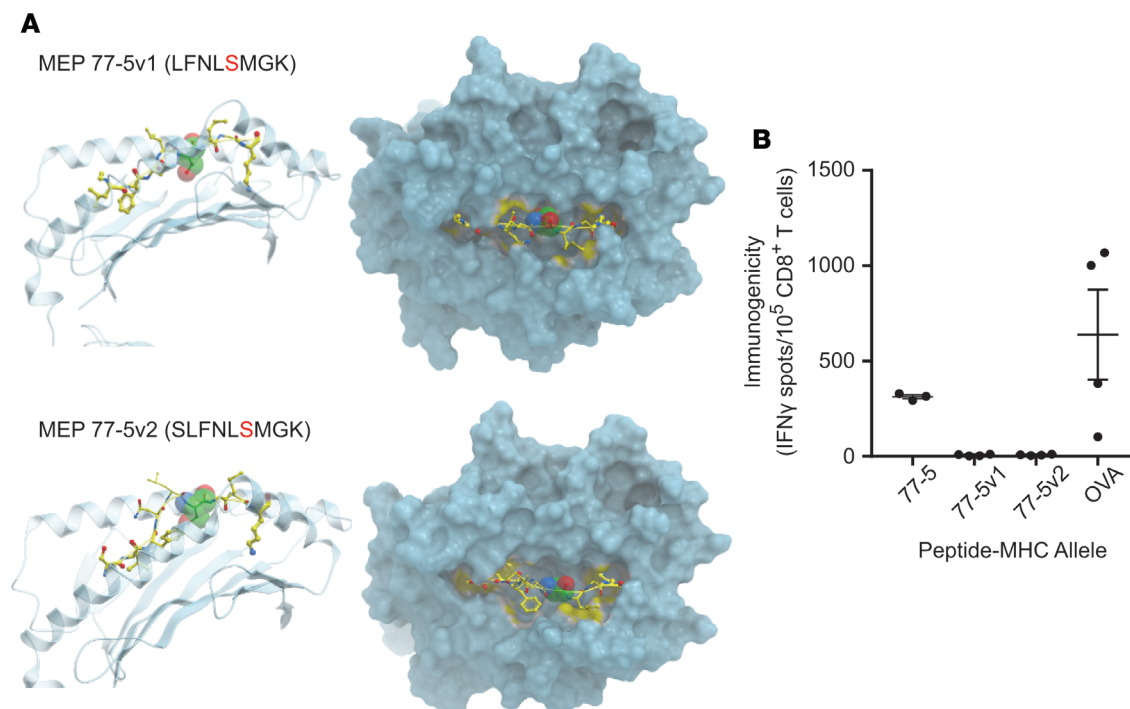
That these atom-level modifications alter immunogenicity would again make it unlikely for current versions of NetMHC, to detect subtle differences in the position of the mutant residue at the MHC I groove, without further training on 3D events.

A recent open-source package, MHCflurry, has been presented as a more accurate predictor of MHC I binding compared with 2 publicly available algorithms, NetMHC<sub>4.0</sub> and NetMHC<sub>pan</sub>, particularly for non-9 amino acid peptides (i.e., peptides of lengths 8 and 10–15) and portends a ~400-fold higher speed (28). It is different from the NetMHC versions in that it uses a publicly trainable architecture and mass spectrometry-identified peptides for model selection (28). While we await more data on outcomes from the new algorithm, it is unclear whether these algorithms can be trained to seek information on protein structure. Given our data, we find that any new module or automated learning machine to more accurately predict neoantigen immunogenicity must have access to PDB and be able to extract the best template. A major challenge is likely to be centered on the selection of the correct template from several high-sequence homology templates for each minimal epitope.

**Table 6. MHC I templates selected from Protein Data Base (PDB) for each variant peptide**

Name	Mutation	MEP	Minimal epitope	Template homolog PDB ID	Sequence similarity %	C $\alpha$ RMSD with template
<i>Rasa3</i>	Y740S	77-5v1	LFNLSMGK	1G7P	50	0.78
		77-5v2	SLFNLSMGK	1S7T	50	0.73

Gene name, point mutation, MEP and its annotation, selected MHC-I template (PDB ID), percent sequence similarity, and C $\alpha$  RMSD values are shown.



**Figure 3.  $\beta$ -Testing of MEP variants for SASA and immunogenicity.** The position of the Ser residue in MEP 77-5 was shifted by 1 amino acid in 2 variant models, MEP 77-5v1 (FNLSMGKL  $\rightarrow$  LFNL\$\\textcolor{red}{S}\$MGK) and MEP 77-5v2 (FNLSMGKL  $\rightarrow$  SLFNL\$\\textcolor{red}{S}\$MGK). **(A)** Side (left panels) and top views (right panels) of modeled complexes between MEPs and selected MHC I template. The ball represents the mutated residue. Calculated SASA and other GETAREA parameters are shown in Table 4. **(B)** Immunogenicity of variant MEPs by ELISPOT for IFN- $\gamma$ -producing T cells. An H-2Kb-binding OVA (OVA-Kb) peptide is used as control (for details, see Methods) ( $n = 3$  mice per group).

## Methods

**Neoantigen identification pipeline.** Extracted DNA and RNA from a murine pancreatic cancer cell line Panc02 (29) was subject to WES and RNA-seq, respectively. The raw sequencing files from WES were aligned to the mm9 reference genome using bowtie2. Variants were called using freebayes and subsequently annotated using Annovar (2). Tumor-specific peptide sequences were extracted using R and analyzed by locally maintained NetMHC 3.2, 3.4, and pan 2.8 algorithms ([https://github.com/rosgood/Panc02\\_Variant\\_ID/commit/899bdd743f671822ef17749769d100ae7f65e698](https://github.com/rosgood/Panc02_Variant_ID/commit/899bdd743f671822ef17749769d100ae7f65e698)) (2). All mutated epitopes predicted to bind to H-2Kb or H-2Db with an affinity of  $\leq 1000$  nM were manually curated by Integrated Genomics Viewer. Pipeline is available on git repository at [https://github.com/rosgood/Panc02\\_Variant\\_ID/commit/899bdd743f671822ef17749769d100ae7f65e698](https://github.com/rosgood/Panc02_Variant_ID/commit/899bdd743f671822ef17749769d100ae7f65e698). For neoepitopes with predictions of  $\leq 1000$  nM, corresponding 20-mer peptides with mutations centered at position 11 were synthesized initially at 70% purity by Peptide 2.0. Further details are available in ref. 15.

**Immunogenicity testing.** Groups of 6-week-old male C57BL/6 mice (3–4 mice per group) each received a pool of 5–6 SLPs at 50  $\mu$ g per peptide, the known H-2Kb-binding OVA (amino acid positions 152–171) peptide GLEQLESIINFEKLTEWTSS (10  $\mu$ g), and poly IC:LC (10  $\mu$ g) (InVivoGen). An initial injection of the vaccine into the hind limb was followed by a boost at day 7 and IFN- $\gamma$  ELISPOT at day 13 to evaluate CD8<sup>+</sup> T cell responses (30). We also mapped the minimal epitopes of the respective 20-mers by similarly testing 8- to 11-amino acid-long peptides corresponding to the expected minimal immunogenic epitopes as predicted by NetMHC to have binding affinities of  $\leq 1000$  nM (15).

**T cell isolation and ELISPOT.** CD8<sup>+</sup> T cells were isolated from freshly harvested splenocytes by first creating a single-cell suspension by passing the spleen through a 40- $\mu$ m filter in CTL media (RPMI with 10% FBS [Thermo Fisher Scientific], 0.5% L-glutamine, 1% penicillin/streptomycin [Invitrogen], and 0.05 mM 2-mercaptoethanol [Thermo Fisher Scientific]). Erythrocytes were removed by ACK lysis, and the resulting T cells were washed, counted, and isolated using the Dynal CD8<sup>+</sup> negative isolation kit (Dynal, Invitrogen) per manufacturer's instructions. Multiscreen 96-well filtration plates (MilliporeSigma) were coated overnight at 4°C with 100  $\mu$ L/well of anti-mouse IFN- $\gamma$  mAb AN18 (100  $\mu$ g/mL, Mabtech, catalog 3321-3-250).

Wells were washed 3 times each with PBS and blocked for 2 hours with CTL media at 37°C. A total of  $1 \times 10^5$  T2 APCs was pulsed with 2 µg peptide in 100 µL CTL media for 2 hours at 37°C, 5% CO<sub>2</sub>. After T cell isolation,  $1 \times 10^5$  CD8<sup>+</sup> T cells were added to the capture plate in 100 µL CTL, followed by the addition of T2 APCs with peptide and incubation for 18 hours at 37°C at 5% CO<sub>2</sub>. Cells were removed from the plate by washing 6 times, 2 minutes per wash, with PBS plus 0.05% Tween-20 (MilliporeSigma). Wells were incubated for 2 hours at room temperature with 10 µg/mL biotinylated anti-mouse IFN-γ mAb R4-6A2 (Mabtech, catalog 3321-6-250) in 0.05% FBS diluted in PBS. Wells were washed as before, incubated with avidin peroxidase complex (Vectastain ELITE ABS kit; Vector Laboratories) for 1 hour at room temperature, and washed again. AEC substrate was added, and wells were developed for 10–15 minutes at room temperature. The reaction was stopped with tap water, and plates were allowed to dry for 24 hours before they were counted using an automated image ELISPOT reader (ImmunoSpot).

**Structural modeling.** To understand the structural influence of the mutated residues in the MEP as a function of its spatial position and configuration, we constructed multiple MEP-HLA complexes. The template structures to model the peptides were identified by searching the PDB using PSI-BLAST (31) for their corresponding homolog in H-2Kb and H-2Db structural sets. MEP 44 and 66 were modeled as MHC I H-2Db, while MEP 77, 84, 175, and 237 were modeled as MHC I MHC I H-2Kb complexes. Hits that exhibited highest sequence identity were selected. The MEP and template sequences were aligned using ClustalX (32). The final template was chosen from the selected set, wherein the mutated residue was solvent exposed.

We generated models of each MEP using MODELLER v9.20 (33) and evaluated stereochemical properties using PROCHECK (34) and ProSA (35). Multiple models were generated, with a final model being chosen based on low-energy function and a low Cα RMSD overlap between the template and the modeled MEP. The peptide in the protein-peptide complex was then optimized using a Rosetta-based algorithm implemented in the FlexPepDock server (36). In short, the starting structure was refined in 200 independent FlexPepDock simulations. Of these, one half was run in the high-resolution model, while the other 100 included a low-resolution preoptimization step, followed by high-resolution refinement. Models generated were then ranked based on their Rosetta generic full-atom energy score. The final model was selected based on solvent accessibility of the mutated residue in the MEP and the lowest energy score from the FlexPepDock optimization.

**Derivative measures.** We derived SASA of each residue of a given MEP using GETAREA, a program implemented within FANTOM (<http://curie.utmb.edu/getarea.html>). In addition to SASAs, this calculation yielded atomic solvation energies and their gradients for the macromolecules. Briefly, the server accepts PDB coordinates as an input file and reports several outputs, including total surface area, apolarity, and contribution of the backbone and side chains in a postprocessed PDB file (Table 3, Table 4, and Table 5). Furthermore, an indication of whether the amino acid is buried (in) or solvent exposed (out) is annotated for each residue in the PDB file (Table 3, Table 4, and Table 5). All structural calculations and figures were generated using ICM-Pro software (25).

**Study approval.** All animal studies were reviewed and approved by the Johns Hopkins IACUC and Biohazards Committee. Animals were kept in pathogen-free conditions and were treated in accordance with institutional and American Association of Laboratory Committee policies. All efforts were made to limit animal pain and discomfort.

## Author contributions

Study conception and design were contributed by NZ, SH, and EMJ; acquisition of data was contributed by NZ, MS, FC, HK, JM, MY, TDA, and SH; analysis and interpretation of data were contributed by NZ, MS, FC, HK, and SH; drafting of manuscript was contributed by NZ, SH, and EMJ; study supervision and critical revision were contributed by EMJ, SH, and NZ; and all authors read and approved the manuscript.

## Acknowledgments

The authors gratefully acknowledge the James W. and Frances Gibson McGlothlin Foundation; The Skip Viragh Center for Pancreatic Cancer at Johns Hopkins; The Bloomberg-Kimmel Institute for Cancer Immunotherapy at Johns Hopkins; The Lustgarten Foundation for Pancreatic Cancer Research; and a Stand Up To Cancer-Lustgarten Foundation Pancreatic Cancer Convergence Dream Team Translational Research Grant (SU2C-AACR-DT1414 to EMJ). NZ was the recipient of the ASCO Young Investigator Award, the GI SPORE Career Enhancement Award, and the Linda Rubin Fellowship for Pancreatic Cancer Research and Patient Care.

Address correspondence to: Elizabeth M. Jaffee or Neeha Zaidi, 4M07 (EMJ), 487 (NZ) Bunting Blaustein Cancer Research Building, 1650 Orleans Street, Baltimore, Maryland 21287, USA. Phone: 410.955.2957; Email: ejaffee@jhmi.edu (EMJ). Phone: 410.614.9006; Email: nzaidi1@jhmi.edu (NZ).

1. Laheru DA, Pardoll DM, Jaffee EM. Genes to vaccines for immunotherapy: how the molecular biology revolution has influenced cancer immunology. *Mol Cancer Ther.* 2005;4(11):1645–1652.
2. Popovic A, Jaffee EM, Zaidi N. Emerging strategies for combination checkpoint modulators in cancer immunotherapy. *J Clin Invest.* 2018;128(8):3209–3218.
3. Zaidi N, Jaffee EM. Immunotherapy transforms cancer treatment. *J Clin Invest.* 2019;129(1):46–47.
4. Karasaki T, et al. Identification of Individual Cancer-Specific Somatic Mutations for Neoantigen-Based Immunotherapy of Lung Cancer. *J Thorac Oncol.* 2016;11(3):324–333.
5. Katsnelson A. Mutations as munitions: Neoantigen vaccines get a closer look as cancer treatment. *Nat Med.* 2016;22(2):122–124.
6. van Rooij N, et al. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol.* 2013;31(32):e439–e442.
7. Verdegaal EM, et al. Neoantigen landscape dynamics during human melanoma-T cell interactions. *Nature.* 2016;536(7614):91–95.
8. Bailey P, et al. Exploiting the neoantigen landscape for immunotherapy of pancreatic ductal adenocarcinoma. *Sci Rep.* 2016;6:35848.
9. Keskin DB, et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature.* 2019;565(7738):234–239.
10. Zaidi N, Jaffee EM. Immune cells track hard-to-target brain tumours. *Nature.* 2019;565(7738):170–171.
11. Osipov A, Zaidi N, Laheru DA. Dual Checkpoint Inhibition in Pancreatic Cancer: Revealing the Limitations of Synergy and the Potential of Novel Combinations. *JAMA Oncol.* 2019;5(10):1438–1439.
12. Ott PA, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* 2017;547(7662):217–221.
13. Sahin U, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature.* 2017;547(7662):222–226.
14. Hilf N, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature.* 2019;565(7738):240–245.
15. Kinkad HL, et al. Combining STING-based neoantigen-targeted vaccine with checkpoint modulators enhances antitumor immunity in murine pancreatic cancer. *JCI Insight.* 2018;3(20):122857.
16. Twardowski P, et al. A phase I trial of GEN-009, a neoantigen vaccine using ATLAS™, an autologous immune assay, to identify immunogenic and inhibitory tumour mutations. *Ann. Oncol.* 2019; 30(Supplement 5):doi:10.1093/annonc/mdz253.
17. Balachandran VP, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature.* 2017;551(7681):512–516.
18. Hopkins A, Jaffee E. Pancreatic cancer: Next-generation algorithms for neoantigen selection. *Nat Rev Gastroenterol Hepatol.* 2018;15(3):135–136.
19. Yadav M, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature.* 2014;515(7528):572–576.
20. Riley TP, et al. Structure Based Prediction of Neoantigen Immunogenicity. *Front Immunol.* 2019;10:2047.
21. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* 2008;36(Web Server issue):W509–W512.
22. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics.* 2008;24(11):1397–1398.
23. Sylvester-Hvid C, et al. SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation. *Tissue Antigens.* 2004;63(5):395–400.
24. Nielsen M, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003;12(5):1007–1017.
25. Abagyan R, Frishman D, Argos P. Recognition of distantly related proteins through energy calculations. *Proteins.* 1994;19(2):132–140.
26. McConkey BJ, Sobolev V, Edelman M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics.* 2002;18(10):1365–1373.
27. Ochoa D, García-Gutiérrez P, Juan D, Valencia A, Pazos F. Incorporating information on predicted solvent accessibility to the co-evolution-based study of protein interactions. *Mol Biosyst.* 2013;9(1):70–76.
28. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst.* 2018;7(1):129–132.e4.
29. Corbett TH, et al. Induction and chemotherapeutic response of two transplantable ductal adenocarcinomas of the pancreas in C57BL/6 mice. *Cancer Res.* 1984;44(2):717–726.
30. Thomas AM, et al. Mesothelin-specific CD8(+) T cell responses provide evidence of in vivo cross-priming by antigen-presenting cells in vaccinated pancreatic cancer patients. *J Exp Med.* 2004;200(3):297–306.
31. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–3402.
32. Larkin MA, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–2948.
33. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. *Meth Enzymol.* 2003;374:461–491.
34. Laskowski RA, Moss DS, Thornton JM. Main-chain bond lengths and bond angles in protein structures. *J Mol Biol.* 1993;231(4):1049–1067.



35. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35(Web Server issue):W407–W410.
36. Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins.* 2010;78(9):2029–2040.