

# Neue Ansätze zum computergestützten Entwurf epitopbasierter Impfstoffe

Nora C. Toussaint

Immunology Program & Computational Biology Program  
Lucille Castori Center for Microbes, Inflammation & Cancer  
Sloan-Kettering Institute  
Memorial Sloan-Kettering Cancer Center  
1275 York Ave, New York, NY 10065, U.S.A.  
toussain@mskcc.org

**Abstract:** Trotz zahlreicher Erfolge des traditionellen Impfstoffentwurfs, gibt es immer noch Krankheiten, gegen die bisher kein geeigneter Impfstoff entwickelt werden konnte. Die wohl bekanntesten Beispiele sind HIV-Infektion und Krebs. Hier sind neue, rational entworfene Impfstoffe, wie zum Beispiel epitopbasierte Impfstoffe, eine vielversprechende Alternative. Der gezielte Einsatz computergestützter Methoden im epitopbasierten Impfstoffentwurf verspricht darüber hinaus verbesserte Impfstoffe bei kürzerer Entwicklungszeit und geringeren Kosten.

Im Rahmen der hier zusammengefassten Dissertation wurden relevante Probleme des epitopbasierten Impfstoffentwurfs formalisiert und mit Hilfe von Methoden der kombinatorischen Optimierung und des maschinellen Lernens gelöst. Die Anwendung der vorgestellten Methoden in realistischen Impfstoffentwurfstudien lieferte vielversprechende Ergebnisse, die das große Potenzial des computergestützten rationalen Impfstoffentwurfs verdeutlichen.

## 1 Einführung

Die Entwicklung von Impfstoffen gehört zu den bedeutendsten Fortschritten in der Geschichte der modernen Medizin. Die Grundidee von Impfungen ist es, Immunität zu generieren, ohne die eigentliche Krankheit hervorzurufen. Dazu machen sich Impfstoffe das Gedächtnis des adaptiven Teils des Immunsystems zu Nutze. Ziel der Immunoinformatik ist es, über die Modellierung des Immunsystems zu einem besseren Verständnis immunologischer Prozesse beizutragen. Mit Hilfe dieser Modelle soll eine gezielte Veränderung des immunologischen Gedächtnisses und damit eine optimale Behandlung möglich werden.

Das adaptive Immunsystem wird aktiv, wenn es ein immunogenes Proteinfsegment erkennt. Im Inneren einer Wirtszelle werden die Proteine eines Pathogens – ebenso wie alle anderen Proteine – in kleine Proteinfsegmente, die *Peptide*, zerlegt. Einige dieser Peptide binden an Moleküle des Haupthistokompatibilitätskomplexes (MHC - *major histocompatibility complex*), die sie an der Zelloberfläche präsentieren. Erkennt eine T-Zelle einen

solchen MHC-Peptid-Komplex wird eine Immunantwort hervorgerufen (Abbildung 1). Peptide, die eine Immunantwort hervorrufen können, werden *immunogene Peptide* oder auch *Epitope* genannt. Die Proteine, aus denen sie hervorgegangen sind, heißen *Antigene*.

Epitope stellen die kleinsten immunogenen Teile eines Antigens dar. Die Verwendung von Epitopen als Komponenten eines Impfstoffs bietet viele Vorteile. Epitopbasierte Impfstoffe (EBIs) können gezielt hochimmunogene Regionen von targetspezifischen Antigenen ins Visier nehmen. Darüber hinaus können EBIs auf die MHC-Ausprägungen einer spezifischen Person zugeschnitten werden und sind somit im Bereich der personalisierten Medizin einsetzbar. EBIs haben noch viele weitere gute Eigenschaften, die sie für die Pharmaindustrie besonders interessant machen. Es ist daher kaum verwunderlich, dass sie in den letzten Jahren sehr viel Aufmerksamkeit erweckt haben.

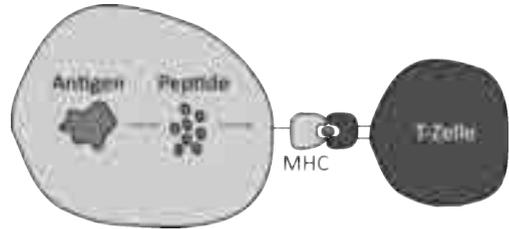


Abbildung 1: Im Inneren der Wirtszelle wird ein Antigen in kleinere Stücke, die Peptide, zerlegt. MHC-Moleküle binden solche Peptide und präsentieren sie an der Zelloberfläche. Wird ein MHC-Peptid-Komplex von einer T-Zelle erkannt, wird eine Immunantwort hervorgerufen.

Der Entwurf eines EBIs kann grob in drei Schritte unterteilt werden: Epitopbestimmung, Epitopselektion und Epitopassemblierung (Abbildung 2). In jedem dieser Schritte können computergestützte Methoden eingesetzt werden, um die Arbeit von Immunologen zu vereinfachen und sie in ihren Entscheidungen zu unterstützen. Die Anwendung solcher *in-silico*-Methoden im EBI-Entwurf verspricht verbesserte Impfstoffe bei einer kürzeren Entwicklungszeit und geringeren Kosten.

Das noch recht junge interdisziplinäre Feld der Immunoinformatik ist geprägt von proprietären Daten – nicht zuletzt aufgrund der Bedeutung immunologischer Daten für die pharmazeutische Industrie – und von einer Grundskepsis der Immunologen gegenüber Versuchen, das Immunsystem durch Gleichungen zu beschreiben. Die hier zusammengefasste Dissertation [Tou11] versucht zur Etablierung informatischer Methoden in der



Abbildung 2: Epitopbasierter Impfstoffentwurf. Ausgehend von einer Menge von targetspezifischen Antigenen werden Kandidatenepitope bezüglich einer Zielpopulation bestimmt (*Epitopbestimmung*). Aus der resultierenden Menge von Kandidatenepitopen muss die für die Verwendung in einem EBI am besten geeignete Teilmenge ausgewählt werden (*Epitopselektion*). Die ausgewählten Epitope werden zum EBI zusammengesetzt (*Epitopassemblierung*). (Die Abbildung basiert auf [TK09b].)

Immunologie beizutragen, indem sie immunologische Probleme formalisiert und mit theoretisch fundierten Methoden reproduzierbar und, soweit möglich, optimal löst. Um neue und verbesserte *in-silico*-Methoden zur Epitopbestimmung zu ermöglichen, wurden Methoden des maschinellen Lernens verwendet und weiterentwickelt. Darüber hinaus wurden die Probleme der Epitopselektion und -assemblierung formalisiert und Methoden der kombinatorischen Optimierung verwendet, um diese Probleme erstmalig optimal zu lösen.

## 2 Epitopbestimmung

Ausgehend von einer Menge von Antigenen werden im ersten Schritt des EBI-Entwurfs Kandidatenepitope bestimmt und experimentell validiert. Kandidatenepitope sind diejenigen targetspezifischen Peptide, die eine adaptive Immunantwort in der Zielpopulation hervorrufen können. Die Einbindung von *in-silico*-Methoden in diesen Schritt kann die Anzahl der durchzuführenden Experimente drastisch reduzieren.

Aufgrund der komplexen Abhängigkeit der T-Zell-Reaktivität vom Immunsystem des Wirts sowie des unvollständigen Verständnisses der zugrundeliegenden Prozesse ist die Vorhersage von T-Zell-Epitopen ein äußerst schwieriges Problem. Da T-Zellen immunogene Peptide nur erkennen können, wenn diese von MHC-Molekülen präsentiert werden, ist die Bindung von Peptiden an MHC-Moleküle notwendig – allerdings nicht hinreichend – für die Auslösung einer Immunantwort. Die Prozesse, die für die Bindung eines Peptids an ein MHC-Molekül zuständig sind, sind gut verstanden. Da gezeigt wurde, dass das Potenzial eines Peptids, eine T-Zell-basierte Immunantwort hervorzurufen, also die *Immunogenität* eines Peptids, gut mit der MHC-Bindeaffinität des Peptids korreliert, wird das Epitopbestimmungsproblem üblicherweise auf das Problem der MHC-Bindevorhersage reduziert.

In der hier zusammengefassten Dissertation werden drei Ansätze zur Epitopvorhersage vorgestellt: zwei Ansätze zur MHC-Bindevorhersage und ein Ansatz zur Verbesserung der Vorhersage von Immunogenität.

### 2.1 MHC-Bindevorhersage

Das Hauptproblem bei der Vorhersage von Peptiden, die an ein bestimmtes MHC-Molekül binden, ist der Mangel an (frei) verfügbaren experimentellen Bindedaten. Mehrere tausend genetische MHC-Varianten, die sogenannten *Allele*, sind bekannt. Unterschiedliche MHC-Varianten binden unterschiedliche Peptidrepertoires und unterschiedliche Individuen tragen unterschiedliche MHC-Varianten. Daraus folgt, dass in jedem Menschen unterschiedliche Peptide von MHC-Molekülen präsentiert werden. Da T-Zellen Epitope nur im Komplex mit MHC-Molekülen erkennen, kann ein Peptid, das in einem Individuum eine T-Zell-Antwort hervorruft, in einem anderen Individuum vom Immunsystem unbeachtet bleiben. Darüber hinaus sind bestimmte MHC-Allele in einer Population häufiger als andere und die Verteilung von MHC-Allelen variiert zwischen unterschiedlichen Populationen. Die Berücksichtigung der MHC-Verteilung in der Zielpopulation ist für den Entwurf eines

effektiven EBIs daher unumgänglich.

Klassische Ansätze zur MHC-Bindevorhersage basieren auf allelspezifischen Modellen und benötigen eine gewisse Menge an experimentellen Bindedaten für das jeweilige Allel [PBC94, BLW<sup>+</sup>03]. Für einen Großteil der bekannten MHC-Allele sind jedoch nicht genügend Bindedaten verfügbar. Im Rahmen der hier zusammengefassten Dissertation werden zwei Ansätze zur Überwindung dieses Problems vorgestellt.

### Neue Kernfunktionen

Ein Ansatz zur Umgehung des Problems der mangelnden Bindedaten für ein bestimmtes MHC-Allel ist die Einbindung zusätzlichen biologischen Wissens. MHC-Moleküle binden Peptide in gestreckter Form (Abbildung 3). Innerhalb des Komplexes interagieren die Peptidbestandteile, d.h. die Aminosäuren, sowohl untereinander als auch mit den angrenzenden Aminosäuren des MHC-Moleküls. Jede der Aminosäuren des Peptids trägt somit zur Bindeaffinität zwischen Peptid und MHC-Molekül bei. Der jeweilige Beitrag hängt von den physikochemischen Eigenschaften der Aminosäure (z.B. Ladung, Größe, Hydrophobizität), von den physikochemischen Eigenschaften der Aminosäuren in der näheren Umgebung und auch von der Position der Aminosäure innerhalb des Peptids ab.

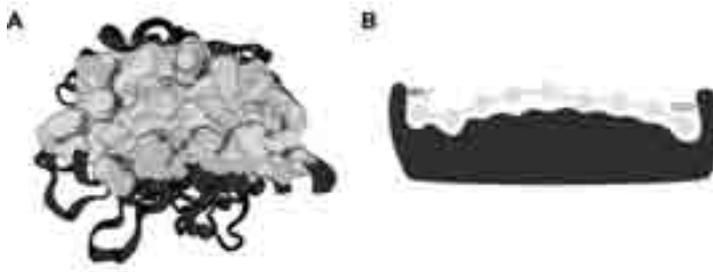


Abbildung 3: MHC-Peptid-Bindung. A) Kristallstruktur eines MHC-Peptid-Komplexes. B) Schematische Darstellung der gestreckten Bindung eines Peptids (grün) an ein MHC-Molekül (rot).

Das Wissen über physikochemische Eigenschaften von Aminosäuren wurde erfolgreich verwendet, um die MHC-Bindevorhersage zu verbessern (z.B. [NLW<sup>+</sup>03]). Die Anordnung der Aminosäuren im Peptid wurde jedoch nicht berücksichtigt. Die hier zusammengefasste Dissertation stellt einen Ansatz basierend auf Supportvektormaschinen (SVMs) [Vap95] vor, der erstmalig das Wissen über physikochemische Aminosäureeigenschaften mit Informationen über die Anordnung der Aminosäuren im Peptid kombiniert [TWKR10, WTA<sup>+</sup>10].

Ausgehend von einer Menge von  $N$  Trainingsbeispielen – in unserem Fall Peptide  $\mathbf{x}_i$  mit ihrem Label  $y_i \in \{\pm 1\}$  – lernen SVMs eine Entscheidungsfunktion, mit der Vorhersagen über ungesehene Peptide getroffen werden können:

$$f(\mathbf{x}) = \text{sgn} \left( \sum_{i=1}^N \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right).$$

Die Trainingsbeispielgewichte  $\{\alpha_i\}_{i=1}^N$  und der Bias  $b$  werden durch Lösen eines quadratischen Optimierungsproblems bestimmt. Die Funktion  $k(\mathbf{x}, \mathbf{x}')$  wird als *Kern* bezeichnet. Sie stellt ein Ähnlichkeitsmaß zwischen zwei Peptidsequenzen dar und hat einen wesentlichen Einfluss auf die Vorhersagequalität der SVM.

Ein Kern, der sich besonders gut eignet, die gestreckte Anordnung der Peptide im MHC-Peptid-Komplex zu berücksichtigen, ist der sogenannte *Weighted-Degree-Kern* (WD-Kern) [RS04]. Der WD-Kern vergleicht zwei Sequenzen gleicher Länge  $L$  basierend auf den Substrings, aus denen sie sich zusammensetzen. Der WD-Kern von Grad  $d$  ist folgendermaßen definiert:

$$k_d^{wd}(\mathbf{x}, \mathbf{x}') = \sum_{\ell=1}^d \beta_{\ell} \sum_{i=1}^{L-\ell-1} \mathbf{I}(\mathbf{x}_{[i:i+\ell]}, \mathbf{x}'_{[i:i+\ell]}),$$

wobei  $I$  die Identitätsfunktion ist,  $\mathbf{x}_{[i:i+\ell]}$  der Substring der Länge  $\ell$  an Position  $i$  und  $\beta_{\ell}$  ein Gewicht, das Übereinstimmungen in längeren Substrings niedriger gewichtet. Ein wesentlicher Nachteil des WD-Kerns bei der Vorhersage von MHC-Bindung ist jedoch, dass für ihn alle Aminosäuren gleich ähnlich sind. Eine Berücksichtigung physikochemischer Eigenschaften von Aminosäuren ist nicht ohne Weiteres möglich.

In der Dissertation wird eine Gruppe von Kernfunktionen vorgestellt, die die Vorteile des WD-Kerns mit denen von physikochemischen Deskriptoren für Aminosäuren verbindet. Ausführliche empirische Untersuchungen im Rahmen der Dissertation haben gezeigt, dass sich der *WD-RBF-Kern* besonders gut für die MHC-Bindevorhersage eignet. Dieser Kern ist wie folgt definiert

$$k_{d,\sigma}^{wd,\Psi}(\mathbf{x}, \mathbf{x}') = \sum_{\ell=1}^d \beta_{\ell} \sum_{i=1}^{L-\ell+1} \exp\left(-\frac{\sum_{j=1}^{\ell} \|\Psi(\mathbf{x}_j) - \Psi(\mathbf{x}'_j)\|^2}{2\sigma^2}\right),$$

wobei  $\mathbf{x}_j$  die Aminosäure an Position  $j$  im Peptid  $\mathbf{x}$  ist. Die Abbildung  $\Psi : \Sigma \rightarrow \mathbb{R}^m$  bildet eine Aminosäure  $x \in \Sigma$  auf  $m$  physikochemische Deskriptoren ab. Effiziente Implementierungen der in der Dissertation vorgestellten Kerne sind als Teil der Open-Source-Machine-Learning-Toolbox Shogun (<http://www.shogun-toolbox.org>) [SRH<sup>+</sup>10] öffentlich verfügbar gemacht worden.

Ein statistischer Vergleich mit dem WD-Kern auf 35 unterschiedlichen MHC-Allelen zeigt, dass die Verwendung des WD-RBF-Kerns zu einer Verbesserung der Vorhersagequalität führt: der WD-RBF-Kern liefert in 23 Fällen bessere Ergebnisse als der WD-Kern, der nur in acht Fällen die besseren Vorhersageergebnisse liefert. Die Verbesserung durch den WD-RBF-Kern ist besonders deutlich, wenn nur wenig Trainingsdaten zur Verfügung stehen.

## Neue Ansätze des Transferlernens

Der zweite in der Dissertation vorgestellte Ansatz zur MHC-Bindevorhersage kommt aus dem Bereich des sogenannten Transferlernens und nutzt zur Umgehung des Problems der mangelnden Daten strukturelle Ähnlichkeiten zwischen unterschiedlichen MHC-Varianten aus. Diese strukturellen Ähnlichkeiten erlauben es, experimentelle Bindedaten einer MHC-Variante für die Vorhersage der Bindeeigenschaften einer anderen zu verwenden.

Der Unterschied zwischen zwei MHC-Varianten zeigt sich in der Bindetasche, also in dem Bereich, in dem das MHC-Molekül das Peptid bindet. Ausgehend von einer Menge von 3D-Strukturen von MHC-Peptid-Komplexen konnten wir feststellen, welche Positionen in der Sequenz eines MHC-Moleküls mit dem gebundenen Peptid interagieren. Die Aminosäuren an den jeweiligen Positionen bilden das Profil der Bindetasche – eine Art Fingerabdruck, anhand dessen MHC-Varianten unterschieden werden können. Je ähnlicher das Profil zweier MHC-Varianten, umso ähnlicher sind auch die Peptide, die sie binden.

Während klassische Ansätze zur MHC-Bindevorhersage ein Modell pro MHC-Variante trainieren, trainieren wir ein einziges SVM-Modell für alle MHC-Varianten gemeinsam. Als Eingabe dient ein MHC-Peptid-Paar, repräsentiert durch einen Vektor, der sich aus physikochemischen Deskriptoren für das Profil der Bindetasche der MHC-Variante und für die Aminosäuren des Peptids zusammensetzt. Dieser Ansatz ermöglichte erstmalig Bindevorhersagen für alle bekannten MHC-Varianten, unabhängig von der Menge der für das jeweilige Allel zur Verfügung stehenden Bindedaten. Die Vorhersagegenauigkeit der Methode ist sowohl für Allele mit Bindedaten als auch für Allele ohne Bindedaten vergleichbar oder sogar besser als die von allelspezifischen Methoden.

## 2.2 Immunogenitätsvorhersage

Die Bindung eines Peptids an ein MHC-Molekül ist lediglich eine notwendige Voraussetzung für die Induktion einer T-Zell-Antwort, d.h. für die Immunogenität eines Peptids. Die Vorhersage von MHC-bindenden Peptiden löst das Epitopbestimmungsproblem im EBI-Entwurf demnach nicht vollständig.

Das Hauptproblem bei der Vorhersage von immunogenen Peptiden ist die komplexe Abhängigkeit der T-Zell-Reaktivität vom Immunsystem des Patienten. Bereits existierende Methoden zur Vorhersage von Immunogenität [BR04, TH07] berücksichtigen diese Abhängigkeit nicht, sondern beziehen lediglich die Aminosäuresequenz des Peptids in die Vorhersage mit ein. Diese sehr starke Vereinfachung des Problems spiegelt sich in einer geringen Vorhersagegenauigkeit wieder.

In der Dissertation wird ein Ansatz vorgestellt, der erstmalig zusätzlich zur Peptidsequenz Wissen über eine relevante Eigenschaft des Immunsystems einbezieht: das adaptive Immunsystem ist *selbsttolerant*, d.h. körpereigene Proteine und Peptide rufen im Allgemeinen keine Immunantwort hervor. Daraus lässt sich folgern, dass körperfremde Peptide, die eine hohe Ähnlichkeit zu Selbstpeptiden haben, wahrscheinlich keine T-Zell-gesteuerte Immunantwort hervorrufen.

Unter Verwendung einer effizienten Trie-Struktur und einer geeigneten Ähnlichkeitsfunktion ermittelt unser Ansatz für jedes betrachtete Peptid die Ähnlichkeit zu den Peptiden des menschlichen Proteoms. Einbeziehung dieser Information in die Immunogenitätsvorhersage führt zu einer deutlichen Verbesserung der Vorhersagegenauigkeit [TFZ<sup>+</sup>11].

### 3 Epitopselektion

Nachdem im ersten Schritt des EBI-Entwurfs eine Menge von Kandidatenepitopen bestimmt wurde, wird im zweiten Schritt aus dieser Menge die Teilmenge ausgewählt, die die beste Immunantwort in der Zielpopulation verspricht. Aufgrund von regulatorischen, ökonomischen und praktischen Überlegungen kann nur eine kleine Menge der im vorhergehenden Schritt bestimmten Kandidatenepitope in den EBI einbezogen werden. Da der Erfolg des Impfstoffs von den ausgewählten Epitopen abhängt, ist es äußerst wichtig, die optimale Kombination von Peptiden zu identifizieren, also die Menge von Peptiden, die die bestmögliche Immunantwort in der Zielpopulation hervorruft.

Weder die übliche manuelle Selektion durch Experten noch die bisher publizierten computergestützten Methoden [GMB<sup>+</sup>05, VSRL07] können garantieren, dass die getroffene Auswahl und somit der resultierende Impfstoff optimal sind. Wir stellen einen auf ganzzahliger linearer Programmierung (ILP - *integer linear programming*) basierenden Ansatz vor, der eine beweisbar optimale Lösung liefert. Unser Ansatz erlaubt eine elegante und flexible Formulierung unterschiedlicher Anforderungen an die auszuwählenden Epitope. Für typische Problemgrößen beträgt die Laufzeit nur wenige Sekunden [TDK08].

Ein guter Impfstoff zeigt eine hohe Gesamtimmunogenität, d.h. er hat das Potenzial starke Immunität in einem Großteil der Zielpopulation hervorzurufen. Die Zielfunktion maximiert also die Gesamtimmunogenität der ausgewählten Epitope. Zusätzlich werden je nach Fall unterschiedliche Anforderungen an die auszuwählenden Epitope gestellt. Diese Anforderungen werden im ILP als Bedingungen formuliert. Ein Beispiel ist in ILP 1 gegeben. Das ILP wählt  $k$  Epitope aus (1a) – aus jedem der  $n$  Antigene mindestens eins (1b) – die insgesamt im Kontext von mindestens  $m$  MHC-Varianten immunogen sind (1c, 1d).

---

#### ILP 1 Ausschnitt aus einem ILP zur Epitopselektion

---

$$\begin{aligned} \text{Maximiere} \quad & \sum_{e \in E} x_e \sum_{a \in A} p(a) i(e, a), \\ \text{so dass} \quad & \sum_{e \in E} x_e = k \quad (1a) \\ & \forall i \in \{1, \dots, n\} : \sum_{e \in E_i \cap I} x_e \geq 1 \quad (1b) \\ & \forall a \in A : \sum_{e \in I_a} x_e \geq y_a \quad (1c) \\ & \sum_{a \in A} y_a \geq m \quad (1d) \end{aligned}$$

**DEFINITIONEN**

- $A$  Menge der MHC-Allele der Zielpopulation
- $E_i$  Menge der Kandidatenepitope von Antigen  $i$
- $E$  Menge aller Kandidatenepitope ( $E = E_1 \cup \dots \cup E_n$ )
- $I_a$  Menge von Epitopen, die im Kontext von MHC-Variante  $a$  immunogen sind
- $I$  Menge aller bezüglich  $A$  immunogenen Epitope ( $I = \bigcup_{a \in A} I_a$ )

**PARAMETER**

- $i(e, a)$  Immunogenität von Epitop  $e$  im Kontext von MHC-Variante  $a$
- $k$  Anzahl der auszuwählenden Epitope
- $p(a)$  Wahrscheinlichkeit, dass die MHC-Variante  $a$  in der Zielpopulation vorkommt
- $m$  Mindestanzahl von MHC-Varianten, die berücksichtigt werden sollen

**VARIABLEN**

- $x_e = 1$  falls Epitop  $e$  zur optimalen Menge gehört, sonst  $x_e = 0$
  - $y_a = 1$  falls MHC-Variante  $a$  von der Epitopmenge berücksichtigt wird, sonst  $y_a = 0$
-

Die Formalisierung des Epitopselektionsproblems und die Formulierung als ILP führen zu deutlich besseren Ergebnissen und damit – zumindest theoretisch – zu deutlich besseren Impfstoffen als bereits existierende Lösungsansätze (Abbildung 4). Um diese Methoden Immunologen zugänglich zu machen, wurde ein öffentlich verfügbarer Webservice entwickelt (<http://www.epitoolkit.org/optitope>) [TK09a].

## 4 Epitopassemblierung

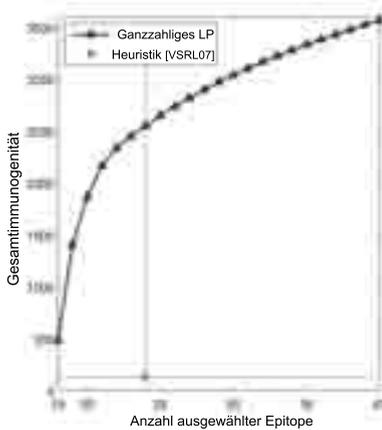


Abbildung 4: Epitopselektion. Vergleich des hier vorgestellten optimalen Ansatz mit der in [VSRL07] präsentierten Heuristik.

toppaares. Das Gewicht  $w_{ab}$  der Kante  $(a, b)$  entspricht dem Logarithmus der Wahrscheinlichkeit, dass die Epitope  $a$  und  $b$  im Konstrukt  $a - b$  abgebaut werden. In dieser Formulierung entspricht die optimale Epitopanordnung dem minimalen Hamiltonpfad, d.h. dem Pfad, der jeden Knoten genau einmal besucht und minimales Gewicht hat. Durch Hinzufügen eines weiteren Knotens, der mit allen anderen Knoten über jeweils eine Hin- und eine Rückkante verbunden wird, wird das Problem in die Suche nach dem minimalen Hamiltonkreis umgewandelt. Die Bestimmung des minimalen Hamiltonkreises entspricht dem wohlbekannten und ausgiebig studierten *Traveling Salesman Problem* (TSP). Verwendung einer bereits publizierten ILP-Formulierung des TSP [Pat03] erlaubt es, das Epitopassemblierungsproblem für realistische Epitopzahlen in vertretbarer Zeit optimal zu lösen.

In einer theoretischen Impfstoffentwurfstudie wurde gezeigt, dass bei einer optimierten Anordnung deutlich mehr der ausgewählten Epitope von MHC-Molekülen an der Zelloberfläche präsentiert werden können als bei einer zufälligen Anordnung [TMKL11].

Der letzte Schritt des EBI-Entwurfs befasst sich mit der Verabreichung des Impfstoffs. Dazu werden die Peptide üblicherweise zu einem einzelnen langen Polypeptid zusammengefügt. Da eine ungünstig gewählte Epitopanordnung zum Abbau der gewünschten Epitope im Wirt führen kann, ist die optimale Anordnung von wesentlicher Bedeutung für den Erfolg des Impfstoffs. Wir präsentieren den ersten computergestützten Ansatz zur Lösung dieses komplexen Problems. Unsere graphentheoretische Formulierung ermöglicht es, für realistische Problemgrößen die optimale Anordnung der Epitope in vertretbarer Zeit zu bestimmen.

Sei  $G = (V, E, w)$  ein vollständiger, gerichteter und gewichteter Graph mit Knoten  $V$ , Kanten  $E$  und Gewichten  $w$ . Jeder Knoten repräsentiert ein Epitop und jede Kante eine mögliche Konkatenation des jeweiligen Epi-

## 5 Zusammenfassung und Ausblick

EBIs haben das Potenzial, eine wichtige Rolle bei der Bekämpfung von Krebs aber auch von neu auftretenden Infektionskrankheiten wie zum Beispiel der Schweinegrippe zu spielen. Beim EBI-Entwurf müssen zahlreiche Probleme gelöst und Entscheidungen getroffen werden. Hier kann die Informatik einen wesentlichen Beitrag leisten.

Die hier zusammengefasste Dissertation [Tou11] befasst sich mit der Formalisierung wesentlicher Probleme des EBI-Entwurfs und der Entwicklung theoretisch fundierter Methoden, um diese Probleme reproduzierbar und optimal zu lösen. Die hier vorgestellten Methoden zur Epitopbestimmung ermöglichen deutlich verbesserte Bindevorhersagen für MHC-Allele mit wenigen oder keinen Bindedaten, sowie eine verbesserte Immunogenitätsvorhersage. Die Verwendung von ILP-basierten Ansätzen zur Lösung des Epitopselektions- und des Epitopassemblierungsproblems liefert erstmalig garantiert optimale EBI-Konstrukte.

Langfristig wird die Verwendung von *in-silico*-Methoden zur Epitopbestimmung, -selektion und -assemblierung den Impfstoffentwurfsprozess drastisch verändern. Die Verwendung standardisierter Ansätze für den Entwurf von EBIs verspricht insbesondere kürzere Entwicklungszeiten, was essenziell sein kann, wenn es zu neu auftretenden Infektionskrankheiten kommt. Darüber hinaus ist eine schnelle und auch kostengünstige Impfstoffentwicklung unverzichtbar für vollständig personalisierte Impfstoffe, insbesondere für die Immuntherapie von Krebs.

## Literatur

- [BLW<sup>+</sup>03] S. Buus, S. L. Lauemøller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm und S. Brunak. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens*, 62(5):378–384, 2003.
- [BR04] M. Bhasin und G. P. S. Raghava. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine*, 22(23-24):3195–3204, 2004.
- [GMB<sup>+</sup>05] A. S. De Groot, L. Marcon, E. A. Bishop, D. Rivera, M. Kutzler, D. B. Weiner und W. Martin. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine*, 23(17-18):2136–2148, 2005.
- [NLW<sup>+</sup>03] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemøller, K. Lamberth, S. Buus, S. Brunak und O. Lund. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*, 12(5):1007–1017, 2003.
- [Pat03] G. Pataki. Teaching integer programming formulations using the traveling salesman problem. *SIAM review*, 45(1):116–123, 2003.
- [PBC94] K. C. Parker, M. A. Bednarek und J. E. Coligan. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol*, 152(1):163–175, 1994.
- [RS04] G. Rätsch und S. Sonnenburg. Accurate Splice Site Detection for *Caenorhabditis elegans*. In B. Schölkopf, K. Tsuda und J.-P. Vert, Hrsg., *Kernel Methods in Computational Biology*, Seiten 277–298. MIT Press, 2004.

- [SRH<sup>+</sup>10] S. Sonnenburg, G. Rätsch, S. Henschel, C. Widmer, J. Behr, A. Zien, F. de Bona, A. Binder, C. Gehl und V. Franc. The SHOGUN Machine Learning Toolbox. *J Mach Learn Res*, 11:1799–1802, 2010.
- [TDK08] N. C. Toussaint, P. Dönnies und O. Kohlbacher. A mathematical framework for the selection of an optimal set of peptides for epitope-based vaccines. *PLoS Comput Biol*, 4(12):e1000246, 2008.
- [TFZ<sup>+</sup>11] N. C. Toussaint, M. Feldhahn, M. Ziehm, S. Stevanović und O. Kohlbacher. T-cell epitope prediction based on self-tolerance. In *Proceedings of the Second Immunoinformatics and Computational Immunology Workshop*, 2011.
- [TH07] C.-W. Tung und S.-Y. Ho. POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. *Bioinformatics*, 23(8):942–949, 2007.
- [TK09a] N. C. Toussaint und O. Kohlbacher. OptiTope – a web server for the selection of an optimal set of peptides for epitope-based vaccines. *Nucleic Acids Res*, 37(Web Server issue):W617–22, 2009.
- [TK09b] N. C. Toussaint und O. Kohlbacher. Towards in silico design of epitope-based vaccines. *Expert Opin Drug Discovery*, 4(10):1047–1060, 2009.
- [TMKL11] N. C. Toussaint, Y. Maman, O. Kohlbacher und Y. Louzoun. Universal peptide vaccines – optimal peptide vaccine design based on viral sequence conservation. *Vaccine*, 29(47):8745–8753, 2011.
- [Tou11] N.C. Toussaint. *New approaches to in silico design of epitope-based vaccines*. Dissertation, Universität Tübingen, 2011.
- [TWKR10] N. C. Toussaint, C. Widmer, O. Kohlbacher und G. Rätsch. Exploiting physicochemical properties in string kernels. *BMC Bioinf*, 11 Suppl 8:S7, 2010.
- [Vap95] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.
- [VSRL07] T. Vider-Shalit, S. Raffaelli und Y. Louzoun. Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Mol Immunol*, 44(6):1253–1261, 2007.
- [WTA<sup>+</sup>10] C. Widmer, N. C. Toussaint, Y. Altun, O. Kohlbacher und G. Rätsch. Novel machine learning methods for MHC class I binding prediction. In T. Dijkstra, E. Tsivtsivadze, E. Marchiori und T. Heskes, Hrsg., *Pattern Recognition in Bioinformatics*, Jgg. 6282 of *Lecture Notes in Computer Science*, Seiten 98–109. Springer Berlin / Heidelberg, 2010.



**Nora C. Toussaint** studierte Diplom-Informatik mit Nebenfach Biologie an der Humboldt-Universität zu Berlin. Bis Ende 2011 war sie als wissenschaftliche Mitarbeiterin an der Universität Tübingen tätig und promovierte im Bereich Immunoinformatik bei Prof. Dr. Oliver Kohlbacher. Während ihrer Promotion hat sie Forschungsaufenthalte am Centrum Wiskunde & Informatica (CWI) in Amsterdam und an der Bar-Ilan Universität in Israel verbracht. Derzeit arbeitet sie als *Postdoc* in den Forschungsabteilungen Immunologie und Bioinformatik des Memorial Sloan-Kettering Cancer Centers in New York.