

OPTIMIZATION OF THERAPEUTIC PROTEINS TO DELETE T-CELL EPITOPES WHILE MAINTAINING BENEFICIAL RESIDUE INTERACTIONS

Andrew S. Parker¹

Karl E. Griswold^{2,*}

Chris Bailey-Kellogg^{1,*}

¹ *Department of Computer Science, Dartmouth College
6211 Sudikoff Laboratory, Hanover, NH 03755, USA
phone: 603-646-3385; fax: 603-646-1672
{asp,cbk}@cs.dartmouth.edu*

² *Thayer School of Engineering, Dartmouth College
8000 Cummings Hall, Hanover, NH 03755, USA
phone: 603-646-2127; fax: 603-646-3856
karl.e.griswold@dartmouth.edu*

Exogenous enzymes, signaling peptides, and other classes of non-human proteins represent a potentially massive but largely untapped pool of biotherapeutic agents. Adapting a foreign protein for therapeutic use poses numerous design challenges. We focus here on one significant problem: modifying the protein to mitigate the immune response mounted against “non-self” proteins, while not adversely affecting the protein’s stability or therapeutic activity. In order to propose such variants suitable for experimental evaluation, this paper develops a computational method to select sets of mutations predicted to delete immunogenic T-cell epitopes, as evaluated by a 9-mer potential, while simultaneously maintaining important residues and residue interactions, as evaluated by one- and two-body potentials. While this design problem is NP-hard, we develop an integer programming approach that works very well in practice. We demonstrate the effectiveness of our approach by developing plans for biotherapeutic proteins that, in previous studies, have been partially deimmunized via extensive experimental characterization and modification of limited segments. In contrast, our global optimization technique considers an entire protein and accounts for all residues, residue interactions, and epitopes in proposing candidates worth subjecting to experimental evaluation.

1. Introduction

Efforts to employ exogenous proteins as therapeutic agents pose significant challenges for protein engineering. Many therapeutic proteins, particularly those derived from non-human sources, are recognized as “non-self” and elicit an immune response in human patients. For the special case of therapeutic antibodies generated in animal models, decades of research and extensive knowledge of their structure and function has yielded grafting techniques that effectively reduce immunogenicity^{11, 6, 8}. However, to effectively leverage other protein classes (enzymes, signaling peptides, etc.) as next-generation drugs, more advanced and widely applicable strategies for functional deimmunization are necessary.

The immune response against a foreign protein is driven by molecular recognition of immunogenic peptides, or *epitopes*, that are found within the protein’s primary sequence. The recognition process begins with proteolytic processing of the offending protein and subsequent loading of the peptide fragments within the groove of type II major histocompatibility complex (MHC II) proteins in antigen presenting cells. Surface display of immunogenic peptides then

enables formation of ternary peptide-MHC II-T-cell receptor complexes via interaction with surface receptors of cognate white blood cells. The ensuing signaling cascade ultimately leads to a coordinated anti-protein immune response.

In the context of biotherapeutic deimmunization, the recognition of and response against the foreign protein can be averted by identifying immunogenic peptide fragments within the protein and mutagenizing key residues so as to disrupt the fragments’ capacity to complex with the MHC II and/or T-cell receptors. This *epitope deletion* approach has been successfully applied to several therapeutic candidates including staphylokinase²⁷, erythropoietin²¹, and factor VIII⁷ (all of which serve as case studies for our results). However, epitope deletion typically requires extensive experimentation, including synthesizing and testing the immunogenicity of large peptide panels taken from the native protein sequence, performing alanine scanning mutagenesis on the most immunogenic fragments to pinpoint critical MHC II binding residues, incorporating deimmunizing mutations into the full length protein, and finally testing the functionality and immunogenicity of the

*Corresponding authors.

engineered protein variants, only a small fraction of which are likely to retain high activity and/or constitute globally deimmunized candidates.

In order to avoid much of the extensive experimental effort previously required, we seek to computationally identify sets of mutations that are predicted to delete epitopes while maintaining the protein’s stability and therapeutic activity. For evaluating epitopes, we make use of the burgeoning development of sequence-based T-cell epitope predictors^{20, 17, 13, 3, 26, 2}, which employ profiles representing amino acid choices favorable for interaction with the MHC II binding groove. These epitope predictors have been successfully used in the analysis of biotherapeutic candidates^{1, 10}, as well as in vaccine development^{4, 9, 12}. For evaluating effects on stability and activity, we employ a position-specific statistical model (with terms for single positions and pairs of positions) learned from a family of related proteins, taking advantage of the fact that constraints on amino acid choices required to maintain structure and function are likely to be manifested in the sequence record. The application of one- and two-body sequence potentials has also been essential in a number of protein engineering contexts^{25, 16, 18}.

This paper presents the first method that optimizes variants of a target protein by deleting immunogenic epitopes (as evaluated by epitope potentials) while maintaining important residues and residue interactions (as evaluated by one- and two-body potentials). It builds upon, but goes significantly beyond, our earlier dynamic programming algorithm that identified conservative mutations predicted to delete T-cell epitopes¹⁴. Our previous dynamic programming approach evaluated the structural and functional consequences of each mutation independently of other mutations. While that is suitable for small numbers of relatively well-separated mutations, it does not account for residue interactions that may underlie stability and activity, and that are increasingly important with higher mutational loads. Unfortunately, the use of a residue interaction score results in an NP-hard optimization problem¹⁵. However, we develop here an integer programming approach that concisely represents the problem constraints and objectives and works quite

well in practice. We demonstrate the effectiveness of our method in application to redesigning proteins of therapeutic value, short limited segments of which have been the target of previous deimmunization efforts.

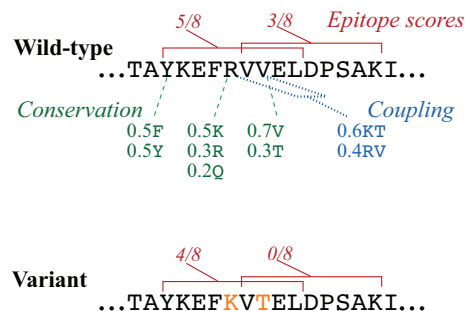


Fig. 1. We employ T-cell epitope predictors to score each 9-mer peptide for potential immunogenicity. In this example two peptides are deemed immunogenic, as they are predicted to be recognized by several of the 8 most representative MHC II alleles. We analyze a family of related proteins for single-position conservation and pairwise coupling (covariation) statistics. Our algorithm then selects a specified number of mutations (here two mutations, underlined in the variant), to optimize a weighted combination of the resulting epitope score and a sequence score capturing consistency with the conservation and coupling statistics. In the example, the variant is predicted to be less immunogenic for one peptide (with one fewer allele recognizing it) and completely deimmunized for the other. The selected substitutions score well under conservation and coupling; while K/V would score somewhat better than the selected K/T under conservation, it would score much worse under coupling and thus is less likely to result in a stable, active protein.

2. Methods

Given a target protein, our goal is to design a variant predicted to have fewer epitopes, by making a fixed number of substitutions that are predicted to maintain the protein’s stability and activity (Fig. 1).

Problem 2.1 (Deimmunization). *Given a protein sequence S of length n , along with a number m of mutations to make, determine an m -mutation variant S' minimizing*

$$\alpha \left(\sum_{i=1}^n \phi_i(S'[i]) + \sum_{i=1}^n \sum_{j=i+1}^n \phi_{i,j}(S'[i], S'[j]) \right) + (1 - \alpha) \sum_{i=1}^{n-8} e(S'[i..i+8])$$

where we have the following contributions to the score

(lower is better for each):

- $e : \mathcal{A}^9 \rightarrow \mathbb{R}$ gives the epitope score for a peptide (we assume a 9-mer; see below)
- $\phi_i : \mathcal{A} \rightarrow \mathbb{R}$ give the position-specific conservation score for an amino acid at position i
- $\phi_{i,j} : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$ gives the positions-specific coupling score for a pair of amino acids at pair of positions i, j

Here and throughout, we use $\mathcal{A}=\{\text{A,C},\dots,\text{Y}\}$ for the set of amino acids; sequences are 1-indexed; and the notation $X[i..j]$ indicates the substring of X from position i to j , inclusive.

2.1. Epitope scores

T-cell epitope predictors encapsulate the underlying specific recognition of an epitope by an MHC II protein. We focus here on the human leukocyte antigen group DR (HLA-DR) of MHC II proteins, since they are the predominant isotype. HLA-DR proteins have a recognition groove whose pockets form energetically favorable interactions with specific side-chains of peptides approximately 9 residues in length. Numerous methods are available for epitope prediction, and they have been shown to be predictive of immunogenicity²⁶. Our current implementation supports two state-of-the-art tools, ProPred²⁰ and SMM-align¹³, but is modular and can readily support others. For brevity, the results presented in this manuscript are all based on ProPred, so we briefly overview it.

Sturniolo et al.²⁰ experimentally measured the binding affinity between individual residues and individual pockets of the MHC II binding groove on a limited set of alleles. They then created binding profiles for untested alleles through sequence and structure alignment with tested alleles. In this “pocket profile” method, TEPITOPE, the sum of position-specific weights for each residue in a 9-mer provides a score that is compared against a threshold to determine whether or not the peptide is in a given percentile of the best-recognized peptides. The approach was experimentally validated by comparing its predictions against HLA-DR selected and non-selected peptide repertoires; up to 80% of the selected peptides were correctly predicted at a threshold that yielded $< 5\%$ false positives. Singh and Raghava

then built a tool, ProPred, to expand the scope of TEPITOPE and make it more easily accessible and applicable¹⁷. In a recent independent evaluation²⁶, ProPred did quite well in epitope prediction, achieving an average 0.73 area under the curve (AUC) across 14 different alleles. ProPred has also been successfully employed in a number of different studies; e.g., it has recently helped identify antigenic sites on a mosquito midgut glycoprotein⁴, immunoreactive peptides in prostatic acid phosphatase⁹, and promiscuous T-cell epitopes of three major secreted antigens of *Mycobacterium tuberculosis*¹². In all three of these examples, ProPred facilitated the rapid identification of potential vaccine targets that were then experimentally characterized in detail. In our case study of Erythropoietin, we found a quite striking match between ProPred predictions and published ELISPOT assay immunogenicity results.

While there are over 50 different HLA-DR alleles, we have focused on 8 common alleles (DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*0801, DRB1*1101, DRB1*1301, and DRB1*1501) that represent the majority of human populations world-wide¹⁹. Thus our epitope score is the fraction of these 8 alleles predicted to recognize a peptide.

2.2. Sequence scores (conservation and coupling)

We base the conservation and coupling scores on the statistical framework from our earlier site-directed recombination work²⁸, but our planning method can use any score of the same form. In fact, it could readily employ a $\Delta\Delta G^\circ$ predictor⁵ instead of a sequence-based score. For consistency with epitope scores, we establish the sequence score so that it is to be minimized. We include both position-dependent residue conservation and coupling (covariation). If a residue is highly conserved to a single amino acid type, it may be important not to mutate it at all; likewise if a few amino acid types dominate the family for a particular residue position, perhaps only those amino acids should be considered for possible substitutions. Similarly, if two residues are highly correlated, it may be necessary to ensure that we make compensating mutations—if we mutate one, we must also mutate the other in order to maintain the covariation that

presumably underlies stability or activity. Our sequence score quantitatively codifies these intuitive conservation and coupling constraints.

Let us consider conservation and covariation in a multiple sequence alignment \mathcal{S} for the family related to our target protein S . For notational convenience, we define $n_i(a)$ as the number of sequences in the family that contain an amino acid a at position i , and similarly $n_{i,j}(a,b)$ as the number of sequences in the family that contain amino acids a and b at positions i and j respectively:

$$n_i(a) = |\{P \in \mathcal{S} : P[i] = a\}| \quad (1)$$

$$n_{i,j}(a,b) = |\{P \in \mathcal{S} : P[i] = a \wedge P[j] = b\}| \quad (2)$$

For the single-position terms, we define $\phi_i(a)$ as the negative log probability (so that more negative is better) of amino acid type a at residue position i :

$$\phi_i(a) = -\log \frac{n_i(a)}{|\mathcal{S}|} \quad (3)$$

For the pairwise terms, we define $\phi_{i,j}(a,b)$ as the negative log probability of the pair of amino acid types a and b , vs. what would be expected if they were independent:

$$\phi_{i,j}(a,b) = -\log \frac{n_{i,j}(a,b)}{|\mathcal{S}|} - \phi_i(a) - \phi_j(b) \quad (4)$$

By subtracting the independent terms from the joint term, $\phi_{i,j}$ contains only the additional information regarding the correlation between the two positions. We can correctly compute a total score by summing up all the singleton and pairwise terms without “double-counting” the singleton contributions.

If we include all pairwise terms, we run the risk of using an overfit model, capturing (and then enforcing) too many and too specific constraints on residue interactions. One approach to limiting the number of pairwise terms is to place a contact restriction: only consider residue pairs that are in contact in a representative structure for the protein family (assumed common to all, by homology), since such pairs have the greatest direct impact on establishing the local environment for a possible mutation. Another approach is to enforce a statistical significance test, such as a χ^2 . For a pair of positions i,j , the χ^2 can be written in terms of our conservation and coupling

scores.

$$\chi^2 = \sum_{a \in S_i} \sum_{b \in S_j} \frac{\left(n_{i,j}(a,b) - \frac{n_i(a) \cdot n_j(b)}{|\mathcal{S}|}\right)^2}{\frac{n_i(a) \cdot n_j(b)}{|\mathcal{S}|}} \quad (5)$$

We compute a p -value according to the χ^2 distribution with $(r-1)(c-1)$ degrees of freedom (where r and c are numbers of different amino acids in the sequence record for columns i and j , respectively), giving the probability that the two residues are actually independent rather than coupled. We employ a simple Bonferroni correction for multiple hypothesis testing, dividing the desired p -value (0.01 for the results shown below) by the number of pairs being tested.

While we have implemented both approaches, for the results we use the χ^2 restriction instead of the contact restriction, since in previous work^{22–24} we have found the purely statistical models to outperform contact-restricted ones in predictive ability, particularly in cases where long-range sequence constraints capture important covariation that is indicative of function.

2.3. Integer programming approach

Due to the pairwise coupling terms, our optimization problem (Problem 2.1) is NP-hard (like other such protein design problems¹⁵). Thus we develop an integer programming approach that incorporates the pairwise terms along with the conservation and epitope terms.

Define singleton binary variable $s_{i,a}$ to indicate whether or not the residue at position i is of amino acid type a . Define pairwise binary variable $p_{i,j,a,b}$, derived from $s_{i,a}$ and $s_{j,b}$ to indicate whether or not the residues at positions i and j are of amino acid types a and b , respectively. Finally, define window binary variable $w_{i,X}$, derived from $s_{i,a}$ through $s_{i+9,b}$, to indicate whether or not the residues in the 9-residue window starting at position i are of the amino acid types in $X \subset \mathcal{A}^9$.

We rewrite our objective function in terms of these binary variables:

$$\Phi = \alpha \left(\sum_{i,a} s_{i,a} \cdot \phi_i(a) + \sum_{i,j,a,b} p_{i,j,a,b} \cdot \phi_{i,j}(a,b) \right) + (1-\alpha) \sum_{i,X} w_{i,X} \cdot e(X) \quad (6)$$

In order to guarantee that the variable assignments yield a valid set of mutations, we impose the following constraints:

$$\forall i : \sum_a s_{i,a} = 1 \quad (7)$$

$$\forall i, a, j > i : \sum_b p_{i,j,a,b} = s_{i,a} \quad (8)$$

$$\forall j, b, i < j : \sum_a p_{i,j,a,b} = s_{j,b} \quad (9)$$

$$\forall i, a \forall h \in 1..9 : \sum_{X: X[h]=a} w_{i,X} = s_{i+h-1,a} \quad (10)$$

$$\sum_{i,a: S[i] \neq a} s_{i,a} = m \quad (11)$$

Eq. 7 ensures that only one amino acid is assigned to a given position. Eq. 8 and Eq. 9 maintain consistency between singleton and pairwise variables, while Eq. 10 maintains consistency between singleton and window variables. Eq. 11 enforces the desired mutation load.

We employ several restrictions to limit the number of variables and make the system amenable to standard IP solvers. We assume that only a small set of substitutions should be considered at each position, namely those that have appeared with frequency greater than 5% at the position in the family. Let us denote by A_i the set of such allowed amino acids at position i . Then we only have $s_{i,a}$ variables for $a \in A_i$. We similarly restrict variables $p_{i,j,a,b}$ to $a \in A_i$ and $b \in A_j$, and create $w_{i,X}$ only for X composed from the allowed amino acid types. We further note that positions P5 and P8 are considered uninformative in matrix-based epitope predictors such as ProPred¹⁷, thereby further reducing the number of variables required (X only has 7 positions rather than 9). Finally, due to the nature of MHC-II binding, epitopes must have a hydrophobic residue from the set {F,I,L,M,V,W,Y} in the first position of the peptide. Thus we only create epitope variable $w_{i,X}$ for X with $X[1]$ in the hydrophobic set.

We implemented and evaluated a further variation on the epitope variables, only creating variables for peptides with non-zero epitope scores (since only they can affect the objective function). We then relax the epitope constraints (Eq. 10) to account for the

fact that only immunogenic peptide variables need appear:

$$\forall i, X : 1/8 \sum_{h=1}^9 s_{h+i-1,X[h]} - w_{i,X} \leq 1 \quad (12)$$

In practice, we found this version to have reduced the structure of the problem sufficiently that the solver took much longer to identify the solution; the reduction in memory usage was not worth the reduction in constraint.

2.4. Implementation

The integer program is solved by the open source COIN-OR cbc solver (<https://projects.coin-or.org/Cbc>). The code to generate the problem formulation and read the solution is implemented in Java. We have placed a limited-capability demonstration at <http://www.cs.dartmouth.edu/~cbk/deimm/>; the full code is available for academic use by contacting the authors.

3. Results

We illustrate the effectiveness of our approach on three biotherapeutic proteins, SakSTAR, Epo, and Factor VIII, previously subjected to expert rational design as well as computational studied with our previous dynamic programming framework¹⁴. In all three cases we performed sequence analysis based on the appropriate Pfam-curated MSA (SakSTAR: PF02821; Epo: PF00758; Factor VIII: PF00754), removing sequences with less than 35% identity to the target and filtering for a maximum of 90% pairwise identity. We considered a substitution to be acceptable if it appeared in at least 5% of the remaining sequences. We then studied the impact of incorporating a pairwise coupling score along with the single-position conservation score. We also considered the impact of the two main parameters to our algorithm: m , the number of substitutions to make, and α , the weight applied to the sequence score (with $1-\alpha$ to epitope score).

The score range for the epitope potential is from 0, when there are no detectable immunogenic peptides, to $(n-8) \cdot a$, when every 9-mer peptide binds to each MHC-II allele (n is the length of the sequence and a the number of alleles). The number of coupling

constraint violations is from 0, when every mutation is benign, to $m \cdot (n - 1)$, when every mutation violates a coupling constraint with every other column (m is the number of mutations). The magnitude of the sequence score depends on the frequencies of the chosen substitutions in the MSA as described in the Methods.

We evaluate here the efficacy of our *computational* method for *global* optimization of protein variants, using three target proteins which have been *experimentally* evaluated under *local* modifications. It is not straightforward to compare and contrast such different methodologies; e.g., since our results are guaranteed to be optimal under the scoring functions, but must account for more residues and interactions. Therefore, we use the literature exclusively as a *qualitative* benchmark, characterizing what was done and the effects found. For a *quantitative* benchmark, we examine the results of our system under the different parameters mentioned above. The primary focus of this work is demonstration of a novel, provably accurate global optimization algorithm operating with several validated bioinformatics models.

3.1. SakSTAR

Warmerdam et al.²⁷ sought to deimmunize the fibrinolytic agent staphylokinase, specifically the SakSTAR wild-type variant derived from a lysogenic *S. aureus* strain. They targeted the C3 region, spanning residues 71–87, which was recognized by 90% of the T-cells cloned from a set of donors. Based on results from alanine scanning mutagenesis, sets of 2–4 alanine substitutions were selected to produce new variants designed to reduce immunogenicity. T-cell proliferation response to this short segment was reduced in the variants.

To assess the potential impact of accounting for coupling in optimizing the SakSTAR C3 region, we computed the sequence score for all 1209 double-mutant plans, using the allowable substitutions as discussed above. See Fig. 2. The mean sequence score is 291, with a standard deviation of 112; the best score is about 41. Even with only double mutants in a small peptide, it is easy to violate important residue interactions; the problem is exacerbated with higher mutational loads and more interactions. As a specific example, in our previous

work¹⁴, which assumed residue independence, we designed a 6-substitution, conservation-based plan that included mutations V112K and F125E. Our new coupling analysis shows that these two positions are significantly coupled (p -value = $3 \cdot 10^{-7}$). In addition, evaluation of the double mutant with Fold-X⁵ finds their $\Delta\Delta G^\circ$ contributions to be non-additive by approximately 1 kcal/mol.

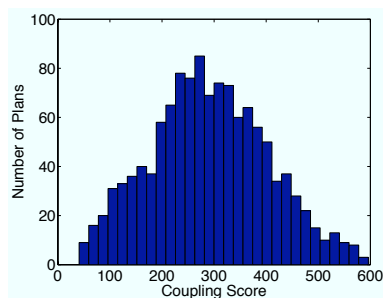


Fig. 2. Histogram of residue coupling scores for all double mutants (using substitutions allowed by conservation statistics) of the C3 region of SakSTAR.

We applied our method to optimize the *entire* SakSTAR protein, addressing a weakness identified in the published experimental methodology: the “vast majority of humans recognize additional immunogenic SakSTAR regions”²⁷. Fig. 3 summarizes the plans and scores. Plans optimized for both conservation and coupling trend downward exponentially in sequence score as the objective function is progressively weighted more heavily towards it. Plans optimized only for conservation trend flat or erratically as the weight is changed; these plans miss critical pairwise interactions.

By placing weight on the sequence score we trade off some minimization of immunogenicity for maximization of stability preservation. For example, in Fig. 3 the first marker in the 2-substitution coupling series (a) is a plan with $\alpha = 0$, which deletes 22 epitopes but incurs a sequence penalty of about 200 (including over 30 pairwise residue interaction clashes). The fourth marker in that series is a plan with $\alpha = 0.1$, which deletes four fewer epitopes, but with a sequence score an order of magnitude better at around 30. At this inflection point in the series most of the pairwise residue interactions have been eliminated in the sequence score. The remaining part

of the score is the independent residue conservation penalty for each substitution. Increasing the weight on the sequence score in the objective function has diminishing returns for stability improvement. Interestingly, the 2- and 4-substitution coupling series intersect near this point. As expected, with the pairwise constraints satisfied, the 2-substitution series can finish best by choosing pairs of mutations with small conservation penalties, but the 4-substitution series must sum over sets of four.

By examining the contributions to the coupling score, we can gain insights into the important residue interactions that might affect the stability and activity of a variant. For example, 2-substitution plans with α of 0 (no sequence score) and 0.1 differ by selecting V112K or L127T, respectively. They both choose M26S. Residue 112 is coupled with over 20 other residues, while position 26 is relatively independent and position 127 is completely independent. This results in more “conflicts” (positive contributions to the coupling score) in the $\alpha = 0$ plan (Fig. 4, left). Furthermore, shifting weight from epitope score to a conservation-only (no coupling) score doesn’t help; a similar property can be observed comparing plans using conservation only vs. those with both conservation and coupling (Fig. 4, right).

Our method enables an engineer to assess the trade offs between optimizing for deimmunization vs. for stability. In addition to the trends observable in Fig. 3, specific plans can be assessed for what is contributing to the objective function. For example, Fig. 5 shows a 6-substitution plan at $\alpha = 0.1$, which deletes 32 epitopes at a sequence score of 80. This sequence score is better than 98% of all 2-substitution plan scores, shown in Fig. 2, and this epitope score is only slightly worse (5 epitopes) than that of a conservation-only plan. Importantly, the plan avoids the central region of the primary sequence, which is known to contain segments critical for therapeutic activity²⁷. While Warmerdam et al. focus on this region, our system targets flanking areas of higher predicted immunogenicity and fewer coupling constraints to achieve global optimization.

3.2. Epo

Tangri et al.²¹ focused on two regions in the protein therapeutic erythropoietin (Epo), residues 101–

115 and 136–150, which they experimentally determined to be immunogenic during an intensive analysis of peptide fragments spanning the entire length of the protein. They engineered four variants targeting the anchor residues of identified T-cell epitopes in these regions: L102P/S146D (named G2), T107D/S146D (G3), L102G/T107D/S146D (G4), and L102S/T107D/S146D (G5). Variants G3 and G4 reduced response in the ELISPOT assay. However, variants G2 and G5 were not bioactive, possibly because of non-conservative substitutions at L102.

As with SakSTAR, optimizing for a combination of epitope and sequence scores allows us to explore the trade-offs between deimmunizing and maintaining stability (Fig. 6). Again we see that it is necessary to explicitly include a pairwise coupling score in general, in order to produce variants with predicted good residue interactions. In this case, though, we see that the conservation-only and conservation + coupling plans reunite at both extremes of α , implying that for those values the optimal residues for sequence score alone and for epitope score alone are relatively independent. However, in striking a balance between those two factors, the conservation-only plans score very poorly in terms of residue interactions.

We considered a number of plans optimized by our algorithm with varying mutation numbers and sequence score weights. An 8-substitution, $\alpha = 0.1$ plan eliminates 50 epitopes with a sequence score of about 30. Ordinarily an engineer may cautiously stop at 8 mutations, but because our method can evaluate pairwise interactions, we can more confidently increase the mutational load and adjust the weight to ensure that sufficient attention is paid to maintaining sequence constraints predictive of stability and activity. In so doing, we found a 16-substitution solution that dominates the 8-substitution solution, deleting 50% more epitopes while still maintaining a low sequence score of about 30 (Fig. 7). This plan reduces epitopes in all immunodominant regions of the protein, including those targeted by Tangri et al., while respecting pairwise residue interactions. This is one example of how our algorithm can be employed to maintain predicted protein stability and function under higher mutational loads, thus permitting more efficient epitope

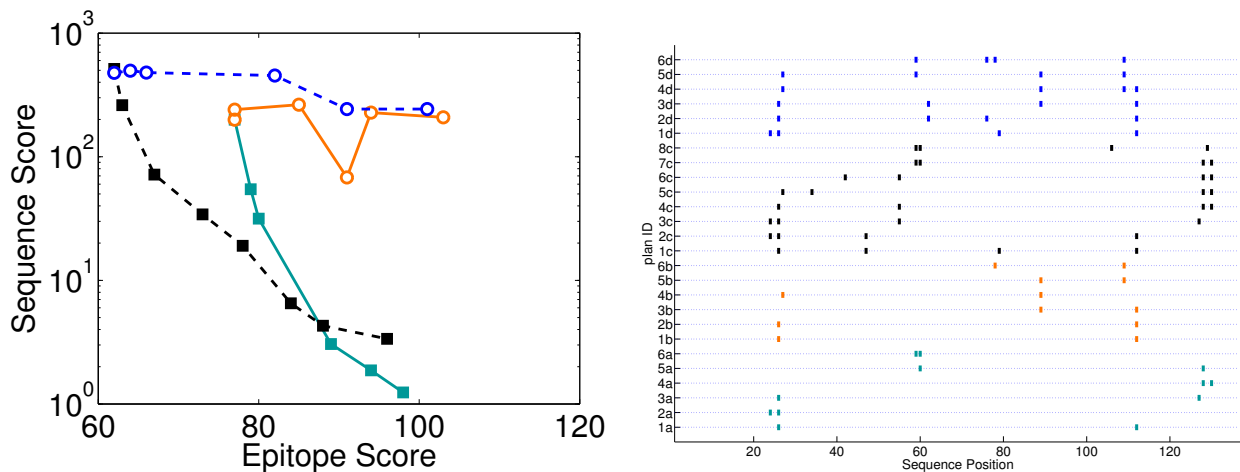


Fig. 3. SakSTAR plans optimized for varying tradeoffs between epitope and sequence scores. **(left)** Scores. The weight α increases monotonically but not uniformly from left to right. Cyan square solid (a): 2 mutations, conservation and coupling; Orange circle solid (b): 2 mutations, conservation only; Black square dashed (c): 4 mutations, conservation and coupling; Blue circle dashed (d): 4 mutations, conservation only. **(right)** Sequence positions of the substitutions.

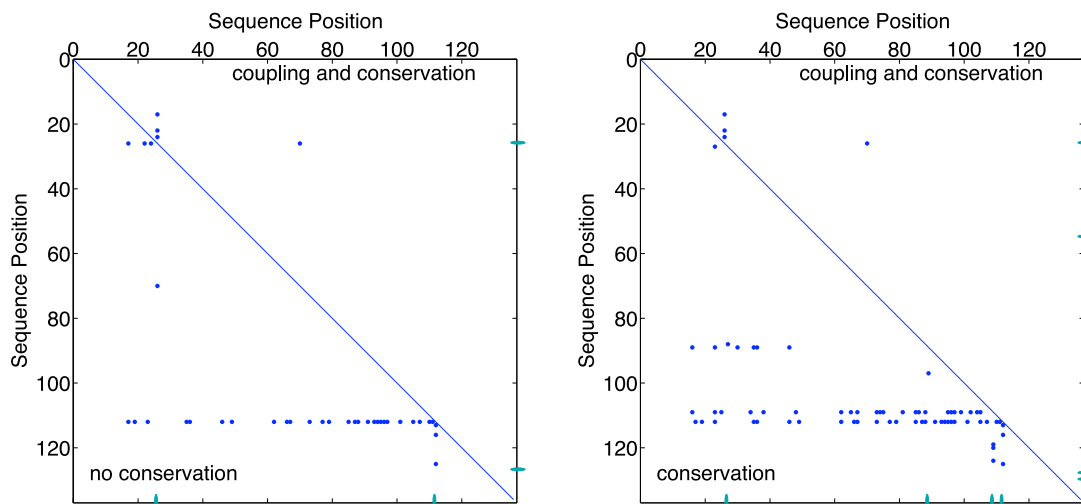


Fig. 4. Pairwise conflicts incurred by SakSTAR protein variants. **(left)** 2 substitutions. Lower triangle: no sequence score ($\alpha = 0$); upper triangle: some sequence score ($\alpha = 0.1$). **(right)** 4 substitutions. Lower triangle: conservation only; upper triangle: both conservation and coupling; both have $\alpha = 0.1$.

deletion.

3.3. Factor VIII

Jones et al.⁷ sought to remove CD4+ T-cell epitopes from the C1 domain of Factor VIII (FVIII). Frequent administration of replacement FVIII treats hemophilia successfully, but in 20%-50% of cases, it leads to the development of neutralizing FVIII antibodies. Jones et al. screened 475 peptides but chose only one for further study. They made a double mutant, M2104K / L2107N, which mitigated the stim-

ulation of CD4+ T-cells in that peptide only and maintained protein activity *in vitro*.

Our method globally searched the *entire* C1 domain for substitutions that optimally reduced epitopes and respected the sequence constraints from conservation and coupling statistics. The 4-mutation, $\alpha = 0.2$ plans under conservation only and with conservation plus coupling both made substitutions in three immunodominant regions and reduced the total epitope score by about 30%. However, the conservation-only approach violates 28 coupling con-

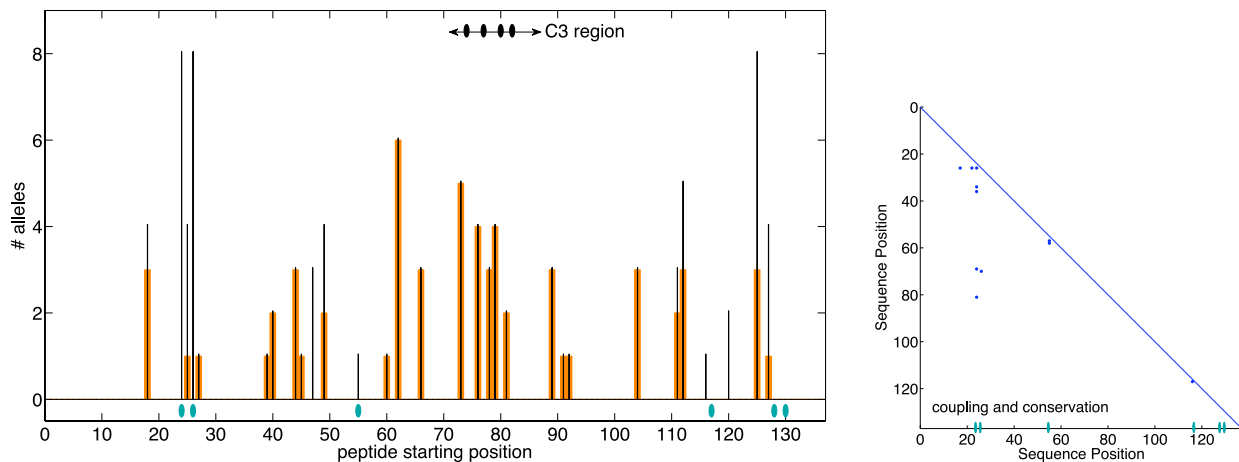


Fig. 5. Optimized 6-substitution, $\alpha = 0.1$ SakSTAR variant. **(left)** Epitope profile. x -axis: starting position of each 9-mer; y -axis: predicted number of alleles recognizing the 9-mer. Thin black bars indicate wild-type scores and thick orange bars indicate variant scores. Note: wild-type epitope scores are always greater than or equal to corresponding variant ones; i.e., we never introduce new epitopes. Blue ellipses indicate mutated positions. Black ellipses at top indicate Warmerdam et al. mutated positions. **(right)** Coupling violations.

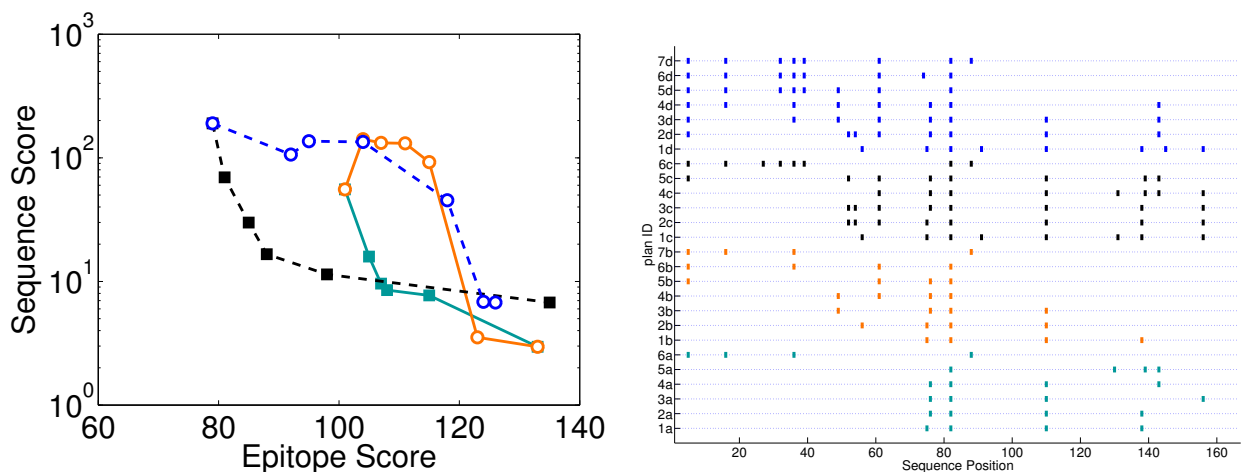


Fig. 6. Epo plans optimized for varying tradeoffs between epitope and sequence scores. **(left)** Scores. The weight α increases monotonically but not uniformly from left to right. Cyan square solid (a): 2 mutations, conservation and coupling; Orange circle solid (b): 2 mutations, conservation only; Black square dashed (c): 4 mutations, conservation and coupling; Blue circle dashed (d): 4 mutations, conservation only. **(right)** Sequence positions of the substitutions.

straints, while the conservation plus coupling method violates only 3. As a result, the sequence score of the former is an order of magnitude worse: 125 versus 14, respectively. Similar to Epo and SakSTAR, a higher mutational load eliminates more epitopes at the cost of a higher sequence score; and again we observe that considering coupled residues dramatically lowers sequence score compared to conservation only plans at both 2 and 4 mutations (Fig. 8). Our FVIII results further establish that computational protein

redesign in the context of epitope deletion demands incorporation of pairwise residue interactions.

FVIII is predicted to be highly immunogenic relative to SakSTAR or Epo. The sequence of the C1 domain of FVIII has a greater proportion of large hydrophobic residues that potentially anchor component peptides into the MHC-II binding groove. Furthermore, the sequence is tightly coupled. High scores in both dimensions present a challenging global optimization problem. Our method

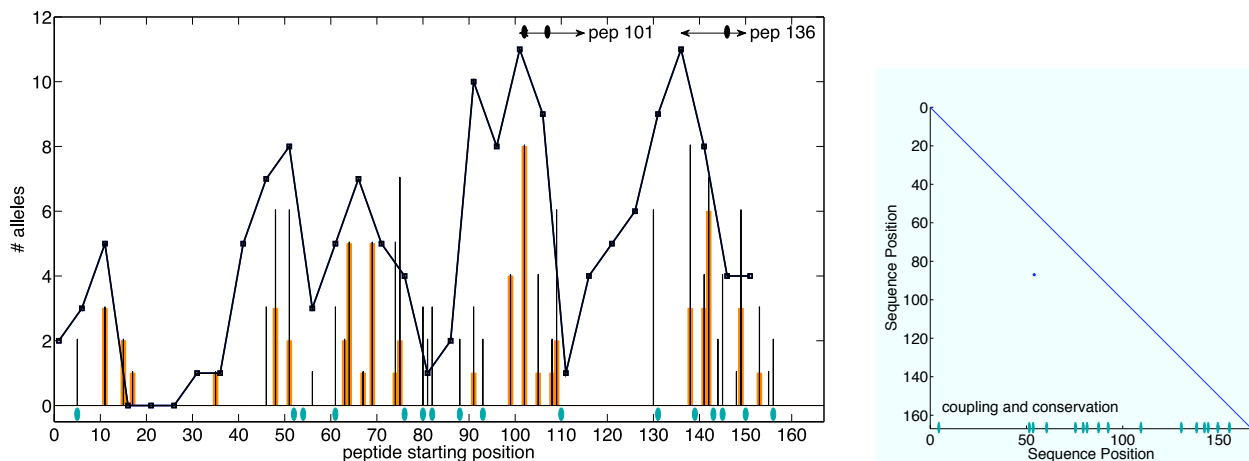


Fig. 7. Optimized 16-substitution, $\alpha = 0.25$ Epo variant. **(left)** Epitope profile. x -axis: starting position of each 9-mer; y -axis: predicted number of alleles recognizing the 9-mer. Thin black bars indicate wild-type scores and thick orange bars indicate variant scores. Note: wild-type epitope scores are always greater than or equal to corresponding variant ones; i.e., we never introduce new epitopes. Blue ellipses indicate mutated positions. Black ellipses at top indicate Tangri et al. mutated positions. The line plot, from Tangri et al., displays wild-type Epo antigenicity using ELISPOT assays, with black squares giving the number of alleles bound to overlapping 15-mers; it trends very well with the ProPred-predicted epitopes. **(right)** Coupling violation (only one).

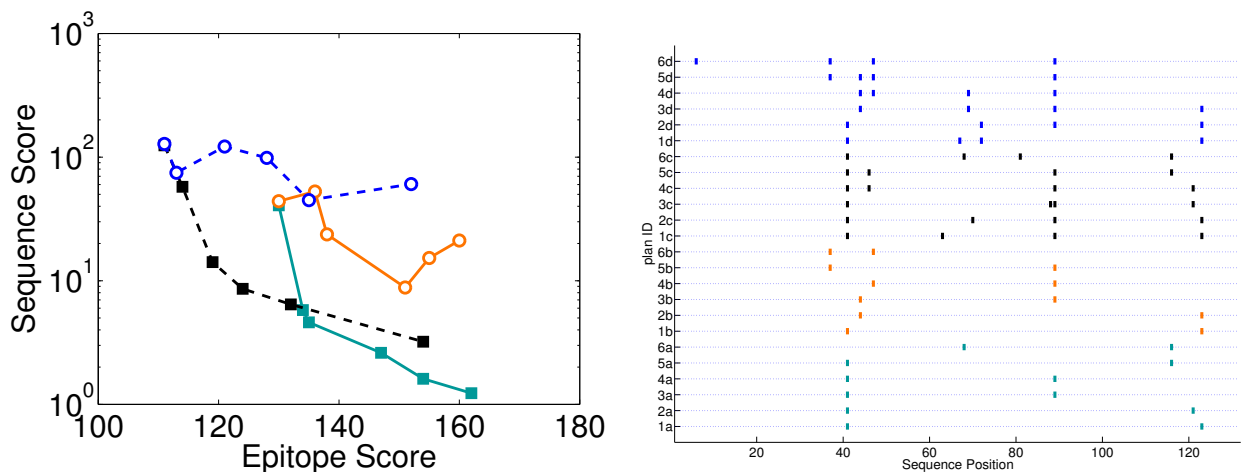


Fig. 8. FVIII plans optimized for varying tradeoffs between epitope and sequence scores. **(left)** Scores. The weight α increases monotonically but not uniformly from left to right. Cyan square solid (a): 2 mutations, conservation and coupling; Orange circle solid (b): 2 mutations, conservation only; Black square dashed (c): 4 mutations, conservation and coupling; Blue circle dashed (d): 4 mutations, conservation only. **(right)** Sequence positions of the substitutions.

designed an 8-substitution FVIII plan with $\alpha = 0.25$ (Fig. 9). The plan eliminates 61 of 162 ProPred predicted pitopes and 25 of 74 SMM-align predicted epitopes while largely respecting pairwise constraints. Jones et al. focus on positions 63-77. We make one mutation in that area as well as many more in flanking areas with equivalently high epitope scores but lower coupling penalties. Our computational method

quickly (< 1 hour) produced a design targeting immunodominant regions of a highly immunogenic and tightly coupled sequence.

As described in Methods, our system optimized all of the plans in this paper with respect to the ProPred immunogenicity model. Another, quite different, technique to predict CD4+ T-cell epitopes works by learning a model from a large database of

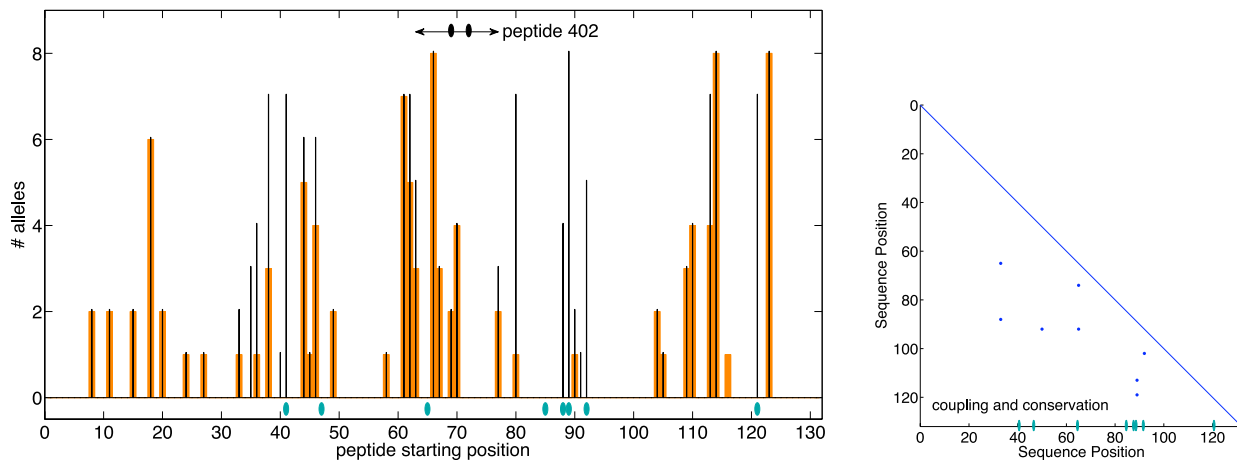


Fig. 9. Optimized 8-substitution, $\alpha = 0.25$ FVIII variant. **(left)** Epitope profile. x -axis: starting position of each 9-mer; y -axis: predicted number of alleles recognizing the 9-mer. Thin black bars indicate wild-type scores and thick orange bars indicate variant scores. Note: wild-type epitope scores are always greater than or equal to corresponding variant ones; i.e., we never introduce new epitopes. Blue ellipses indicate mutated positions. Black ellipses at top indicate Jones et al. mutated positions. **(right)** Coupling violations.

peptides with known MHC-II allele binding affinities. One such successful method is called SMM-align¹³. Tab. 1 shows the 2-mutation plans utilizing conservation and coupling from Fig. 8(a). After optimization, we evaluated these plans with SMM-align. Notice that the number of epitopes predicted by ProPred correlates well with the number of epitopes predicted by SMM-align across the range of α values. We found similar trends in the other test cases.

Table 1. 2-mutation plans using conservation and coupling for FVIII; E : total number of ProPred epitopes; C : conservation penalty for substitutions based on the frequency of a residue in the Family sequence record; S : number of SMM-align epitopes.

α	FVIII 1-131	E	C	S
wild type		162	0.00	74
0.00	L41E, I123E	130	40.75	62
0.25	L41E, Y121A	134	5.79	63
0.85	L41E, M89K	135	4.61	68
0.87	L41Q, M89K	147	2.61	69
0.90	L41Q, L116I	154	1.61	67
1.00	I68V, L116I	162	1.23	74

4. Discussion and Conclusion

We have developed a novel method for the design of biotherapeutics optimized to delete MHC-II epitopes while satisfying sequence constraints on individual

residues and pairs of residues. The combined epitope and sequence-constraint potential enhances the likelihood that designed variants will maintain wild-type stability and therapeutic activity, while displaying reduced immunogenicity. The global optimization of these factors enables us to more confidently pursue plans with higher mutational loads, deleting more epitopes. While the computational problem is intractable in general, we have developed an integer programming approach that is sufficiently fast in practice (plans shown here took at most a few hours of wall-clock time), while guaranteeing optimality.

Our computational system comprises an algorithm and models of immunogenicity and stability; we employ models that have previously been experimentally validated. Significantly, the algorithm is guaranteed to produce results that are optimal with respect to the models. Thus if designed variants prove not to be sufficiently stable, active, or deimmunized, the fault lies in the models, not in the lack of search through the design space. This further provides an opportunity to update and refine the models accordingly. It also enables our system to take advantage of other published improvements in models of immunogenicity and stability.

Previous efforts focused on short fragments of target proteins. In contrast, for each target our method globally searched the entire protein and

chese sets of substitutions spanning several immunodominant regions. The selected mutations obtain the most epitope score reduction for the least sequence score penalty. Quantitatively comparing the results of our computationally determined plans with those in the literature is not appropriate. On the one hand, our plans are guaranteed optimal with respect to the scoring models; the literature plans would have worse scores. On the other hand, experimentally constructing and assaying our plans is beyond the scope of the present paper, which is about a new computational method that integrates well-characterized sequence potentials to uncover promising candidates in a large, complex design space. Ultimately, our method provides the engineer with an effective mechanism for exploring the space of good variants, trading off goals of deimmunization and maintenance of stability and activity.

Acknowledgements. This work is supported in part by US NSF grant CCF-0915388 to CBK, and the William H. Neukom 1964 Institute for Computational Science at Dartmouth College.

References

1. A. S. De Groot, P. M. Knopp, and W. Martin. Deimmunization of therapeutic proteins by T-cell epitope modification. *Dev. Biol. (Basel)*, 122:171–94, 2005.
2. A. S. De Groot and W. Martin. Reducing risk, improving outcomes: Bioengineering less immunogenic protein therapeutics. *Clinical Immunology*, 131:189–201, 2009.
3. A. S. De Groot and L. Moise. Prediction of immunogenicity for therapeutic proteins: State of the art. *Curr. Opin. Drug Discov. Devel.*, 10:332–340, 2007.
4. R. R. Dinglasan, D. E. Kalume, S. M. Kanzok, A. K. Ghosh, O. Muratova, A. Pandey, and M. Jacobs-Lorena. Disruption of Plasmodium falciparum development by antibodies against a conserved mosquito midgut antigen. *PNAS*, 104:13461–13466, 2007.
5. R. Guerois, J. E. Nielsen, and L. Serrano. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, 320:369–387, 2002.
6. P. T. Jones, P. H. Dear, J. Foote, M. S. Neuberger, and G. Winter. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*, 321:522–525, 1986.
7. T. D. Jones, W. J. Phillips, B. J. Smith, C. A. Bamford, P. D. Nayee, T. P. Baglin, J. S. H. Gaston, and M. P. Baker. Identification and removal of a promiscuous CD4+ T cell epitope from the C1 domain of factor VIII. *J. Thromb. Haemost.*, 3:991–1000, 2005.
8. S. V. S. Kashmiri, R. De Pascalis, N. R. Gonzales, and J. Schlom. SDR grafting—a new approach to antibody humanization. *Methods*, 36:25–34, 2005.
9. E. N. Klyushnenkova, D. V. Kouiyavskaia, J. A. Kodak, A. A. Vandembark, and R. B. Alexander. Identification of HLA-DRB1*1501-restricted T-cell epitopes from human prostatic acid phosphatase. *Prostate*, 67:1019–1028, 2007.
10. E. Koren, A. S. De Groot, V. Jawa, K. D. Beck, T. Boone, D. Rivera, L. Li, D. Mytych, M. Koscec, D. Weeraratne, S. Swanson, and W. Martin. Clinical validation of the “in silico” prediction of immunogenicity of a human recombinant therapeutic protein. *Clinical Immunology*, 124:26–32, 2007.
11. S. L. Morrison, M. J. Johnson, L. A. Herzenberg, and V. T. Oi. Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains. *PNAS*, 81:6851–5, 1984.
12. A. S. Mustafa and F. A. Shaban. ProPred analysis and experimental evaluation of promiscuous T-cell epitopes of three major secreted antigens of Mycobacterium tuberculosis. *Tuberculosis*, 86:115–124, 2006.
13. M. Nielsen, C. Lundegaard, and O. Lund. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinf.*, 8:238, 2007.
14. A.S. Parker, W. Zheng, K.E. Griswold, and C. Bailey-Kellogg. Optimization algorithms for functional deimmunization of therapeutic proteins. *BMC Bioinf.*, 11:180, 2010.
15. N. Pierce and E. Winfree. Protein design is NP-hard. *Protein Eng.*, 15:779–782, 2002.
16. W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffee, and R. Ranganathan. Natural-like function in artificial WW domains. *Nature*, 437:579–583, 2005.
17. H. Singh and G.P.S. Raghava. ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, 17:1236–1237, 2001.
18. M. Socolich, S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner, and R. Ranganathan. Evolutionary information for specifying a protein fold. *Nature*, 437:512–518, 2005.
19. S. Southwood, J. Sidney, A. Kondo, M. F. del Guercio, E. Appella, S. Hoffman, R. T. Kubo, R. W. Chesnut, H. M. Grey, and A. Sette. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J. Immunol.*, 160:3363–3373, 1998.
20. T. Sturniolo, E. Bono, J. Ding, L. Radrizzani, O. Tuereci, U. Sahin, M. Braxenthaler, F. Gallazzi, M. P. Protti, F. Sinigaglia, and J. Hammer. Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnol.*, 17:555–561, 1999.

21. S. Tangri, B. R. Mothe, J. Eisenbraun, J. Sidney, S. Southwood, K. Briggs, J. Zinckgraf, P. Bilsel, M. Newman, R. Chesnut, C. LiCalsi, and A. Sette. Rationally engineered therapeutic proteins with reduced immunogenicity. *J. Immunol.*, 174:3187–3196, 2005.
22. J. Thomas, N. Ramakrishnan, and C. Bailey-Kellogg. Graphical models of residue coupling in protein families. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 5:183–197, 2008.
23. J. Thomas, N. Ramakrishnan, and C. Bailey-Kellogg. Graphical models of protein-protein interaction specificity from correlated mutations and interaction data. *Proteins*, 76:911–929, 2009.
24. J. Thomas, N. Ramakrishnan, and C. Bailey-Kellogg. Protein design by sampling an undirected graphical model of residue constraints. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 6:506–516, 2009.
25. C. A. Voigt, C. Martinez, Z. G. Wang, S. L. Mayo, and F. H. Arnold. Protein building blocks preserved by recombination. *Nat. Struct. Biol.*, 9:553–558, 2002.
26. P. Wang, J. Sidney, C. Dow, B. Mothe, A. Sette, and B. Peters. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comp. Biol.*, 4:e1000048, 2008.
27. P. A. M. Warmerdam, S. Plaisance, K. Vanderlick, P. Vandervoort, K. Brepoels, D. Collen, and M. De Maeyer. Elimination of a human T-cell region in staphylokinase by T-cell screening and computer modeling. *J. Thromb. Haemost.*, 87:666–673, 2002.
28. X. Ye, A. M. Friedman, and C. Bailey-Kellogg. Hypergraph model of multi-residue interactions in proteins: sequentially-constrained partitioning algorithms for optimization of site-directed protein recombination. *J. Comput. Biol.*, 14:777–790, 2007.

