

Schlussbericht zum Vorhaben

Verbundprojekt: Exploration und Translation modifizierter, glycosilierter Vakzine Epitope; Teilprojekt: Algorithmen zur Identifizierung von Mimotopen durch subtraktives Epitop-Fingerprinting

im Rahmen des Eurostars Projekts
E - SVEET
„Sugar-modified Vaccine Epitopes; Exploration and Translation“

Michael Szardenings, Marcus Puder

epitopic GmbH
Deutscher Platz 5E
04103 Leipzig

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin / beim Autor.

Förderkennzeichen: 01QE2109B
Projektlaufzeit: 01.04.2021-31.03.2024

Teil I Kurzbericht

Im Projekt SVEET sollte untersucht werden, wie sich die bakterielle Glykosylierung eines potenziellen Impfstoff-Antigens auf die Bildung einer schützenden Immunantwort auswirkt. Bei dem Protein handelte es sich um das YghJ Protein, ein Protein auf der Oberfläche von Bakterien, das es diesen ermöglicht, sich an das Epithel von Blase und Harnleitern anzuheften und erst dadurch Harnwegsinfekte zu verursachen.

Herkömmliche rekombinante Proteine aus Bakterien sind in der Regel nicht modifiziert und der Partner GlyProVac hatte genau dieses bei den Proteinen der infektiösen Keime nachgewiesen. Dazu wurden glykosyliertes und nicht glykosyliertes Protein hergestellt und zur Immunisierung von Kaninchen und in einem Infektionsmodell an der Odense Universität verwendet. Die nach Verabreichung beider Varianten gebildeten Antikörper wurden dahingehend untersucht, ob sie verschiedene Erkennungsmuster der von ihnen im YghJ erkannten Bindestellen/Epitope aufweisen. Dazu wurde eine Methode zur Kartierung von Epitopen verwendet, die vom Fraunhofer IZI Partner mitentwickelt wurde. Um die bei dieser Methode entstehenden Daten auswerten zu können bedarf es einer speziellen Auswertesoftware, die von der epitopic GmbH beständig weiterentwickelt wird.

Am einfachsten lässt sich die verwendete Variante einer Peptid-Phagedisplay Technik dadurch beschreiben, dass aus mehr als 5 Milliarden Varianten eines Peptids von 16 Aminosäuren durch Bindung an die Antikörper eines Serums bindende Sequenzen herausgefiltert werden. Das resultiert in über einer Millionen bindender Sequenzen. Über den Umweg des Next Generation Sequencing der Peptidgene der gebundenen Phagenpartikel kann praktisch jedes einzelne gebundene Peptid bestimmt werden. Daten für bindende Peptide an hunderte von verschiedenen Antikörpern können gesammelt werden. Unter den gewählten experimentellen Bedingungen lassen sich einzelne Gruppen von Peptiden mit Hilfe statistischer Berechnungen anhand ihrer Aminosäuremotive aus den Peptiddatensätzen zusammenfassen und bei bekannter Proteinsequenz Epitope verschiedener Antikörper identifizieren. Das war Stand der Technik zu Beginn des Projekts.

Um nun die Unterschiede in der Immunisierung mit verschiedenen Antigenen herauszuarbeiten war es notwendig die bestehende Software und Arbeitsabläufe in Bezug auf den Vergleich möglichst vieler Datensätze zu optimieren. Aus Datensätzen von Selektionsexperimenten mit den Seren am IZI wurden zunächst die Epitope der Antikörper ermittelt, die das Protein direkt über Aminosäuresequenzen erkennen. Diese wiesen bereits im Vergleich von verschiedenen immunisierten Kaninchen deutlich unterschiedliche Muster auf, auch wenn viele Epitope erwartungsgemäß gleich waren. Auf Basis der Analysen der epitopic konnten beim Partner IZI dann auch die passenden Peptide getestet und validiert werden.

Die besondere Herausforderung für die statistische Analyse war es dann aber, dass auch Peptide identifiziert werden, die molekulare Mimikry der auf den Proteinen vorhandenen Glykosylierung sind. Die Datensätze aus den immunisierten Kaninchen boten hierzu die einmalige Gelegenheit einer kontrollierten Immunisierung, um eine generell anwendbare Methode zur Identifizierung solcher Mimotope zu entwickeln. Antikörper erkennen ja nicht nur Proteine, sondern in allen Datensätzen sind auch Informationen zu anderen Molekülen vorhanden, die von Antikörpern erkannt werden, an die wiederum Peptidstrukturen/Mimikry binden.

In der Literatur werden auch immer wieder subtraktive Verfahren bei der experimentellen Selektion der Peptide beschrieben. Diese Experimente am IZI wurden mit einer statistischen Analyse der verschiedenen Datensätze von (nicht)glykosylierten Antigenen verglichen. Nach einer sehr

aufwendigen und nicht immer gleich erfolgreichen Suche nach brauchbaren statistischen Ansätzen konnten verschiedene erfolgreiche Methoden gefunden werden, die es ermöglichen Gruppen von Peptiden mit ähnlichen Motiven zu identifizieren. Hauptproblem ist dabei der Umgang mit Daten in der Größen von einigen Terabyte, denn letztlich wurden ca. 61 Mio. Peptidsequenzen aus 221 Datensätzen bearbeitet.

Tatsächlich konnten Peptide gefunden werden, die ausschließlich Antikörper binden, die spezifisch die Glykosylierung des verwendeten Antigens binden. Das konnte von den Partnern am IZI mit entsprechenden Experimenten belegt werden. Die verschiedenen Methoden zur Identifizierung dieser Peptide wurden standardisiert. Letzten Endes wurden die dazu notwendigen Algorithmen in die Auswertesoftware so integriert, dass sie in Zukunft auch in Routineprojekten für Kunden eingesetzt werden können.

Dass dieses nur mit diesen speziellen Serumproben möglich war, zeigte die Analyse der Daten aus Seren von Schweinen und letztlich auch Patienten. Dort ist es so, dass Kontakt zu weitverbreiteten Bakterien mit YghJ oder nahverwandten Proteinen (AcfD) gleicher Funktion einer ganzen Reihe von anderen bei Harnwegsinfektion auftretenden Bakterien, bereits in der Vergangenheit eine Immunantwort ausgelöst hat. Diese Bakterien finden sich nämlich auch im Verdauungstrakt von Menschen und Schweinen. Die Gedächtniszellen des Immunsystems werden bei Kontakt mit dem YghJ aktiviert und eine differenzierte Analyse wie oben ist nicht mehr möglich.

Als weiteres Ergebnis des Gesamtprojekts wäre daher auch hervorzuheben, dass zwar alle untersuchten Seren (Menschen und Schweine mit Harnwegsinfekten und immunisierte Kaninchen) die gleichen Arten von Antikörpern gegen verschiedene Teile des YghJ Proteins aufweisen, aber die Immunsysteme der ersten beiden bereits gelernt haben, entsprechende Antikörper zu binden. Insbesondere überraschte bei einer Analyse der Sequenzen, dass in den Patienten mit Harnwegsinfekten durchweg Peptidsequenzen dominieren, die den in allen AcfD konservierten Sequenzen entsprechen. Dadurch ist zumindest theoretisch eine optimale Immunantwort auf verschiedene Pathogene gegeben.

Die wesentlichen Ergebnisse des gesamten Projekts zusammengefasst:

- Die glykosylierten Proteine rufen auch auf Ebene der Epitope eine „natürliche“ Immunantwort hervor und sind den nicht glykosylierten Proteinen daher als Impfstoff überlegen.
- Für Bewertung solcher Studien ist die Wahl der Versuchstiere absolut entscheidend, eine Studie der Epitope insbesondere der vorhandenen Antikörper ist wichtig für die Bewertung von Impfstudien dieser Art.
- Peptidmimotope für nicht proteinogene Antigene lassen sich auf statistischem Wege und damit minimalem Probenvolumen bestimmen.

Dadurch ergeben sich Möglichkeiten Antigene aus komplexen organischen Molekülen durch vergleichsweise leicht hergestellte Peptidmimotope ersetzen und zum Beispiel in der Identifizierung solcher Antikörper im der Immundiagnostik einzusetzen.

Die ersten Ergebnisse wurden bereits im Journal *Frontiers in Immunology* (2023, 14:1258136. doi:10.3389/fimmu.2023.1258136) veröffentlicht: Kern K., Delaroque N., Boysen A., Puder M., Wendt R., Kölsch A., Ehrentreich-Förster E., Stärk K., Andersen TE., Andersen K., Lund L., Szardenings M.: Glycosylation of bacterial antigens changes epitope patterns.

Teil II

Die Arbeiten an dem Projekt bei epitopic liefen in enger Kooperation mit dem Fraunhofer IZI und den anderen Partnern. Die Arbeiten bestanden aus 3 Teilprojekten, die zeitlich und thematisch abgegrenzt waren.

1. Die Identifizierung der Epitope von Antikörpern in verschiedenen Immunsystemen (Mensch, Kaninchen, Schwein) gegen das Protein YghJ zur Überprüfung der Hypothese einer Änderung der Immunantwort nach bei glykosyliertem Protein.
2. Erarbeitung von statistischen Algorithmen, die Motive von Peptiden identifizieren, die an Antikörper binden und dabei aber die Struktur zum Beispiel eines Zuckermoleküls imitieren, den sogenannten Mimotopen.
3. Integration der als sinnvoll erkannten Vorgehensweisen zur Identifizierung von Mimotopen in die Software, die bereits zur Identifizierung normaler Epitope verwendet werden.

Der Hintergrund des Projekts ist die berechnete Annahme, dass Modifikationen rekombinanter Impfstoffe durch die posttranslationale Modifizierung, wie zum Beispiel die Glykosylierung, zu unterschiedlichen Immunantworten führen kann. Der Partner GlyProVac aus Odense (DK) hatte diese Glykosylierungen für das bakterielle Protein YghJ nachgewiesen. Dieses gehört zu einer Klasse von Proteinen, die es Bakterien ermöglichen die eigentlich „aseptisch“ konstruierten Innenseiten zum Beispiel des Epithels der Harnwege zu besiedeln. GlyProVac ist es gelungen spezielle Bakterienstämme zu schaffen, die dieses Protein quasi „naturidentisch“ nachbauen. Dazu sind alle anderen im Labor verwendeten Bakterien normalerweise nicht in der Lage. Das gesamte Projekt sollte zeigen, ob und welche Antikörper in Falle einer Immunisierung mit Proteinvarianten im Detail gebildet werden. Dazu sollten die Muster der von den Antikörpern erkannten Epitope erkannt werden und weitere gegen die Glykosylierungsstrukturen gebildete Antikörper über Peptide-Mimotope nachgewiesen werden.

Epitopic hat dazu zusammen mit dem Fraunhofer IZI Methoden entwickelt, mit denen sich Epitope identifizieren lassen. Der Prozess läuft über die Bindung einer extrem großen Zahl von Peptidvarianten auch an Antikörpermischungen, wie z.B. Impfseren, Erfassung von hunderttausenden von gebundenen Sequenzen und die statistische Auswertung von angereicherten Peptidmotiven. Dieses beruht auf der Basis der Statistik von Motiven, genauer gesagt der Veränderung der Häufigkeit bestimmter Motive. Im folgenden Text wird dieses in der Regel als Anreicherung beschrieben.

Das Teilprojekt IV.1 wurde gemeinsam von epitopic und Fraunhofer IZI bearbeitet. Dabei war epitopic für die Datenanalyse zuständig und das IZI für die Generierung der Daten durch Selektionen und Verifizierung der gefundenen Peptidepitope durch Peptid-Mikroarraymessungen im Labor.

Um möglichst viele Epitope zu erfassen und auch einen ersten Eindruck von den Auswirkungen der Glykosylierung zu erhalten, wurden Kaninchen mit glykosyliertem und nicht glykosyliertem YghJ-Protein immunisiert. Eine vergleichende Untersuchung der experimentell angereicherten Peptide zeigte rasch, dass sich zum Teil deutlich unterschiedliche Antikörper gebildet hatten.

Das kann man zum Beispiel wie in Abb. 1 visualisieren, indem man die Anreicherung der 4-mer des Proteins über seine gesamte Länge als Graphik darstellt. Augenscheinlich werden Peptide mit 4-mer Motiven aus einigen Bereichen des YghJ durch Bindung an Serumantikörper je nach Art des verwendeten Proteins eher angereichert. Ob durch die Glykosylierung bestimmte Epitope auf dem Protein besser zugänglich werden oder andere eher geschützt oder ob andere Effekte hier eine Rolle spielen lässt sich aus diesen Daten zunächst nicht erschließen

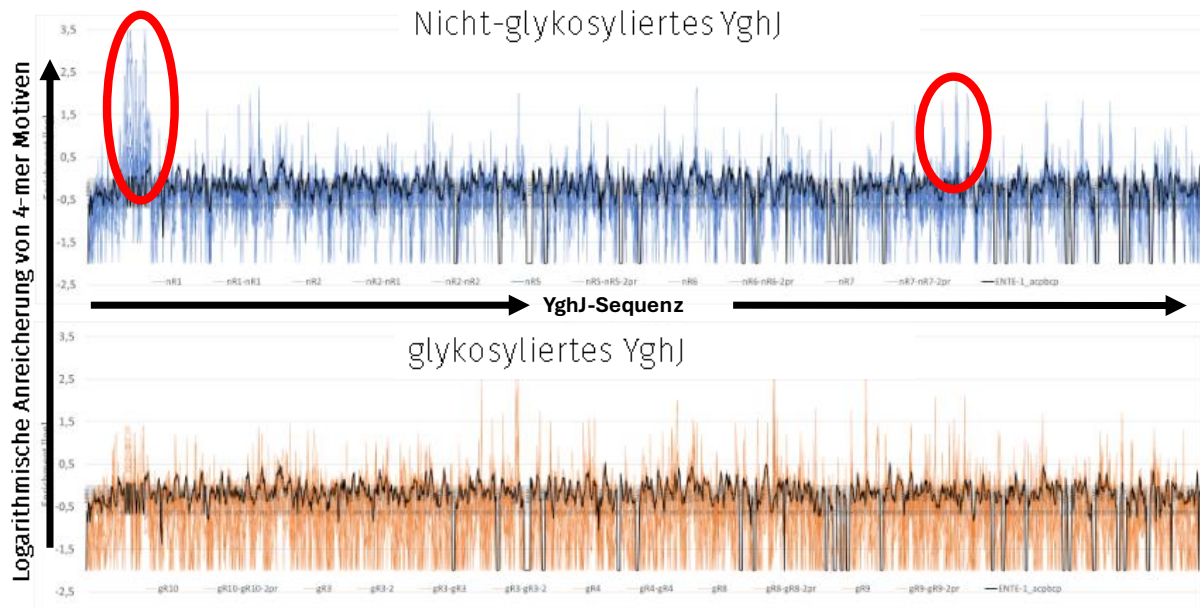


Abbildung 1: Anreicherungskurve von Motiven des YghJ-Proteins zum Zeitpunkt nach der Immunisierung (T1) mit glykosyliertem YghJ (orange Kurve) und nicht-glykosyliertem YghJ (blaue Kurve). Rote Kreise markieren die durch die Glykosylierung hervorgerufene Veränderung bei der Erkennung des Antigens durch Kaninchen-Antikörper.

Beim genauen Hinsehen wird man aber in den Kurven aber auch in dem jeweils anderen Immunisierungsexperiment eine deutliche, d.h. >5000x, Anreicherung finden. Das bedeutete zunächst einmal, dass es hier kein Schwarz und Weiß gibt, sondern dass die Übergänge eher fließend sind. Die Ursachen dafür sind allerdings wohl darin zu suchen, dass

1. Die Glykosylierung der rekombinanten Proteine nicht perfekt ist.
2. Einzelne Kaninchen eher zufällig dem in der Natur sehr verbreiteten Antigen schon zuvor ausgesetzt waren.

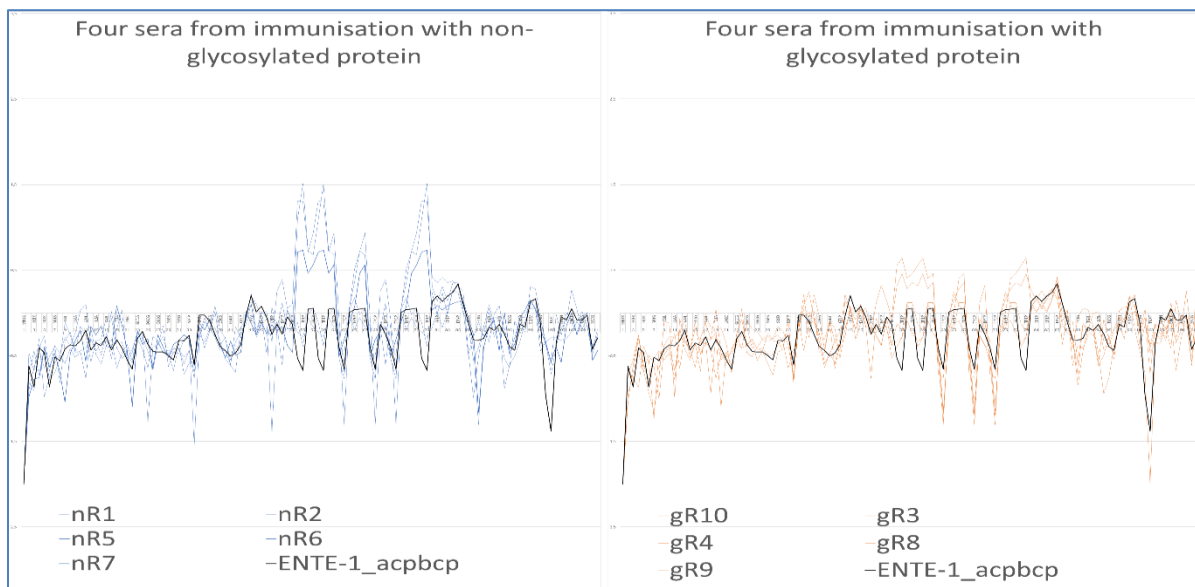


Abbildung 2: Bei den Versuchstieren waren die Immunantworten teils recht individuell. Diese Anreicherung von Motiven am N-terminus des YhJ ist typisch für die Antwort auf glykosyliertes Protein. Trotzdem wird sie auch bei einzelnen Tieren nach der Immunisierung mit glykosylierten Protein oder sogar davor beobachtet, obwohl die Kaninchen eigentlich immunologisch naiv sein sollten.

Während Kaninchen normalerweise eher nicht als Modell für Harnwegsentzündungen geeignet sind, können Schweine von Größe und Physiologie her durchaus mit den Menschen verglichen werden. An der Universität Odense wurden solche Tiermodelle entwickelt. Epitop-Daten aus den Seren sowohl mit YghJ immunisierter als auch mit Bakterien infizierter Schweine wurden bei epitopic analysiert. Es zeigte sich schnell, dass aus der Perspektive der B-Zellantwort das Bild völlig anders ist, als bei den Kaninchen. Wie auch später in den Peptidarrays bestätigt wurde, haben alle Tiere bereits Kontakt zu YghJ und anverwandten Bakterien (s.u. bei Patienten) gehabt. Das bedingt auch, dass die Seren der Tiere auch Antikörper gegen bakterielle Glykosylierungsstrukturen enthielten. Das erschwert die Auswertung der Daten insbesondere auch in Hinblick auf die Wirkung der verschiedenen Impfstoff-Kandidaten.

Last not least wurden die Seren von Patienten untersucht. Diese hatten zuvor zumindest einmal eine Harnwegsinfektion überstanden und wurden nach verschiedenen Pathogenen unterschieden. Tatsächlich stellte sich aber heraus, dass die identifizierten Epitope alle mit den Teilen der YghJ Proteinsequenz zusammenfallen, in denen das Protein mit den verwandten AcfD anderer Pathogene von Harnwegsinfektionen identisch ist. *AcfD* steht dabei accessory colonization factor D, ursprünglich aus *Vibrio cholerae*, und ist für alle diesen Bakterien notwendig zur Besiedelung der Epithelschichten der Harnwege.

Die praktische Testung von über 100 Peptiden aus der Standardbestimmung von Epitopen und mehr als 20 Mimotopen (Identifizierung siehe nächster Abschnitt) bestätigte die Ergebnisse der rein statistischen Analysen.

Zusammenfassen lassen sich diese Ergebnisse wie folgt:

- Ein Vorimmunisierung kann das Ergebnis von Impfstoffstudien verfälschen, mit einzelnen Epitopes lassen sich die Effekte aber wesentlich besser verstehen.
- Längere Exposition führt zu differenzierten B-Zell-Populationen. Die Ergebnisse für eine immunogene Region mit Patientendaten lassen unterschiedliche Bindestellen für die Antikörperpopulationen verschiedener Patienten erkennen.



Abbildung 3: Die Anreicherung von Peptiden auf Antikörpern der Patienten PS12 und PS15 lässt für das Epitop-Motiv 1178_PKFYKDGWEWK unterschiedliche Bereiche der Relevanz von Teilbereichen erkennen.

- Schweine sind in diesem Fall als Versuchstiere besser geeignet als Kaninchen, weil sie wie die Menschen ein unter anderem aus coli-formen Bakterien bestehendes Mikrobiom besitzen. Daher sind sie deren Proteinen und auch dem YghJ/AcfD ausgesetzt. Wie beim Menschen kann hier eine Impfung in erster Linie zur Stärkung der Immunabwehr führen.

- d) Bei Patienten, die ja in der Regel eher häufiger unter Harnwegsinfektionen leiden, bilden sich letztlich vor allem Antikörper, die die verschiedenen Varianten der eng verwandten AcfD-Proteine verschiedener Pathogene erkennen.

Consensus	L I Y - - g - - - - - f - G - - - A p - y - - - k d g - W - n - l - - s - A P - g E - - S -											
SV1-N-1165	- - - - - D D - - - V S A N F T F T - - - - - A P F Y - - - - - K D G E W K N D L - D - - - -											
Y-1179-N1	- - - - - - - - - - - - - - - - - - - - - F Y - - - - - K D G E W K N D L - - - - -											
SV1-N-1180	- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -											
YghJ sequence	L I Y I K G - - - - - D S K D D - - - V S A N F T F T G V V K A P F Y - - - - - K D G E W K N D L - D S P A P L G E L E S A											
Sequence Ruler	1,160 1,170 1,180 1,190 1,200											
YghJ [Escherichia coli]	L I Y I K G - - - - - N S S T N - - - E S A S F T F T G V V K A P F Y - - - - - K D G A W K N D L - N S P A P L G E L E S D											
AcfD Escherichia fergusonii	L I Y I K G - - - - - D S K E V - - - Q S A D F T F T G V V K A P F Y - - - - - K D G K W Q H D L - N S P A P L G E L E S A											
AcfD_E	L I Y I K G - - - - - D S K E V - - - Q S A E F T F T G V V K A P F Y - - - - - K D G Q W K H E L - S S P A P L G E L E S E											
AcfD Shigella flexn	L I Y I K G - - - - - N S S T N - - - E S A S F T F T G V V K A P F Y - - - - - K D G A W K N D L - N S P A P L G E L E S D											
AcfD Shigella sonnei	L I Y I K G - - - - - N S S T N - - - E S A S F T F T G V V K A P F Y - - - - - K D G A W K N D L - N S P A P L G E L E S D											
AcfD Citrobacter freundii	L I Y I K S - - - - - N S Q E V - - - Q S A D F T F T G V V K A P F Y - - - - - K D G K W Q H D L - N S P A P L G E L E S T											
AcfD Enterovibrio norvegicus	L I Y V Q G - - - - - G N A E - - - - - K V A I N L S G T V D A P L Y - - - - - D S A Q K S W V N S I - D S P A P I G D V V S N											
AcfD Photobacterium lutim	L I Y I K P - - - - - L E L Q A G P H S Q A N F T V S G V E K A A F W - - - - - Q D G Q W V H T L Q E A T A P I A E I D T G											
AcfD Photobacterium phosphoreum	L I Y V S G - - - - - N E T G - - - - - N V T V N L N G T V A A P W Y - - - - - K D N A W V N P Q - D S V A P I G E I E S N											
AcfD Photobacterium carnosum	L I Y V S G - - - - - N E I G - - - - - N V T V N L D G T V A A P W Y - - - - - K D N A W V N P Q - D S V A P I G E I E S N											
AcfD Photobacterium kishitanii	L I Y I S G - - - - - N E T G - - - - - N I T V N L D G T V A A P W Y - - - - - K N N A W V N P Q - N S V A P I G E I E S D											
AcfD Shewanella indica	L I Y I Q P - - - - - K E Q G L - - - G E V T F N L E G V E T A A W W - - - - - K E G Q W I H S L E S A T A P I A E I D T G											
AcfD Shewanella schlegeliana	L I Y V Q G Q G I Q N Q E S Q - - - - - T V T L N F T G S V D A P L Y Q Y D G K G G Q W L N P L - D S P A P I G E V I S Q											
AcfD_Vibrio-campb	L I Y I Q P - - - - - L E V D S R - - - D V V T F N F T G V L R A S W W - - - - - K N G S W L N P I - N T D V P L A E I D S G											
AcfD_vibrio-ch_	L I Y A Q G - - - - - G N S Q - - - - - Q V K L T F S G T I D A P L Y - - - - - I D G K W R N P L - L S G A P I G E V V S D											
ACFD_VIBC3	L I Y A Q G - - - - - G N S Q - - - - - Q V K L T F S G T I D A P L Y - - - - - I D G K W R N P L - L S G A P I G E V V S D											
AcfD_Vibrio_paraaem	L I Y A Q G - - - - - G N S E - - - - - S V Q L T F T G T L A A P L F - - - - - I D G S W K N D L - D S P A P V G E V V S K											
AcfD part Klebsiella pneu	- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -											
AcfD partial Klebsiella pneum	- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -											
AcfD Photobacterium leiogn	L I Y V Q G - - - - - N E T N - - - - - P V T L N F S G T V N A P W F - - - - - K I D K W I N D Q - S S P A P I G E I E S N											
AcfD Shewanella sp	L I Y I K P - - - - - T S T D E L - - - N T V E F S F T G V L E A S L W - - - - - R N G K W L T P V - N L E V P L A E I N T G											

Abbildung 4: Die von Antikörpern erkannten Peptid-Epitope SV1-N-1165, Y-1179-N1 und SV1-N-1180 (oben) im Alignment mit YghJ und anderen AcfD Proteinen. Vor allem die letzteren Peptide decken sich mit einem sehr stark konservierten Bereich der Proteinfamilie.

- e) Last not least, das Immunsystem von Menschen und Tieren bildet auch auf Epitop-Ebene fast identische Antikörper. Entscheidend für die Vergleichbarkeit sind aber auch Vorimmunisierungen.

Insgesamt wurden alle Ziele erreicht. Die Identifizierung der Peptidepitope ist eine reine Routinearbeit, die Auswertung der Daten in Bezug auf die Impfstoffentwicklung führte jedoch zum Teil zu überraschenden Ergebnissen, zum Teil widerspiegelten sich Effekte die zuvor bei anderen Infektionskrankheiten beobachtet wurden.

In jedem Fall konnte auch die Mehrzahl der Peptidepitope abgedeckt und vom IZI experimentell durch Messungen mit einer größeren Zahl von Seren bestätigt werden. Das Wissen um diese Epitope war eine Voraussetzung für die Durchführung der folgenden zwei Arbeitspakete, da noch nicht einmal in der sonst recht vollständigen Immune Epitope Database (<https://www.iedb.org/>) Epitope dieser Proteine erfasst sind.

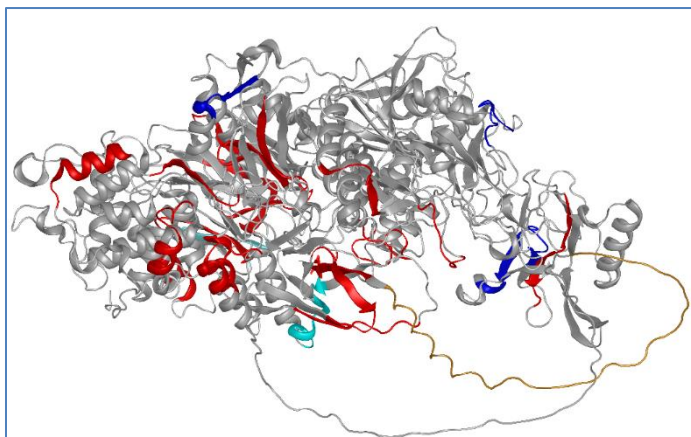


Abbildung 5; Dreidimensionales Modell des YghJ-Proteins mit farblich markierten Epitopen. Die Glykosylierung führt in den hellblauen Bereichen zu geringeren und in den dunkelblauen Bereichen zu höheren Antikörpertitern. Keine Veränderung wurde in den roten Bereichen gefunden. (AF-Modell von YghJ (E3PJ90, <https://alphafold.ebi.ac.uk/entry/E3PJ90>))

entstehen. Insofern war es nicht erstaunlich, dass eine Reihe von insbesondere Schweinen bereits Antikörper gegen diese Mimotope besaß. Bei Kaninchen ist dieses weniger der Fall.

Dadurch dass in diesem Fall alle Details einer größeren Zahl von Datensätzen, sequenzspezifische Epitope und die exakten Antigene bekannt waren, konnten die Berechnungen bereits „in silico“ überprüft werden, was die Erprobung verschiedenster Analysemethoden erlaubte. Das Ergebnis des Arbeitspakets war eine Reihe von Regeln, die bei der globalen Analyse der Anreicherung von Motiven zu beachten sind, und eine Reihe von Möglichkeiten auch ohne Kenntnis des Hintergrunds der Proben, Peptide mit Bindungseigenschaften zu identifizieren. Theoretisch lassen sich damit beliebige Binder an alle möglichen Strukturen identifizieren. Bei hinreichender Qualität der Datensätze sogar zum Beispiel Peptide, die an pharmakologisch relevante Strukturen (Rezeptoren) binden.

Arbeitspaket IV.3 umfasste die Eingliederung der in Paket IV.2 erlangten Erkenntnisse bei der Identifizierung von Mimitopen von YghJ in die bei epitopic bestehende Software „LibDB“ in einer allgemein-gültigen Version. Dementsprechend war die manuelle Analyse am Beispiel von glykosyliertem und nicht-glykosyliertem YghJ nur der „Proof of Concept“ und sollte nun für die Anwendung an beliebigen Kundenprojekten zur routinemäßigen Anwendung angepasst werden. Dafür mussten einige Hürden genommen werden. Allen voran musste eine Vielzahl an strukturellen Anpassungen vorgenommen werden, da das Kernprogramm noch auf älteren Bibliotheken und Abhängigkeiten basiert. Außerdem war die Software am Anfang dafür ausgelegt gewesen, einen Datensatz zu einem Zeitpunkt zu verarbeiten, was inzwischen bereits erweitert worden ist, aber noch nicht in einem für dieses Projekt angemessenen Umfang.

Experiment 2022 - 01 - score testing

Please choose **UP TO 4** datasets (2 pairs of 1st and 2nd selection round) where motifs s identified (save your selection!)

save selected experiments

Name	Seq Count	Valid Count	Motif Count	Pattern	
██████████	557063	361562	136439	ENTE1	<input checked="" type="checkbox"/>
██████████	992899	634635	135359	ENTE1	<input type="checkbox"/>
██████████	612074	403395	139365	ENTE1	<input type="checkbox"/>
██████████	445083	287893	139293	ENTE1	<input type="checkbox"/>

save selected experiments

Please select **UP TO 8** reference datasets to compare against! (save your selection!)

save selected references

Name	Seq Count	Valid Count	Motif Count	Pattern	
██████████	557063	361562	136439	ENTE1	<input type="checkbox"/>
██████████	992899	634635	135359	ENTE1	<input type="checkbox"/>
██████████	612074	403395	139365	ENTE1	<input type="checkbox"/>
██████████	445083	287893	139293	ENTE1	<input type="checkbox"/>

save selected references

Please select **only ONE** library dataset to compare against! (save your selection!)

save selected naive dataset

Name	Seq Count	Valid Count	Motif Count	Pattern	
██████████	1231651	985371	141494	ENTE-1-allCodons	<input type="checkbox"/>
██████████	2976762	2275840	143713	ENTE-1-allCodons	<input type="checkbox"/>
██████████	1745111	1290469	142331	ENTE-1-allCodons	<input type="checkbox"/>
██████████	1473587	1067675	139465	ENTE1	<input type="checkbox"/>
██████████	860390	586611	139455	ENTE1	<input type="checkbox"/>
██████████	619328	386457	139364	ENTE1	<input type="checkbox"/>

Dies sollte durch die neu hinzugefügte Methode so geändert werden, dass nun maximal 4 Analyse-Datensätze, 8 Kontrolldatensätze aus verschiedenen Projekten und 1 naiver Bibliotheks-Datensatz gleichzeitig abgerufen und geöffnet werden können. Dafür musste einerseits die physische Leistung des Host-Servers erweitert werden, aber auch die Menge an Variablen, die die Software zu jedem Zeitpunkt halten kann. Jedes der knapp 140.000 4-mer Motive in jedem der bis zu 13 Datensätze besitzt noch einmal 7 statistische Unterparameter, die während der Analyse gleichzeitig abrufbar sein müssen, damit die formulierten Regeln für die Filterung dieser angewendet werden können. Die Schwellwerte einiger dieser Parameter können vor dem Start der automatischen Methode in einem grafischen Benutzer-Interface eingetragen werden. Sie sind gekoppelt an die Auswahl der jeweiligen Datensätze, die für die Auswertung genutzt werden sollen. Die folgenden Bilder zeigen

Ausschnitte aus der Benutzeroberfläche:

██████████	2788025	1879672	166138	ENTE-4-all codons	<input type="checkbox"/>
██████████	1141785	363835	136958	ENTE2	<input type="checkbox"/>
██████████	594911	351011	139416	ENTE1	<input type="checkbox"/>

save selected naive dataset

Please enter the parameters for the identification process. Otherwise you can just keep the default values for a starting point

Please enter the enrichment threshold for the reference datasets (default: 0.2) (everything below will be considered)

Please enter the enrichment threshold for the library dataset (default: 0.3) (everything below will be considered)

Please enter the maximum number of motifs in the output (default: 1,000) (the output will contain results up to this number)

Search

Abbildung 7: Dargestellt sind zwei Ausschnitte aus der Benutzeroberfläche von LibDB, die die in der Datenbank enthaltenen Datensätze samt einiger Parameter wie Gesamtzahl Sequenzen oder Gesamtzahl Motive zeigen. Ebenfalls zu sehen sind die Eingabemasken der Parameter für die automatische Bestimmung der Epitope/Mimitope. (Datensatz-Namen sind aus Datenschutzgründen geschwärzt)

Wurden alle Datensätze ausgewählt und die entsprechenden Parameter eingetragen, kann die Suche gestartet werden und es wird der Ergebnis-Bildschirm angezeigt:

Identified following motifs in target run (193) compared against reference (193) and library datasets

Export list as CSV
Export seqs as Fasta

Motif	Analysis DS-1	Enrichment-1	Count-1	Analysis DS-2	Enrichment-2	Count-2	Analysis DS-3	Enrichment-3	Count-3	Analysis DS-4	Enrichment-4	Count-4	Reference DS-1	Ref-Enrichment-1	Ref-Count-1	Reference DS-2	Ref-Enrichment-2
STCQ	K12-1pr	0,82623	221	K12-2pr	2,543	3482	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,22478	190	K16-2pr_merged	-0,35458
NSKG	K12-1pr	0,4386	97	K12-2pr	2,70353	27072	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,01981	127	K16-2pr_merged	-0,30775
NSKG	K12-1pr	0,45929	106	K12-2pr	2,68723	27168	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,09407	157	K16-2pr_merged	-0,24057
QHWY	K12-1pr	0,92602	51	K12-2pr	2,59709	1401	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,19491	52	K16-2pr_merged	-0,29408
ITCQ	K12-1pr	0,74191	182	K12-2pr	2,58366	15953	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,28048	216	K16-2pr_merged	-0,28864
HWYN	K12-1pr	0,46477	27	K12-2pr	2,49735	4393	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,07937	38	K16-2pr_merged	-0,08825
QTTC	K12-1pr	0,71153	205	K12-2pr	2,42237	15974	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,2219	228	K16-2pr_merged	-0,45088
HALE	K12-1pr	0,38458	27	K12-2pr	2,37609	4015	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,14979	54	K16-2pr_merged	0,0726
EHAL	K12-1pr	0,31494	23	K12-2pr	2,37174	3975	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,18083	58	K16-2pr_merged	-0,47147
WYNR	K12-1pr	0,41561	33	K12-2pr	2,35924	4395	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,09148	49	K16-2pr_merged	0,23208
QHWY	K12-1pr	0,01767	2	K12-2pr	2,32962	632	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,35692	15	K16-2pr_merged	-0,13401
PTNQ	K12-1pr	0,3799	35	K12-2pr	2,28402	4256	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,15725	72	K16-2pr_merged	-0,13022
TNQP	K12-1pr	0,41561	38	K12-2pr	2,2697	4118	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,58557	193	K16-2pr_merged	1,78036
QJWY	K12-1pr	0,48688	53	K12-2pr	2,2273	4423	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,30937	121	K16-2pr_merged	-0,56537
KGEP	K12-1pr	0,06536	25	K12-2pr	2,21342	5311	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,01242	76	K16-2pr_merged	-0,8822
TCCQ	K12-1pr	0,63232	148	K12-2pr	2,21082	8503	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,43946	326	K16-2pr_merged	-0,43435
TNQP	K12-1pr	0,36676	42	K12-2pr	2,25907	4389	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,05528	70	K16-2pr_merged	-0,50508
GEPT	K12-1pr	0,31172	37	K12-2pr	2,19829	4321	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,27958	118	K16-2pr_merged	1,00681
KGEW	K12-1pr	0,04764	6	K12-2pr	2,17002	1206	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,06221	16	K16-2pr_merged	-0,28014
EPFN	K12-1pr	0,25333	35	K12-2pr	2,13749	4236	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,02471	71	K16-2pr_merged	-0,71342
TQWY	K12-1pr	0,28336	1	K12-2pr	2,13897	401	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,1202	5	empty	-10,000
TCCQ	K12-1pr	0,39323	98	K12-2pr	2,12479	7973	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,2261	228	K16-2pr_merged	-0,71967
ALEY	K12-1pr	0,1072	25	K12-2pr	2,12405	3941	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,00701	66	K16-2pr_merged	-0,50457
RPPG	K12-1pr	0,23563	37	K12-2pr	2,12079	4307	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,10496	58	K16-2pr_merged	-0,05829
STLQ	K12-1pr	0,51486	93	K12-2pr	2,11205	5578	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,13114	132	K16-2pr_merged	0,35533
NRPY	K12-1pr	0,27197	43	K12-2pr	2,09135	4302	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,09798	63	K16-2pr_merged	-0,6101
WHSK	K12-1pr	0,01127	34	K12-2pr	2,09129	6529	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,16475	82	K16-2pr_merged	-0,69443
TCCQ	K12-1pr	0,61848	129	K12-2pr	2,07874	5645	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,3837	258	K16-2pr_merged	-0,47746
NDPI	K12-1pr	0,37089	54	K12-2pr	2,07363	4130	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,23622	136	K16-2pr_merged	0,12364
QPN	K12-1pr	0,27197	43	K12-2pr	2,07269	4121	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,0422	87	K16-2pr_merged	-0,25791
PDHE	K12-1pr	0,17825	22	K12-2pr	2,0571	2524	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,15127	71	K16-2pr_merged	-0,88988
DHEG	K12-1pr	0,28299	28	K12-2pr	2,04912	2478	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,09553	44	K16-2pr_merged	-0,91482
QSTL	K12-1pr	0,498	105	K12-2pr	2,04302	5585	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,15648	165	K16-2pr_merged	-0,62052
PPDH	K12-1pr	0,17825	22	K12-2pr	2,03486	2398	0,000	0	0,000	0	0,000	0	K16-1pr_merged	-0,06452	43	K16-2pr_merged	-0,31985
WHWD	K12-1pr	0,3187	8	K12-2pr	2,02929	633	0,000	0	0,000	0	0,000	0	K16-1pr_merged	0,15855	19	K16-2pr_merged	-0,3101

Abbildung 8: Ein Ausschnitt aus der Ergebnis-Tabelle, die angezeigt wird, wenn die automatische Bestimmung von Kandidaten für Epitope oder Mimotope erfolgreich war. Die Spalte „Motif“ zeigt das jeweilige 4-mer Motiv, dass anhand der Parameter vom Algorithmus bestimmt worden ist. Danach folgen die jeweiligen, beim Start der Methode ausgewählten, Datensätze und die Parameter des Motivs in diesen Datensätzen (Enrichment: Anreicherung des Motivs im Datensatz im Vergleich zur erwartbaren Häufigkeit abgeleitet aus der naiven Bibliothek; Count: absolute Häufigkeit des Motivs in allen Sequenzen des Datensatzes). Die Reihenfolge von links nach rechts ist dabei fest und folgt diesem Muster: Analyse-Datensatz 1,2,3,4; Vergleichsdatsatz 1,2,3,4,5,6,7,8; Naive Bibliothek. Falls weniger als 4 Analysedatensätze bzw. 8 Vergleichsdatsätze ausgewählt wurden, bleiben die übrigen Spalten leer bzw. sind auf 0 gesetzt. Das obige Beispiel zeigt das Ergebnis einer Suche auf Kaninchen-Daten (K12 vs. K16) wo bereits weit oben in der Tabelle die Motiv-Kandidaten NSKG, SKGE und KGEW sichtbar sind (siehe Umrahmung). Sequenzen mit diesen Motiven konnten als Peptide synthetisiert und verifiziert werden, sodass NSKGEW schließlich als Epitop identifiziert wurde.

Anhand der Ergebnis-Tabelle können erste Epitop- oder Mimotop-Kandidaten leicht identifiziert werden, wodurch eine umfangreiche manuelle Suche im Vorfeld vermieden werden kann. Ausgehend von diesem ersten Ergebnis können dann gezieltere Suchen nach einzelnen Sequenzen durchgeführt werden, die schließlich als Alignment dargestellt und hinsichtlich ihrer Plausibilität untersucht werden können. Die Daten können selbstverständlich auch exportiert und im CSV-Format gespeichert werden.

Ein weiteres mathematisches Hilfsmittel, was im Rahmen des Projekts in LibDB integriert worden ist, sozusagen als Vorstufe der automatisierten Methode, war die Berechnung eines gewichteten Scorings sowie weiterer einfacherer statistischer Scores für die einzelnen Motive eines gewählten Datensatzes. Bei der manuellen Analyse der Mimotop-Kandidaten in Paket IV.2 kam es hin und wieder vor, dass es trotz offensichtlicher Unterschiede gleiche Werte für die Anreicherung zweier Motive gab. Das beruht darauf, dass nicht unterschieden wird, ob sich das Motiv nur in einer stark angereicherten Sequenz befindet oder von einer Vielzahl ähnlicher Sequenzen geteilt wird. Im letzteren Fall ist die Wahrscheinlichkeit höher, dass es sich tatsächlich um einen Kandidaten für eine Epitop oder Mimotop handelt. Dementsprechend kann der gewichtete Score trotz des identischen Anreicherungs-Wertes hier eine Unterscheidung vornehmen, sodass die Ermittlung falsch positiver Kandidaten auf ein Minimum beschränkt werden konnte. Weitere Scorings basierend auf dem Mittelwert und dem Median kamen ebenfalls zum Einsatz. Die zugehörigen Daten können ebenfalls exportiert und im CSV-Format gespeichert werden.

Durch die Integration beider Funktionen in die bestehende Software von epitopic, konnte das Ziel des Teilpakets erreicht werden. Eine automatisierte Methode steht für die Bearbeitung von Kundenanfragen zur Verfügung.

Veröffentlichungen / Präsentationen

Eine Publikation wurde bisher veröffentlicht, weitere mit detaillierteren Analysen der Mimotope und zu Auswirkungen des Adjuvans werden folgen:

Kern K., Delaroque N., Boysen A., Puder M., Wendt R., Kölsch A., Ehrentreich-Förster E., Stærk K., Andersen TE., Andersen K., Lund L., Szardenings M.

Glycosylation of bacterial antigens changes epitope patterns.

Front. Immunol., 2023, 14:1258136. doi:10.3389/fimmu.2023.1258136

Im Januar 2023 wurde das Projekt außerdem auf der „PepTalk Conference“ in San Diego vorgestellt:

Michael Szardenings “YghJ: Preclinical Vaccination and Human Disease Studies at the Epitope Level”

Diese Präsentation ist auch online verfügbar.

Zusammenfassend: Das Projekt ermöglichte den Zugang zu einer großen Zahl von kontrollierten Variationen von Immunseren und war daher hilfreich bei der stetigen Weiterentwicklung von Methoden zur

- Analyse komplexer Antikörpermischungen / Seren
- Identifizierung von Antikörper Peptid-Mimotopen
- Identifizierung von Bindern an beliebige Strukturen