*Article*

# Deep Intraclonal Analysis for the Development of Vaccines against Drug-Resistant *Klebsiella pneumoniae* Lineages

Ana Tajuelo [1,2,†], Eva Gato [1,†], Jesús Oteo-Iglesias [3,4], María Pérez-Vázquez [3,4], Michael J. McConnell [5], Antonio J. Martín-Galiano [6,*] and Astrid Pérez [1]

[1] Intrahospital Infections Unit, National Centre for Microbiology, Instituto de Salud Carlos III (ISCIII), Majadahonda, 28220 Madrid, Spain; a.tajuelomp@isciii.es (A.T.); eva.gato@isciii.es (E.G.); astrid.perez@isciii.es (A.P.)

[2] Universidad Nacional de Educación a Distancia (UNED), 28015 Madrid, Spain

[3] Reference and Research Laboratory for Antibiotic Resistance and Health Care Infections, National Centre for Microbiology, Instituto de Salud Carlos III (ISCIII), Majadahonda, 28220 Madrid, Spain; jesus.oteo@isciii.es (J.O.-I.); mperezv@isciii.es (M.P.-V.)

[4] CIBER de Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain

[5] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA; mcconnell.mike75@gmail.com

[6] Core Scientific and Technical Units, Instituto de Salud Carlos III (ISCIII), Majadahonda, 28220 Madrid, Spain

* Correspondence: mgaliano@isciii.es

† These authors contributed equally to this work and should be considered co-first authors.

**Abstract:** Despite its medical relevance, there is no commercial vaccine that protects the population at risk from multidrug-resistant (MDR) *Klebsiella pneumoniae* infections. The availability of massive omic data and novel algorithms may improve antigen selection to develop effective prophylactic strategies. Up to 133 exposed proteins in the core proteomes, between 516 and 8666 genome samples, of the six most relevant MDR clonal groups (CGs) carried conserved B-cell epitopes, suggesting minimized future evasion if utilized for vaccination. Antigens showed a range of epitopicity, functional constraints, and potential side effects. Eleven antigens, including three sugar porins, were represented in all MDR-CGs, constitutively expressed, and showed limited reactivity with gut microbiota. Some of these antigens had important interactomic interactions and may elicit adhesion-neutralizing antibodies. Synergistic bivalent to pentavalent combinations that address expression conditions, interactome location, virulence activities, and clone-specific proteins may overcome the limiting protection of univalent vaccines. The combination of five central antigens accounted for 41% of all non-redundant interacting partners of the antigen dataset. Specific antigen mixtures represented in a few or just one MDR-CG further reduced the chance of microbiota interference. Rational antigen selection schemes facilitate the design of high-coverage and "magic bullet" multivalent vaccines against recalcitrant *K. pneumoniae* lineages.

**Keywords:** antibiotic resistance; betweenness centrality; epitope; immunoinformatics; intrahospital infection; natural processing language; neutralizing antibody; protein hub; virulence factor; reverse vaccinology

## 1. Introduction

Rationally-designed prophylactic strategies have been underutilized against difficult-to-treat pathogens [1,2]. One of these microorganisms is *Klebsiella pneumoniae*, a gram-negative bacterium found in the environment as well as in the gastrointestinal and respiratory tract microbiota of animals and humans [3]. *K. pneumoniae* can cause a range of life-threatening infections, such as sepsis, pneumonia, or meningitis, as well as urinary tract and wound infections. The two most worrisome epidemiological behaviors observed in this pathogen comprise the hypervirulent and the multidrug-resistant (MDR) phenotypes [4].

The former affects young immunocompetent individuals and are usually susceptible to antibiotics [5]. In contrast, the latter, or classical phenotype, involves MDR lineages that cause worldwide opportunistic infections in elderly, immunocompromised, and hospitalized patients [6–9]. MDR *K. pneumoniae* clones are most responsible for increased morbidity and mortality attributed to this species. These are currently estimated to cause a third of intrahospital infections [10]. These lineages, mainly those that are carbapenem-resistant, can cause trans-border outbreaks and are considered critical pathogens by the World Health Organization, for which new treatments are urgently needed [11].

Vaccination may be an effective alternative to conventional chemotherapy for preventing *K. pneumoniae* infections, given the challenges associated with developing new antibiotics. Vaccination use could decrease antimicrobial resistance in several ways, including reduction of carriage, bacterial load, and antibiotic use [12]. However, B-cell epitopes of pathogen antigens often undergo strong selective pressure through mutation, recombination, or deletion to evade the host's immune response [13]. Strategies to develop an effective vaccine against *K. pneumoniae*, including attenuated and inactivated vaccines, outer membrane vesicles, proteins, polysaccharides, and conjugated vaccines, were rather unsuccessful [14,15]. Reasons for failure included lipopolysaccharide toxicity and limited coverage of the 24-valent capsular vaccine. The most advanced prototype is the tetravalent KlebV4 vaccine (identifier: NCT04959344), which is in phase 1/2 trial. As a result, *K. pneumoniae* remains classified into "Group C: pathogens with vaccine candidates either in early clinical trials or with moderate to high feasibility of vaccine development" [16]. This modest biotechnological status, despite the medical relevance of these infections, highlights the need for continued development of novel strategies for combating *K. pneumoniae*.

The limited protection achieved by antigens selected by traditional approaches has led to the use of "omics"-based approaches for antigen identification. On the one hand, experimental techniques such as immunoproteomics have detected real antigens based on patient antisera samples using proteomic tools [17–19]. On the other hand, knowledge-based tools such as reverse vaccinology (RV) use pre-existing data and algorithms to potentially outperform reductionist approaches [20]. RV aims to identify the most protective antigens from complete pathogen genomes after the application of selection criteria that include the presence of epitopes by immunoinformatic tools, sequence conservation, non-homology with human proteins, exposure to the immune system, experimental workability, and functional importance. Protocols have been enriched with a finer knowledge of human immunity as well as further post-genomic criteria involving massive genome utilization and protein structural information [21,22]. Compared to empirical strategies, RV presents advantages such as a major control on the immunological response, the inclusion of whole proteomes in the analysis, the facilitation of updating the vaccine if needed, and savings in terms of labor, cost, and time [20]. Data and algorithmic advances promise to revolutionize antigen selection for vaccines. For example, the availability of tens of thousands of genomic sequences per pathogen enables deep intra-clonal search for epitope drift in lineages of interest. Omic-level expression information allows for the identification of important antigens during critical stages of infection. Machine learning predictive tools using natural processing language and transfer learning [23] have leveraged the accuracy of B-cell epitope predictors [24,25] over preceding methods with moderate performance [26]. These advances utilized in combination may enhance prophylaxis against *K. pneumoniae* by finely tuning antigen selection toward high species coverage as well as limiting cross-immunity effects leading to gut dysbiosis.

In this study, we have taken advantage of a recent algorithm and biological data accessibility to find strongly conserved B-cell epitopes in thousands of proteomes of the six principal highly resistant clones. Protein expression, interactivity, and presence of homologs in gut microbiota species were further considered to identify 11 promising universal antigens to prevent *K. pneumoniae* infections in susceptible populations. Finally, we have designed optimal antigen combinations adequate for multivalent formulations with potentially higher and more accurate protective performance.

## 2. Results and Discussion

### 2.1. Evaluation of Published K. pneumoniae RV Studies

We reviewed the literature to identify omic-level RV antecedents utilized for vaccine design against *K. pneumoniae*. We detected five papers using "*Klebsiella pneumoniae*" and "("reverse vaccinology" or immunoinformatic)" keywords in Pubmed published between 2019 and 2022. Considering its clinical relevance and low vaccine development, these are few RV studies. For example, for other critical pathogens such as *Acinetobacter baumannii*, 11 RV articles were available in 2021 [27].

We selected 24 RV issues, a number of which have been reasonably approached in *K. pneumoniae* studies (Table 1). In general, these strategies were oriented toward the design of multi-epitope hybrid antigens built by peptide linking, as well as structural modeling and molecular dynamic simulation of the resulting polypeptides. These pioneer RV studies mostly resulted in universal antigens for this species, extensible to carbapenem-resistant clones [28–32]. However, other aspects were missing or poorly implemented in the current literature. For instance, no vanguardist B-cell epitope prediction was carried out. Epitope sequence conservation was not profoundly evaluated using thousand-level genome datasets from MDR lineages to anticipate vaccine escape in the actual vaccinable population. Potential interference with the microbiota was weakly approached. The expression pattern and interactomic positioning of antigens were barely considered. Last but not least, no systemic approximation was proposed to find rational antigen combinations that boost the protection of univalent vaccines beyond the empirical merging of top antigens.

**Table 1.** Summary of RV antecedent reports for *K. pneumoniae*. NA: non-applied.

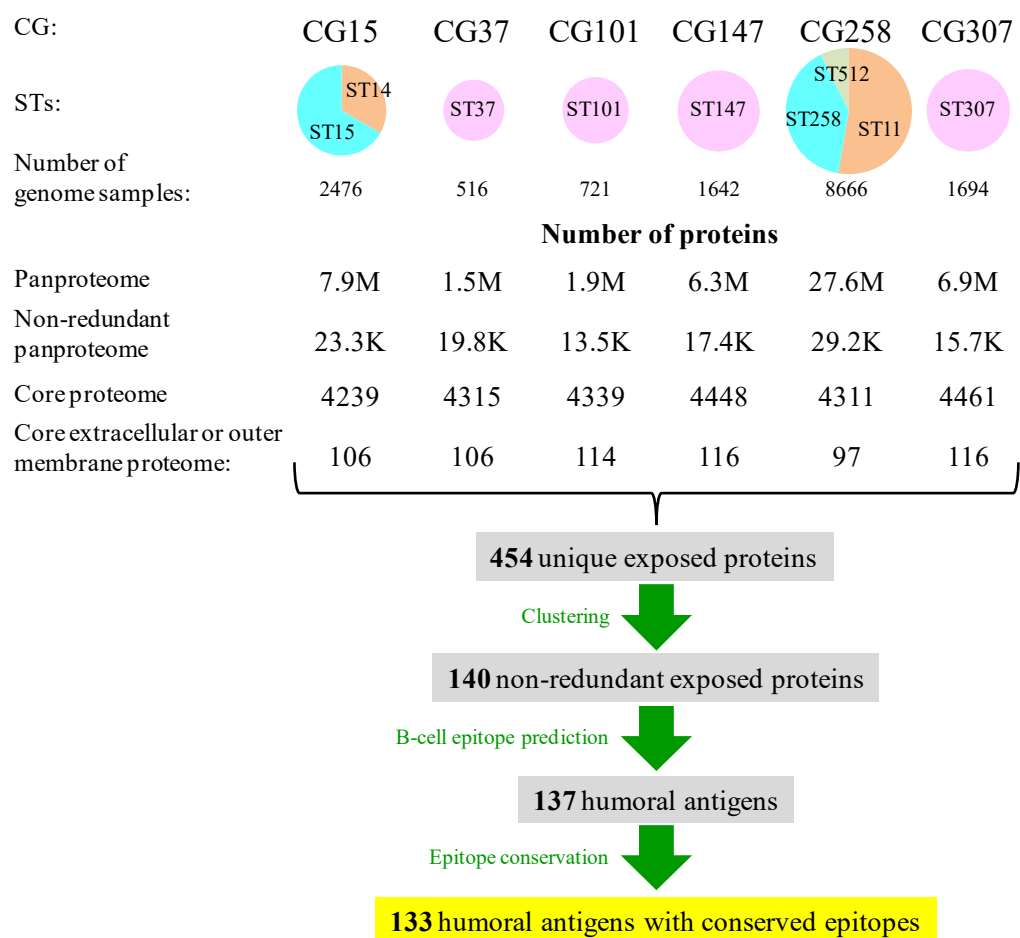| RV Items | Dar et al., 2019 [28] | Mehmood et al., 2020 [29] | Allemailem, 2021 [32] | Wang et al., 2022 [31] | Cuscino et al., 2022 [30] | This Study |
|---|---|---|---|---|---|---|
| **INPUT DATA** | | | | | | |
| Number of genomes considered | 222 | 173 | 973 | 274 | 467 | 15,715 |
| Protein identity threshold | 90% | 50% | NA | Default roary | NA | 70% |
| Clinical phenotype | NA | NA | NA | Prevalent capsular types | 467 carbapenem-resistant isolates | The six most prevalent MDR lineages |
| Antigen clonal prevalence | NA | NA | NA | NA | 95% isolates | 95% isolates within CGs |
| Deep intraclonal analysis | NA | NA | NA | NA | NA | Yes (516–8666 genomes per CG) |
| Protein location | PSORTb 2.0 | PSORTb 3.0 | PSORTb 3.0 | PSORTb 3.0 | PSORTdb 4.0 | PSORTb 3.0.3 |
| **EPITOPE ANALYSIS** | | | | | | |
| Prediction of B-cell epitopes | ABCPred | ABCPred | BepiPred 2.0 | ABCpred | ABCpred | BepiPred 3.0 |
| Epitope coverage | NA | NA | NA | NA | >80% identity (full protein) | Epitope conservation in CGs |
| Epitope sequence conservation | NA | NA | NA | NA | NA | Per residue within CG |
| Prediction of HLA II epitopes | ProPred/MHCPred 2.0 | BcePred | MHCPred 2.0 | IEDB server | NetMHCIIpan-3.2 | NA |
| **SIDE EFFECTS** | | | | | | |
| Homology with human proteins | Whole antigen BLAST | Whole antigen BLAST | NA | NA | NA | Whole antigen and epitope MMSeqs2 |
| Homology with gut microbiota | NA | GFDB | *Lactobacillus* spp. | NA | *E. coli* K-12 | Three-level stratified taxons (307 representative genomes). Whole protein and epitope searching |
| Toxicity | NA | NA | ToxinPred | NA | NA | ToxinPred |
| Allergenicity | AllergenFP 1.0; AllerTOP 2.0 | NA | AllerTOP 2.0 | AllergenFP 1.0; AllerTOP 2.0 | AllergenFP 1.0; AllerTOP 2. | AllerCatPro 2.0 |
| **FUNCTIONAL CONSTRAINTS** | | | | | | |
| Antigen expression | NA | NA | NA | NA | NA | 5 conditions |
| Virulence | MvirDB; VFDB | VFDB | VFDB; VirulentPred; SPAAN | NA | VirulentPred | VFDB |
| Interactomics | NA | NA | NA | NA | NA | Centrality from STRING data |
| Essentiality | DEG | DEG | NA | NA | NA | NA |

**Table 1.** *Cont.*

| RV Items | Dar et al., 2019 [28] | Mehmood et al., 2020 [29] | Allemailem, 2021 [32] | Wang et al., 2022 [31] | Cuscino et al., 2022 [30] | This Study |
|---|---|---|---|---|---|---|
| WORKABILITY | | | | | | |
| In vitro solubility | SOLpro; PROSO II | NA | NA | NA | Protein-sol | SoluProt |
| Transmembrane helices | HMMTOP 2.0; TMHMM 2.0 | NA | NA | NA | NA | Phobius |
| PROPOSED VACCINE | | | | | | |
| Molecular dynamics | GROMACS | NA | AMBER20 | GROMACS | GROMACS | NA |
| Antigen type | NA | NA | Peptides | Epitopes | Epitopes | Antigen combinations |
| Predicted antigenicity | ANTIGENpro; VaxiJen v.2.0 | VaxiJen v.2.0 | NA | VaxiJen | ANTIGENpro; VaxiJen 2.0 | NA |
| Criteria for synergic multivalency | NA | NA | NA | NA | NA | Several |

In conclusion, several RV issues have been reasonably addressed for *K. pneumoniae*. In contrast, other matters should be further explored in order to enhance our understanding of the *K. pneumoniae* vaccine. These have been attempted below.

### 2.2. Core Proteome of Multidrug Clonal Groups and Exposed Proteins

In this study, we focused on the principal MDR lineages that monopolize most MDR intrahospital cases worldwide. These are six CGs, namely CG15 (ST14, ST15), CG37 (ST37), CG101 (ST101), CG147 (ST147), CG258 (ST11, ST258, and ST512), and CG307 (ST307), including nine STs, classified as "problem clones" [9]. In our Spanish Reference Laboratory, these CGs accounted for approximately 75% of all antibiotic-resistant cases [33]. These lineages often carry an array of resistance genes and were named multidrug-resistant CGs (MDR-CGs). Altogether, they were represented by 15,715 NCBI-Assembly genomes (516–8666 range, mean = 2619; ~235 M proteins in total) (Figure 1), which allows for deep intra-clonal analysis. The core proteomes of the six MDR-CGs showed similar sizes (4239–4461, mean = 4349 ± 88 proteins), indicating a strong within-clone genomic redundancy. This is valuable information for detecting pre-existing mutations in epitopes that potentially may lead to vaccine escape.

Immunoprotection against *K. pneumoniae* is mostly humoral [14] and a response, for instance, likely to prevent neonatal sepsis through vaccination of the mother [34]. Searching for antibody-accessible proteins, between 97 and 116 proteins were identified as exposed, either extracellular or in the outer membrane, in the core proteomes of the six MDR-CGs (2.2–2.6% of the respective core proteomes). The aggregation of these searches for all MDR-CGs rendered 454 unique exposed proteins. This became reduced to 140 proteins when clustered at 70% identity and 70% alignment length cutoffs. With these thresholds, proteins within the same cluster were assumed to maintain most epitope residues and domain architectures over more relaxed conditions without significantly affecting the number of representative proteins. More stringent clustering conditions (50% or less identity) only moderately reduced the number of cluster representatives, while more relaxed clustering conditions (e.g., identity = 90%) may divide potential antigens into sub-antigens essentially sharing the same epitopes. In all, 140 non-redundant exposed proteins from at least one MDR-CG core proteome were susceptible to being humoral antigens.

| CG: | CG15 | CG37 | CG101 | CG147 | CG258 | CG307 |
|---|---|---|---|---|---|---|
| STs: | ST14 / ST15 | ST37 | ST101 | ST147 | ST512 / ST258 / ST11 | ST307 |
| Number of genome samples: | 2476 | 516 | 721 | 1642 | 8666 | 1694 |
| **Number of proteins** | | | | | | |
| Panproteome | 7.9M | 1.5M | 1.9M | 6.3M | 27.6M | 6.9M |
| Non-redundant panproteome | 23.3K | 19.8K | 13.5K | 17.4K | 29.2K | 15.7K |
| Core proteome | 4239 | 4315 | 4339 | 4448 | 4311 | 4461 |
| Core extracellular or outer membrane proteome: | 106 | 106 | 114 | 116 | 97 | 116 |

**454** unique exposed proteins

↓ *Clustering*

**140** non-redundant exposed proteins

↓ *B-cell epitope prediction*

**137** humoral antigens

↓ *Epitope conservation*

**133** humoral antigens with conserved epitopes

**Figure 1.** Sizes of genome datasets, pan-proteomes, core proteomes, and exposed proteins in *K. pneumoniae* MDR-CGs. Circles are proportional to the number of genomes and were expressed as sector diagrams when more than one ST was involved to denote the proportion of the genome contribution of each ST to the MDR-CG.

*2.3. Identification and Intraclonal Conservation of B-Cell Epitopes*

The potential capacity of the 140 MDR-CG core exposed proteins described above to elicit humoral prophylaxis was evaluated. For that, continuous B-cell epitopes were predicted with BepiPred 3.0, a transfer learning algorithm from natural processing language. Long-recognized developers of this predictor claimed it was able to capture even discontinuous epitopic traits [24]. The total of 137 (97.9%) MDR-CG core exposed proteins carried epitope spans of at least five residues. This 5mer cutoff was selected as linear runs of five or more residues commonly correspond to complete continuous epitopes or the core section of discontinuous epitopes [35,36]. In total, these 137 antigens accumulated 770 B-cell epitopic zones.

To assess antigenic drift, sequence conservation of these epitopic zones was calculated. This was carried out considering all sequenced samples available of each MDR-CG separately and at the highest per-residue resolution. Although antibody binding shows a range of tolerance to mutations in their respective epitopes [37,38], even minority mutations may anticipate mid-term replacement of present dominant isolates by mutants and subsequent evasion of the immune response elicited by these antigens. Thus, we adopted a conservative approach in which any epitope mutation can invalidate the protection induced by the epitope or reduce its size below the length threshold, as this assumption may favor the selection of broad antibodies. Therefore, a stringent residue conservation threshold of ≥99.5% isolates per MDR-CG was established to permit only occasional sequencing mistakes or transient mutational noise. This step reduced the antigen pool by 2.9% (133 antigens:

40 extracellular and 93 outer membrane proteins) (Table S1) and the total number of epitopic zones by 3.4% (744 epitopic zones, Table S2). Overall, a number of broad epitopes were found, those with sequences extremely conserved in a huge genomic representation of highly resistant clones, which are appropriate for vaccine design.

As clonal coverage greatly influences their clinical applicability, antigens were classified according to the number of core MDR-CGs in which they were present. This value was assessed using the sum of MDR-CGs that carried any conserved epitope of the antigen. Under our conditions, 76 antigens (57.1% of the total) were represented in the six analyzed lineages and, therefore, were deemed global MDR-CG-global antigens (Figure 2a). At the other end, 19 antigens (14.3%) belonged only to the core of one CG and were termed specific MDR-CG antigens. These ranged from no (CG37) to six specific antigens (CG101). Finally, the remaining 38 antigens (28.6%) were represented between two to five CGs and, thus, were called intermediate MDR-CG antigens. Remarkably, the global MDR-CG antigen pool was enriched in outer membrane proteins (81.2%) with respect to non-global (intermediate or specific) ones (54.4%) ($p = 0.0005$, Fisher exact test), which proportionally contained more extracellular antigens. This indicates antigen location is associated with the degree of clonal coverage.
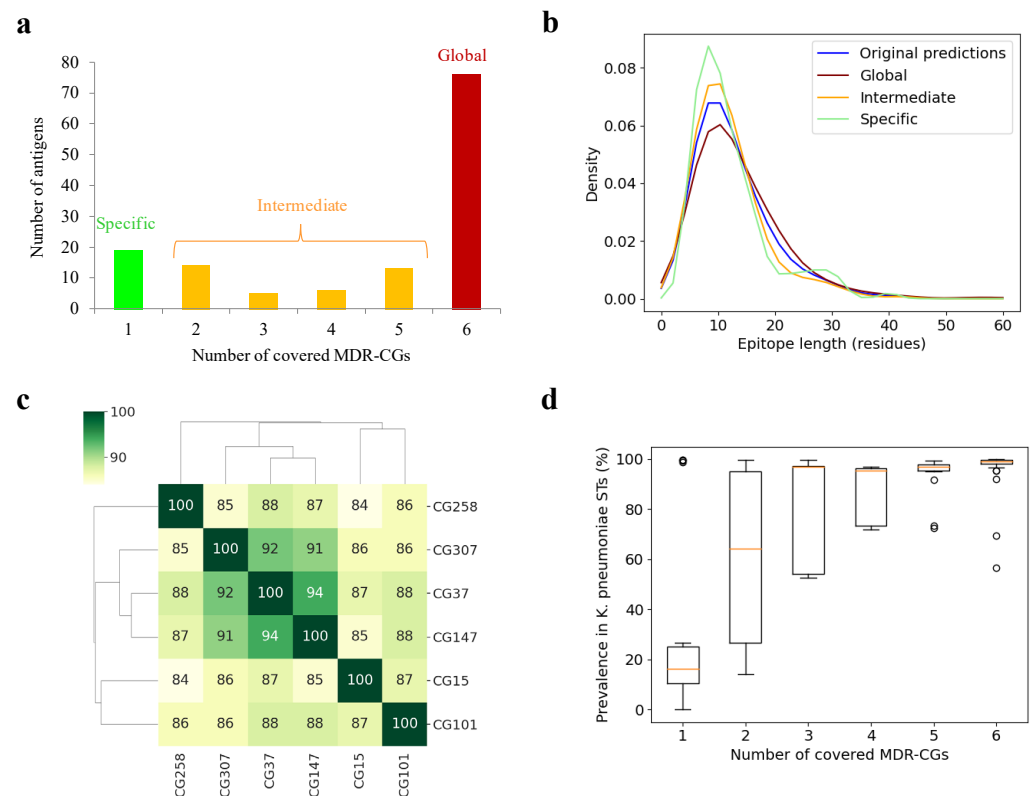


**Figure 2.** MDR-CG antigens and epitopes. (**a**) Distribution of antigens according to MDR-CG coverage. (**b**) Length distribution of epitopes, including raw predictions by BepiPred3 and those conserved in MDR-CGs. (**c**) Dendrogram and heatmap showing antigen co-existence between all MDR-CG pairs. The percentage of antigenic matching is color-ranked and indicated within the cells. (**d**) Boxplot showing the ST *K. pneumoniae* spread of MDR-CG antigens binned by MDR-CG prevalence. Orange dash indicates the median value. Box limits indicate the interquartile range. Whiskers were adjusted to maximal and minimal values if lower than 1.5 times the IQR. Further outliers are indicated as circles.

The epitope length was analyzed in these three coverage bins since, as a result of the application of the sequence conservation cutoff, epitope trimming, splitting, and removal were observed. The raw epitopic zones predicted by BepiPred-3.0 showed an average length

of 13.6 ± 11.1 residues, with the bulk of epitopes between 5 and 15 residues (Figure 2b). However, while broad epitopes in intermediate (12.8 ± 10.3 residues) and specific antigens (12.5 ± 6.8 residues) were shorter than the original BepiPred-3.0 predictions, these were longer for global (15.0 ± 14.2 residues) antigen epitopes. Thus, initial epitopes were prone to be split for specific and intermediate antigens, while some epitopes fell below the five residue cutoff and were removed for global antigens. However, this effect was not strong enough to reach statistical significance and did not noticeably alter the percentage of epitopic residues in the whole antigens.

Next, antigen pools of distinct MDR-CGs were compared. This assessment relied on intermediate MDR-CG antigens and may be relevant for the design of high-coverage multivalent vaccines with a safe relationship with the gut microbiota. The antigenic similarity, defined as the percentage of matching antigens, was calculated for all MDR-CG pair combinations (Figure 2c). Remarkably, CG37, CG147, and CG307 shared 91–94% antigens, which suggests these three lineages may be classified as a vaccitype. In contrast, CG15, CG101, and CG258 were more dissimilar, and their prophylactic control may have benefited by clone-specific initiatives.

It should be mentioned that the fact that antigens were classified as MDR-CG intermediate or specific does not necessarily ascertain the protein is fully absent in other MDR-CGs or other *K. pneumoniae* lineages. Some antigens may be widely present in the species but may not satisfy the stringent pharmacological criteria applied here, for clonal membership and epitope conservation, to qualify them for complete CG protection. Indeed, the distribution of antigens within a set of 512 ST-representative *K. pneumoniae* isolates roughly correlated with the number of MDR-CGs represented (Figure 2d, see Section 3). Antigens covering five or six MDR-CGs were present in nearly all *K. pneumoniae* STs. However, the median presence in this *K. pneumoniae* ST-representative pool only decreased below 70% for antigens covering just one or two MDR-CGs, suggesting that those are the ones most adequate for formulations that do not affect commensal *Klebsiellae*.

In all, 133 proteins harboring humoral epitopes with no or negligible sequence drift were detected in the core proteomes of one to six MDR-CGs. Globally distributed antigens are appropriate for universal MDR-CG vaccines, while non-global ones contributed to the definition of MDR-CG vaccitypes and may facilitate precise vaccine interventions.

*2.4. Potential Cross-Reactions with Host and Human Gut Microbiota*

The existence of close immunological homologs of MDR-CG antigens in humans or in their gut microbiota may produce undesirable side effects if included in vaccines. Only five MDR-CG antigens shared significant similarity (E < 0.00001, 21–34% identity over 52–84% antigen length) with human proteins. Moreover, these antigens did not share any identical epitope tract of ≥5 residues with their respective human counterparts. This indicates potential antibody cross-recognition to human proteome by MDR-CG antigen responses is not negligible but, in general, low.

A similar analysis was conducted with the human gut microbiota, which may be altered by cross-reactivity, causing health problems. In contrast to the human proteome, nearly all MDR-CG antigens (125, 94.0%) showed one or more homologs in gut microbiota isolates stored in the general CGR resource, at least when relaxed thresholds (E < 0.00001, ≥50% antigen length) were applied (Figure 3a). Among these, homologs from *Proteobacteria* isolates (*Enterobacter*, *Escherichia*, and *Citrobacter* species) dominated in terms of number and identity level. Nevertheless, 50 antigens (37.6%) still had homologs in non-*Proteobacteria* phyla, and 31 of them (23.3%) even in gram-positive phyla (*Actinobacteria* and *Firmicutes*). Up to 85 MDR-CG *K. pneumoniae* antigens (63.9%) still had gut microbiota homologs sharing 60% identity or more with CGR isolates, which are those that, in practice, may elicit cross-reactive antibodies. These antigens were relatively enriched in global MDR-CGs (76.5% of the total, *p* < 0.0001, Fisher exact test) (Figure 3b) compared to intermediate or specific ones. Microbiota homologs over 60% identity were only detected in *Proteobacteria* (Figure 3a), except the lipoprotein Blc also present in *Bacteroidetes*.
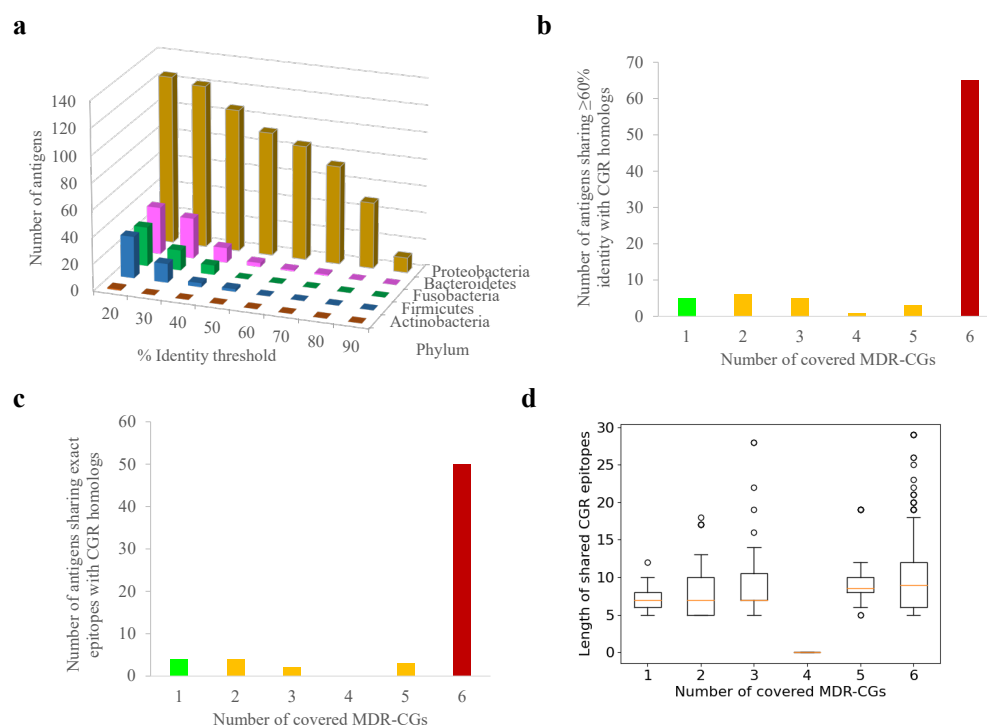
**Figure 3.** Potential cross-reactivity between *K. pneumoniae* MDR-CG antigens and human gut microbiota proteomes. (**a**) Number of antigens showing homologs with gut microbiota isolates at different sequence identity cutoffs. (**b**) Number of antigens showing homologs with gut microbiota isolates above 60% identity according to MDR-CG coverage. Bars for global, intermediate, and specific MDR-CG antigens are colored in red, orange, and green, respectively. (**c**) Number of antigens sharing epitopes with five residues or with homologs in gut microbiota isolates according to MDR-CG coverage. Bars for global, intermediate, and specific MDR-CG antigens are colored in red, orange, and green, respectively. (**d**) Boxplot showing length distribution of MDR-CG epitopes shared with human gut microbiota. Values were organized by the number of MDR-CGs represented by the epitope. Orange dash indicates the median value. Box limits indicate the interquartile range. Whiskers were adjusted to maximal and minimal values if lower than 1.5 times the IQR. Further outliers are indicated as circles.

When exact 5mer epitopes or longer were searched in gut microbiota homologs, the incidence of *Proteobacteria* was exclusive. Global MDR-CG antigens shared proportionally more conserved epitopes (73.7%) than intermediate and specific antigens, including identical epitopes with 15 or more residues (76.6%) (Figure 3c). Remarkably, only 11 global MDR-CG antigens satisfied the confronted forces of covering all MDR-CGs but were not represented (less than 60% identity and no epitopes) in any of the gut microbiota isolates.

Overall, the risk of immunological cross-reactivity after vaccination with MDR-CG antigens is emphasized in relation to global antigens and *Proteobacteria*. In any case, precise cutoffs regarding the degree of homolog closeness in the gut microbiota can be implemented in RV procedures for antigen selection. This would allow us to find a balance by finely tuning sensitivity (protection) and specificity (lack of cross-reactions) when generating antigen candidate lists.

## 2.5. Identification of Antigens with High Functional Constraints

Certain functional aspects of the antigens may be associated with evolutionary constraints that reduce the chances of epitope drift, shift, or loss. Regarding general functional classification, global MDR-CG antigens were proportionally enriched in housekeeping-like functions concerning the "Cell wall/membrane/envelope biogenesis" class, while intermediate and specific antigens were more often classified within "Cell motility" and

"Intracellular trafficking and secretion" categories (Figure 4a). This functional swap may have implications when precise immunoprotection is designed.

**a**



**b**



**c**
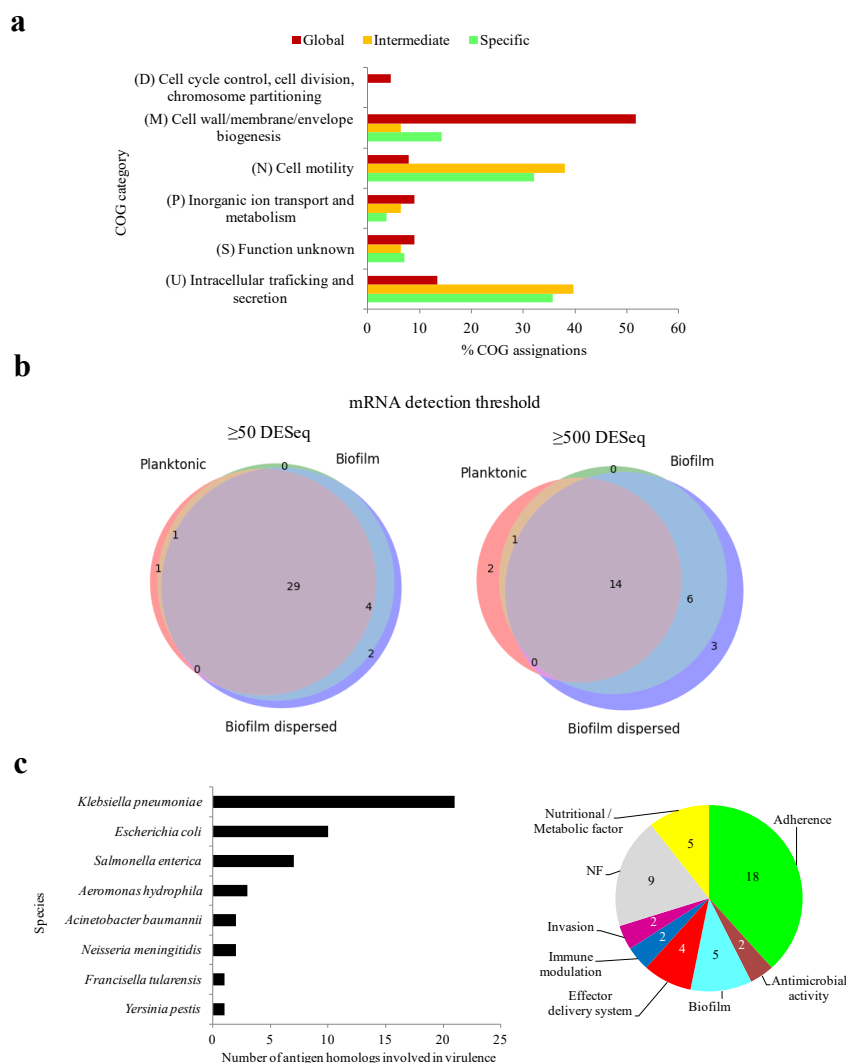


**Figure 4.** Functional constraints of MDR-CG antigens. (**a**) General functional category of antigens according to MDR-CG coverage. Only categories with at least 4% COG assignation in a MDR-CG coverage bin were included. (**b**) Weighted three-class Venn diagrams showing the number of expressed antigens during planktonic, biofilm, and biofilm dispersed stages. Low (≥50 DESeq, **left**) and high (≥500 DESeq, **right**) RNA-Seq detection thresholds were considered. The Venn diagrams were depicted using the *Venn_3* method of the *matplotlib* Python library. (**c**) Species origin of putative virulence antigens (**left**). Prevalence of VF categories for antigens (**right**).

Knowing the expression regime of an antigen is crucial for its ultimate utilization. To address this issue, transcriptomic data previously published from five growth stages, including distinct nutrient availability and environmental conditions, were utilized [39]. Equivalent proteins for 108 out of 133 antigens (81.2%) were detected in the strain CH1034, the one utilized in the original study. From them, transcription for 37 and 26 (27.8% and 19.5%) MDR-CG antigen homologs was observed in at least one condition after applying cutoffs of 50 and 500 DESeq normalized values and deemed expression and high expression, respectively. Up to 31 antigens were expressed in both planktonic conditions (exponential and stationary), 49 in the two biofilm conditions (biofilms of 7 and 13 h), and 54 during the "biofilm dispersed" stage. Expression for 29 antigens was detected during all conditions (Figure 4b, left), and 14 of them were always expressed at high counts and were, therefore, deemed constitutively highly transcribed antigens (Figure 4b,

right). The last group included 12 outer membrane proteins, of which five belonged to the "Cell wall/membrane/envelope biogenesis" and four to the "Inorganic ion transport and metabolism" COG categories.

Next, we studied whether antibodies elicited by the MDR-CG antigens may ameliorate infection through direct neutralization of virulent activities. Forty-seven antigens (35.3%) shared significant homology with virulence factors (VFs). Almost half of them, 21, were genuine *K. pneumoniae* VFs while the remaining were mostly homologs to VFs from other *Enterobacteria* pathogens such as *E. coli* and *Salmonella enterica* subsp. enterica serovar Typhimurium (Figure 4c, left). Eighteen antigens were classified within the "Adherence" category, five were in the "Nutritional/Metabolic factors" category, and five were classified in the "biofilm" virulence category (Figure 4c, right). Thus, the humoral response against one-third of the MDR-CG antigens may directly interfere with several pathogenic mechanisms of *K. pneumoniae*.

Finally, the influence of MDR-CG antigens on the *K. pneumoniae* network of protein-protein interactions (PPIs), the interactome, was analyzed since the types and extent of the connections of a protein correlate with its global relevance on cellular physiology [40]. For that, we applied the principles of network theory, by which proteins are converted into nodes, and their interactions are taken as edges. Equivalent proteins for 101 out of the 133 antigens (75.9%) were identified in HS11286, the *K. pneumoniae* reference strain of the central interactome resource STRING. From those, 90 antigens showed interactions with other *K. pneumoniae* proteins, ranging from 1 to 77 PPIs (Figure 5a). In addition, 46 antigens showed first-rank PPIs to at least one virulence factor (VF-PPIs), suggesting potential indirect neutralization influence of their elicited antibodies. For VF partners interacting with antigens, "Adherence" was also the category that was more prominent (Figure 5b). Therefore, it is possible that the neutralization of this pathogenesis stage through vaccination could be further amplified through functional contacts of antigens with adhesive proteins.
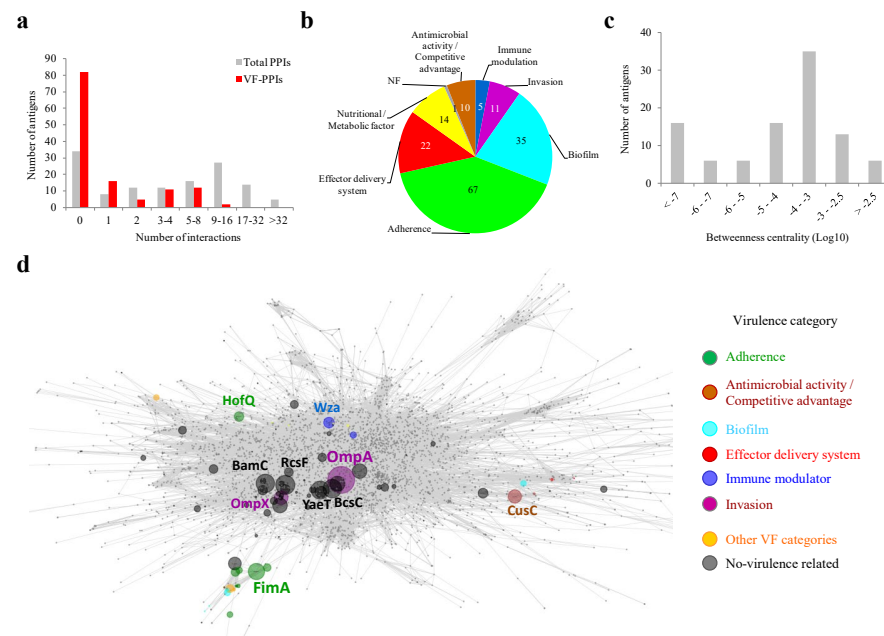


**Figure 5.** Interactomic properties of MDR-CG antigens. (**a**) Antigen distribution according to the total number of PPIs and VF-PPIs. (**b**) Prevalence of VF categories for first-rank interactions of antigens. The cumulative number of PPIs, i.e., one protein can interact with several antigen partners, is shown. (**c**) Antigen distribution according to BC values. (**d**) Topology of the *K. pneumoniae* interacting network. Antigens are colored by virulence categories if applicable. The sphere diameter of antigen nodes is proportional to BC value. Antigen nodes are name-labeled when above a BC cutoff of 0.3% or 0.1% if they are VFs. Non-antigen nodes are shown as small dots in gray. Edges are shown in gray.

Among the principal virulence hub antigens, two Hcp1 homologs from Type VI secretion systems excelled with 11 and 10 VF-PPIs with "Effector delivery system" proteins. Furthermore, two FimD fimbrial proteins exhibited eight interactions, each involving biofilm and adherence factors. An interactome metric of protein importance over the raw PPI count is BC (see Section 3). BC values of antigens followed a rather normal distribution, with a peak between $10^{-3}$ and $10^{-4}$ (Figure 5c). Six antigens showed a betweenness centrality over 0.3% shortest paths (labeled with protein name in the network graph depicted in Figure 5d), including the invasion-related protein OmpA (BC = 0.96%). Thus, there are MDR-CG antigens that show an attractive positioning in the topology network for magnified neutralization of *K. pneumoniae*.

## 2.6. Selection for Global and Constitutive Expressed Antigens for Univalent Vaccines

Despite the evident properties of antigen prioritization, no unique selection scheme can be considered. To dynamically check different selection scenarios, all data generated above were normalized as a single framework. In a first approximation, we studied proteins separately for univalent vaccines. For this, we focused on the clonal scope and conserved epitopicity over the six principal MDR lineages, as well as constitutive expression. Moreover, we established these antigens should show modest conflict with the gut microbiota, i.e., only *Proteobacteria* homologs at most, as these anaerobic, facultative species account for less than 0.1% of bacteria in the oxygen-deprived colon environment [41]. Out of the initial 76 global MDR-CG proteins, only 13 satisfied all these criteria, including several porins and fimbriae. Importantly, this list was further reduced to 11 antigens, as shown in Figure 6, because the BamD and BamE outer membrane assembly proteins were actually oriented inwards of the cell and, thus, were not accessible to antibodies [42]. This indicates that, in conjunction with computational methods, exhaustive manual analyses by human experts are ultimately required to evaluate RV outcomes prior to labor-intensive experimental validation. Notably, vaccination with any of the three extracellular proteins in this subset (EcpD, FimA, and FimG) would predictably elicit virulence neutralization. Likewise, four out of the 11 proteins showed BC values above 0.01% of the shortest paths. Antigens in this list essentially showed attractive experimental predicted solubility, non-toxic, and non-allergenic properties (Table S3). Altogether, these 11 antigens are attractive candidates for univalent approaches for either protein or nucleic acid vaccines.

| Protein | Description | Number of epitope residues | Expression (log2 DESeq values) Planktonic Exp. | Sta. | Biofilm 7h | 13h | Disp. | Interactomic centrality Deg. | Bet. | Host cross-reaction Human | Gut microbiome Phyla | Genera | Isolates | Virulence Class | 1st order interactions |
|---------|-------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| **RafY** | Carbohydrate porin | 151 | 5 | 8 | 6 | 7 | 4 | 6 | 0.0329 | 0 | 1 | 3 | 4 | - | - |
| **Fiu** | Catecholate siderophore receptor | 149 | 15 | 6 | 6 | 7 | 9 | 6 | 0.0002 | 0 | 1 | 3 | 5 | - | - |
| **ChiP** | Chitoporin | 132 | 9 | 10 | 8 | 11 | 8 | 4 | 0 | 0 | 1 | 3 | 5 | - | - |
| **EcpD** | Fimbrial adhesine | 117 | 5 | 6 | 7 | 7 | 8 | 6 | 0 | 0 | 1 | 3 | 4 | Adherence | - |
| **LamB** | Maltoporin | 113 | 8 | 11 | 11 | 14 | 12 | 18 | 0.0993 | 0 | 1 | 3 | 5 | - | - |
| **Tsx** | Nucleoside-specific channel-forming protein | 100 | 12 | 10 | 13 | 12 | 12 | 4 | 0.001 | 0 | 1 | 3 | 5 | - | Immune modulation (1) |
| **FimG** | Type 1 fimbrial protein | 51 | 6 | 5 | 7 | 6 | 8 | 16 | 0.0753 | 0 | 1 | 2 | 4 | Adherence | Adherence (4) |
| **OmpW** | Outer membrane protein W | 49 | 6 | 15 | 10 | 11 | 11 | 2 | 0 | 0 | 1 | 3 | 5 | - | |
| **FimA** | Type 1 fimbrial major subunit | 34 | 12 | 11 | 14 | 13 | 13 | 30 | 0.0051 | 0 | 1 | 3 | 5 | Adherence | Adherence (8) |
| **PagP** | Lipid A palmoyltransferase | 28 | 13 | 11 | 11 | 10 | 12 | 10 | 0.0583 | 0 | 1 | 5 | 7 | - | Invasion (1) |
| **PldB** | Lysophopholipase | 21 | 8 | 10 | 11 | 11 | 10 | 16 | 0.0005 | 0 | 1 | 2 | 4 | - | - |

**Figure 6.** Antigens selected for univalent vaccines against *K. pneumoniae* MDR-CGs. Antigens were inversely sorted by the raw number of conserved B-cell epitope residues. Values for filtering criteria were color-ranked where the intensity was proportional to the positive condition. The number of first-order interactive partners grouped by virulence class is provided in brackets.

### 2.7. Antigen Selection for Multivalent MDR K. pneumoniae Vaccines

Multivalent vaccines achieve high efficacy by limiting potential escape and low coverage, often shown by univalent formulations [43]. Therefore, we studied different rational antigen combinations, taking advantage of the harmonized data described above. Bivalent to pentavalent options were obtained based on the selection reasoning.

We designed a first approach consisting of highly epitopic proteins (>100 B-cell epitope residues). Moreover, among these antigens, those showing high expression certainty in at least one infection stage were chosen. This was based on the assumption that to reach optimal results with nosocomial pathogens, it is necessary to combine antigens expressed during planktonic and biofilm stages [44]. We established the microbiota cross-reactivity and BC filters to two phyla/seven genera and $10^{-6}$ shortest paths, respectively. These criteria resulted in a trivalent combination of the "catecholate siderophore receptor" Fiu (high "exponential planktonic" transcription), the "maltose and maltodextrins transporter" LamB (high "stationary planktonic" and "late biofilm" expression counts) and the "nucleoside-specific channel-forming protein" Tsx (high "biofilm expression" levels). Remarkably, these three proteins are involved in the acquisition of different key nutrients, which may be simultaneously undermined by the humoral protection elicited.

Next, direct neutralization of pathogenicity was analyzed as antibodies against VFs may behave as an "anti-virulence" defense [45]. In this context, two antigen selection paradigms are possible for the VF antigens utilized belonging to a single or several virulence categories. The aims are magnifying the focus on an event or avoiding functional redundancy, respectively. An illustrative example is adhesion, the most predominant activity to be neutralized through vaccination. However, the fact that numerous antigens exert this function, clones that eventually delete adherent-focus antigens to escape the vaccine may remain infective. Nevertheless, a mixture including the tip subunits of Fim and LpfC fimbria, as well as the PilQ porin, may be considered for preliminary assays. Virulence categories less prone to functional overlapping were "biofilm formation" and "Invasion", which are the other two critical phases during *K. pneumoniae* infection. A bivalent option that may elicit immunological interference of biofilm formation consist of the efflux pump AdeH and MrkC, the latter being the most attractive subunit of the Mrk fimbria by epitopicity, expressability, and network centrality criteria. Invasion may be hindered by immunization with the OmpA porin and the Ail evasin.

Finally, a systems biology approach was conducted to maximize neutralization through the synergy of non-related interactomic hubs. The 98 antigens with predicted PPIs, after excluding BamCDE, accounted for a total of 1025 PPIs involving 491 unique protein partners. When the MDR-CG antigens were sorted greatest-to-least by the number of PPIs, a plateau was reached much earlier when only the non-redundant interacting proteins were considered (Figure 7). In fact, a collection of just five antigens, namely OmpA, OmpX, BcsC, BamA, and Hcp amounted to 41% of all unique partners. Thus, antibody-driven neutralization through this pentavalent formulation may be amplified through functional interactions to produce a vast disturbance in the protein network architecture and, subsequently, in the cell physiology.
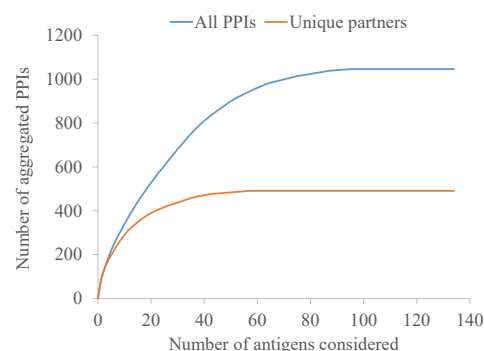


**Figure 7.** Aggregated number of all PPIs and those with non-redundant partners of antigens. Antigens were reversely sorted on the x-axis by their contribution to the interactome.

Altogether, rational antigen combinations concerning high epitopicity, virulent categories, and interactomic warrant experimental validation to gain insights into the most successful strategies in practice.

### 2.8. The Combination of Intermediate MDR-CG Antigens

Another promising combinative approach to produce effective vaccines relies on the fine selection of complementary antigens with intermediate MDR-CG coverage. The advantage of doing this is that cross-reactive side effects may be reduced while maintaining protection against highly resistant clones. Out of the 38 intermediate MDR-CG antigens, 8 (6 were extracellular, and 5 were VFs) were not represented in microbiota phyla other than *Proteobacteria* and in ≤60% of *K. pneumoniae* ST representatives (Figure 8). These antigens involved four homologs of FimA fimbrial protein, two Type VI secretion system tube TssD proteins, one usher fimbrial protein, and one OprD porin. Despite the obvious relationship of homologs belonging to either FimA or TssD families, they only shared moderate identities of ≤30% and 61% with other members of the same family, respectively, which further decreased to ≤14% and 50% identities in the epitopic regions. This supports intermediate MDR-CG antigens with comparable functions that are immunologically independent and adequate for co-selection in the same formulation. Moreover, wide MDR-CG protection may be achieved by adding antigens that cover combos of two, three, or four lineages while maintaining microbiota safety profiles. This combinatorial plan was congruent with the vaccitype paradigm depicted above (Figure 2c), as the CG147-CG307 clonal pair shared up to four of these antigens. Moreover, these two CGs were abundantly considered by this type of strategy (seven and five intermediate antigens, respectively). In contrast, CG101 only had one of these intermediate antigens, which likely required the addition of CG101-specific ones (see below) to obtain sufficient clonal coverage. As a result of this analysis, a trivalent intermediate antigen vaccine with a fimbrial FimA homolog (WP_045327861.1), a TssD homolog (WP_002904614.1), and the only porin (OprD), would cover the six MDR-CGs with high specificity.

| Family | Antigen | Lineage | | | | | | Number of epitopic residues | *% K. pneumoniae* ST coverage |
|---|---|---|---|---|---|---|---|---|---|
| | | CG15 | CG37 | CG101 | CG147 | CG258 | CG307 | | |
| Fimbrial FimA protein | WP_004177744.1 | | ■ | | ■ | | | 79 | 36.52 |
| | WP_045327861.1 | | | | ■ | | | 60 | 29.88 |
| | WP_004213865.1 | | | | ■ | | ■ | 43 | 14.84 |
| | WP_004213863.1 | | | | ■ | | | 5 | 14.06 |
| Outer membrane OprD porin | WP_153932497.1 | ■ | | | ■ | | | 90 | 54.30 |
| Outer membrane usher FimD protein | HBU3631830.1 | | | | ■ | | | 93 | 14.65 |
| Type VI secretion system tube protein TssD | WP_002904614.1 | ■ | | | | ■ | | 20 | 52.54 |
| | WP_032433792.1 | | | | ■ | ■ | | 42 | 25.59 |

**Figure 8.** Intermediate MDR-CG antigens showing low gut microbiota and low general *K. pneumoniae* ST predicted cross-reactivities. Representation (filled cells) and non-representation (empty cells) of the MDR-CGs by the antigen are indicated.

### 2.9. The Utilization of Specific MDR-CG Antigens

Finally, another precise vaccinable strategy consists of the utilization of specific MDR-CG antigens, which may even further reduce the interference with microbiota and commensal *K. pneumoniae* isolates. Out of the 19 MDR-CG-specific antigens, 15 showed extraordinary microbiota safety profiles. In some cases, homologs were detected in ≤1 microbiota *Proteobacteria* isolates and were present in <30% *K. pneumoniae* representative STs (Figure 9). The 15 reliably specific antigens showed epitopicities ranging from 39 to 163 MDR-CG conserved residues, suggesting a sufficient antigenic capacity. All CGs but CG37 had at least one CG-specific antigen, while CG101 accounted for six.

| Family | Antigen | CG15 | CG37 | CG101 | CG147 | CG258 | CG307 | Number of epitopic residues | % *K. pneumoniae* ST coverage |
|---|---|:--:|:--:|:--:|:--:|:--:|:--:|:--:|:--:|
| | | | | Lineage | | | | | |
| Fimbrial FimA protein | WP_004183089.1 | | | ■ | | | | 46 | 23.05 |
| | WP_004183106.1 | | | ■ | | | | 60 | 21.09 |
| | WP_004181296.1 | ■ | | | | | | 72 | 12.89 |
| | WP_032432059.1 | | | | ■ | | | 35 | 5.66 |
| | WP_240184712.1 | ■ | | | | | | 60 | 9.38 |
| Hypothetical protein | WP_032419366.1 | | | ■ | | | | 140 | 16.21 |
| Outer membrane beta-barrel protein | WP_023316483.1 | | | | | | ■ | 137 | 5.66 |
| Outer membrane usher FimD protein | WP_077255459.1 | | | | ■ | | | 60 | 5.66 |
| | WP_004181289.1 | ■ | | | | | | 44 | 13.67 |
| | WP_223175056.1 | | | ■ | | | | 44 | 21.48 |
| | WP_032419371.1 | | | | | | ■ | 82 | 23.63 |
| TolC family outer membrane protein | WP_004179227.1 | | | | | | ■ | 39 | 26.76 |
| TonB-dependent receptor | HBT7858981.1 | | | | | ■ | | 167 | 0.20 |
| | WP_023158563.1 | | | | ■ | | | 97 | 11.33 |
| Winged HTH domain CadC1 protein | WP_004182815.1 | | | | | | ■ | 65 | 15.04 |

**Figure 9.** Specific MDR-CG antigens showing low gut microbiota and low general *K. pneumoniae* ST predicted cross-reactivity. Representation (filled cells) and non-representation (empty cells) of the MDR-CGs by the antigen are indicated.

Functional description for nearly all of these antigens was obtained, which is important to rationalize the mode-of-action of elicited antibodies, i.e., neutralization, over empirical protection. Remarkably, nine antigens from four MDR-CGs were annotated as fimbriae subunits, which, like for intermediate MDR-CG fimbria, had low chances of inducing cross-protection. Seven of these fimbrial antigens were phylogenetically closer to adherent virulence factors from isolates of other enterobacterial pathogens. This suggests these particular antigens were acquired by lateral transfer at *K. pneumoniae* post-speciation stages and maintained through improving the colonization abilities of the resulting clone. Thus, tissue-specific adhesion of CG15, CG101, and CG147 may be approached with high preciseness inducing neutralizing IgA antibodies against their hallmark fimbriae.

Despite reasonable epitopicity of intermediate (50.0 ± 28.5, average ± SD, epitopic residues) and specific (77.6 ± 39.5) MDR-CG antigens, an obvious disadvantage for these approaches is no extensive information, e.g., expression or interacting data, can be inferred for most of these proteins due to lack of representation in working strains. Furthermore, a prohibitive number of antigens must be utilized to achieve full MDR-CG protection. Nevertheless, they may still be considered to complement the intermediate MDR-CG antigen combinations. Moreover, epitopic zones of specific antigens may be selected for peptide-based vaccines or for the construction of hybrid antigens based on loop grafting into scaffold proteins [46]. Importantly, one intermediate and one specific antigen (Type VI secretion system tube TssD protein and TonB-dependent receptor homologs, respectively) were identified for CG258, which is still the most challenging high-resistant clone worldwide. Compared with global MDR-CG antigens, the use of specific MDR-CG antigens embodies the ideal of "magic bullet" vaccination that mostly circumvents cross-reactivity with health-related commensal *Proteobacteria* strains as those concerning K vitamin production and prevention of colonization by pathogens [47]. This strategy may be useful for hospitals with recurrent infections by the same endemic clone [48,49], in particular for patients prone to microbiota alterations.

## 3. Materials and Methods

### 3.1. Sequence Acquisition and Clustering

Non-redundant complete proteomes of *K. pneumoniae* from the Genbank and RefSeq databases were acquired from the NCBI-Assembly resource [50]. Samples that were non-annotated, partial, anomalous, and those that did not pass the taxonomy check were excluded. Isolates were assigned to ST by applying BLAST 2.2.26, with strict 100% identity and 100% alignment coverage, on whole genomes using sequence types (STs) gene allelic

combinations and sequences stored at the MLST database (https://pubmlst.org/) (accessed on 27 March 2023) [51]. Proteins were clustered with CD-HIT 4.6 [52] at 70% identity and 70% alignment length. Proteins showing homologs over these thresholds in ≥95% isolates from a given clonal group (CG) by BLAST were deemed core proteins of that CG.

### 3.2. Sequence Analyses

Among all CG core proteins, those exposed (extracellular or outer membrane cellular locations) were identified using PSORTb 3.0.3 by selecting the "Gram-negative" option [53]. Exposed core proteins from any CG of interest were pooled and subjected to further clustering with CD-HIT applying a 70% identity and 70% alignment length. B-cell linear epitopes were predicted with BepiPred 3.0 [24], applying the default threshold for the rolling mean score. Predicted B-cell epitope spans with lengths ≥ 5 positions were subjected to residue-level conservation analysis within the CG with Clustal Omega 1.2.4 [54]. Only those epitope regions with five or more residues that were strictly conserved in ≥99.5% isolates within the CG were considered broad epitopes and considered for further analyses. To assign the sequence limits of broad epitopes, the highest number of CGs was prioritized over epitope length even when its sequence was reduced by trimming or splitting due to non-conserved positions in any of the CGs, providing epitopes were five residues or longer. For this process, an in-house Python script was programmed. The final MDR-CG coverage of an antigen was defined as the total number of MDR-CGs in which at least one broad epitope of the antigen is present.

Membership in the *K. pneumoniae* core proteome sensu lato (https://www.cgmlst.org/ncs/schema/Kpneumoniae865/) (accessed on 4 August 2023) was determined with MMSeqs2 v.13-45111+ds-2 applying 70% identity and 70% alignment length thresholds. The global occurrence of antigens in the whole *K. pneumoniae* species was estimated by selecting representative samples of 512 STs with five or more sequenced samples in NCBI-Assembly (Table S4) using MMSeqs2 [55] applying 70% identity and 70% alignment length coverage as cutoffs.

Functional categories according to the Cluster of Orthologous Groups (COGs) scheme were assigned with eggNOGmapper v2 [56] against the eggNOG 5 database [57] using default parameters. Protein domains in the Pfam v33.1 database [58] were detected with hmmscan of HMMER 3.1b1 [59] applying score gathering thresholds.

### 3.3. Host and Human Gut Microbiota Homologs

For the detection of antigen homologs in the human proteome, the reviewed *Homo sapiens* SwissProt proteome downloaded from UniProt was used (accessed 29 March 2023). The global presence of antigen homologs in the gut microbiota was assessed using representative genomes covering culturable gut microbiota stored in the Culturable Genome Reference (CGR) archive [60]. CGR contains 306 bacterial isolates stratified into 176 species, 50 genera, and 5 phyla. For both human and CGR proteomes, antigen sequence searches were carried out with MMSeqs2 [55], applying different identity, alignment coverage, or E-value cutoffs. Exact *K. pneumoniae* epitope sequences with five or more residues present in CGR and human proteomes were identified with an in-house Python 3.10.12 script. The only *K. pneumoniae* representative isolates in the CGR was not considered in gut microbiota analyses.

### 3.4. Expression, Virulence, and Interactome Data

Normalized DESeq transcriptomic expression data from exponential (ExPlan) and stationary (StatPlan) planktonic cells, cells after 7 h (7hBiof) and 13 h (13hBiof) biofilm growth, and cells dispersed from biofilms (DispBiof) were acquired from previously published data [39]. Antigen homologs in the original strain of this study, CH1034 (NCBI-Assembly id: GCF_001404095), were identified by search with MMSeq2 applying thresholds of 70% identity and 70% alignment length. Different cutoffs for normalized DESeq values were applied to consider expression.

To detect putative proteins involved in virulence, antigens were compared with the set A content of the virulence factor database (VFDB) using MMSeq2 and applying thresholds of 30% identity and 60% alignment length.

For interactomic analyses, homologs for antigens in the *K. pneumoniae* representative strain available in STRING v11 [61], HS11286, were identified by search with MMSeq2, applying thresholds of 70% identity and 70% alignment coverage. Then, pre-calculated edges showing a combined score of 0.7 or higher ("Confident level") were selected. Betweenness centrality (BC) was calculated based on these data using the betweenness_centrality method of the NetworkX 3.0 Python package [62]. BC is defined as the fraction of shortest paths between all node pairs in which an antigen node is included.

### 3.5. Calculation of Antigen Workability, Allergenicity and Toxicity

Solubility of overexpressed proteins in *Escherichia coli* was assessed with SoluProt 1.0 [63], toxicity with the hybrid model of ToxinPred2 [64], allergenicity with AllerCatPro 2.0 [65] and transmembrane helices with Phobius [66]. All these predictions were carried out after removing the signal peptide when detected with SignalP 6.0 using the non-eukaryotic 'other' model [67].

### 4. Conclusions

In this study, we have applied several antigen selection criteria not considered previously for K. pneumoniae vaccination. For instance, the conservation assessment of epitope sequences using massive genomic data per CG allowed us to enter immunoinformatics into the deep intraclonal age. Consequently, this allowed us to detect pre-existing minority subclones with mutated epitopes that may eventually lead to vaccine escape. Likewise, antigen- and epitope-level conservation analyses in a comprehensive list of gut microbiota taxons rendered a more precise estimation of potential undesirable cross-reactions. In contrast to previous *K. pneumoniae* RV reports, essentiality was not applied here because, unlike antibacterial targets, many vaccine antigens are only indispensable for optimal fitness. Our analyses permitted a most confident estimation of antigen epitopicity and the protective clonal scope for each antigen and revealed a potential MDR vaccitype that included CG37, CG147, and CG307 clades.

Humoral antigens showed a range of clonal coverage, microbiota homologs, number of epitope residues, expression conditions, functions, and interactions. A number of them were directly or indirectly related to adhesion during infection. Of the 133 antigens represented in the MDR lineages, only 11 satisfied most of the general criteria and were deemed promising for univalent vaccines. Remarkably, manual inspection revealed some of the original antigens were very likely not accessible to antibodies, which suggests RV outcomes must be carefully revised by experts. Furthermore, several strategies were utilized concerning clonal coverage, expression, virulence, and positioning in the protein network. For instance, as few as five antigens accumulated 41% of all non-redundant partners in the interactome of conserved-epitope antigens. The reasoned utilization of antigens with middle and exclusive clonal MDR coverage increased the preciseness. Notably, among other functionalities, many promising antigens consisted of porins and fimbriae.

In sum, the insights obtained here provide important clues that will facilitate the design of improved vaccines against this primary human pathogen, in particular for synergic reasoned multivalent and low vaccine-escape approaches.

**Author Contributions:** Conceptualization, E.G., M.P.-V., J.O.-I., M.J.M., A.J.M.-G. and A.P.; methodology, A.T. and A.J.M.-G.; software: A.J.M.-G.; formal analysis, A.T. and A.J.M.-G.; data curation, A.T., E.G. and A.J.M.-G.; writing—original draft preparation, A.T. and A.J.M.-G.; writing—review and editing, M.P.-V., J.O.-I., M.J.M., A.J.M.-G. and A.P.; funding acquisition, M.P.-V., J.O.-I., M.J.M. and A.J.M.-G. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All relevant data are included in the manuscript and its supporting information files.

**Conflicts of Interest:** M.J.M. is a founder and shareholder in the company Vaxdyn, S.L. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The remaining authors declare no relevant financial or non-financial interests to disclose.

## References

1. Bloom, D.E.; Black, S.; Salisbury, D.; Rappuoli, R. Antimicrobial Resistance and the Role of Vaccines. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12868–12871. [CrossRef] [PubMed]
2. Tobuse, A.J.; Ang, C.W.; Yeong, K.Y. Modern Vaccine Development via Reverse Vaccinology to Combat Antimicrobial Resistance. *Life Sci.* **2022**, *302*, 120660. [CrossRef] [PubMed]
3. Wang, G.; Zhao, G.; Chao, X.; Xie, L.; Wang, H. The Characteristic of Virulence, Biofilm and Antibiotic Resistance of Klebsiella Pneumoniae. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6278. [CrossRef]
4. Russo, T.A.; Alvarado, C.L.; Davies, C.J.; Drayer, Z.J.; Carlino-MacDonald, U.; Hutson, A.; Luo, T.L.; Martin, M.J.; Corey, B.W.; Moser, K.A.; et al. Differentiation of Hypervirulent and Classical *Klebsiella pneumoniae* with Acquired Drug Resistance. *mBio* **2024**, *15*, e0286723. [CrossRef]
5. Paczosa, M.K.; Mecsas, J. *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 629–661. [CrossRef]
6. Podschun, R.; Ullmann, U. *Klebsiella* spp. as Nosocomial Pathogens: Epidemiology, Taxonomy, Typing Methods, and Pathogenicity Factors. *Clin. Microbiol. Rev.* **1998**, *11*, 589–603. [CrossRef]
7. Lee, C.-R.; Lee, J.H.; Park, K.S.; Jeon, J.H.; Kim, Y.B.; Cha, C.-J.; Jeong, B.C.; Lee, S.H. Antimicrobial Resistance of Hypervirulent *Klebsiella pneumoniae*: Epidemiology, Hypervirulence-Associated Determinants, and Resistance Mechanisms. *Front. Cell Infect. Microbiol.* **2017**, *7*, 483. [CrossRef]
8. Martin, R.M.; Bachman, M.A. Colonization, Infection, and the Accessory Genome of *Klebsiella pneumoniae*. *Front. Cell Infect. Microbiol.* **2018**, *8*, 4. [CrossRef]
9. Wyres, K.L.; Lam, M.M.C.; Holt, K.E. Population Genomics of *Klebsiella pneumoniae*. *Nat. Rev. Microbiol.* **2020**, *18*, 344–359. [CrossRef]
10. Magill, S.S.; Edwards, J.R.; Bamberg, W.; Beldavs, Z.G.; Dumyati, G.; Kainer, M.A.; Lynfield, R.; Maloney, M.; McAllister-Hollod, L.; Nadle, J.; et al. Multistate Point-Prevalence Survey of Health Care-Associated Infections. *N. Engl. J. Med.* **2014**, *370*, 1198–1208. [CrossRef]
11. Tacconelli, E.; Carrara, E.; Savoldi, A.; Harbarth, S.; Mendelson, M.; Monnet, D.L.; Pulcini, C.; Kahlmeter, G.; Kluytmans, J.; Carmeli, Y.; et al. Discovery, Research, and Development of New Antibiotics: The WHO Priority List of Antibiotic-Resistant Bacteria and Tuberculosis. *Lancet Infect. Dis.* **2018**, *18*, 318–327. [CrossRef] [PubMed]
12. Atkins, K.E.; Lafferty, E.I.; Deeny, S.R.; Davies, N.G.; Robotham, J.V.; Jit, M. Use of Mathematical Modelling to Assess the Impact of Vaccines on Antibiotic Resistance. *Lancet Infect. Dis.* **2018**, *18*, e204–e213. [CrossRef] [PubMed]
13. Andrews, T.D.; Gojobori, T. Strong Positive Selection and Recombination Drive the Antigenic Variation of the PilE Protein of the Human Pathogen *Neisseria meningitidis*. *Genetics* **2004**, *166*, 25–32. [CrossRef] [PubMed]
14. Assoni, L.; Girardello, R.; Converso, T.R.; Darrieux, M. Current Stage in the Development of *Klebsiella pneumoniae* Vaccines. *Infect. Dis. Ther.* **2021**, *10*, 2157–2175. [CrossRef]
15. López-Siles, M.; Corral-Lugo, A.; McConnell, M.J. Vaccines for Multidrug Resistant Gram Negative Bacteria: Lessons from the Past for Guiding Future Success. *FEMS Microbiol. Rev.* **2021**, *45*, fuaa054. [CrossRef]
16. Frost, I.; Sati, H.; Garcia-Vello, P.; Hasso-Agopsowicz, M.; Lienhardt, C.; Gigante, V.; Beyer, P. The Role of Bacterial Vaccines in the Fight against Antimicrobial Resistance: An Analysis of the Preclinical and Clinical Development Pipeline. *Lancet Microbe* **2023**, *4*, e113–e125. [CrossRef]
17. Kurupati, P.; Teh, B.K.; Kumarasinghe, G.; Poh, C.L. Identification of Vaccine Candidate Antigens of an ESBL Producing *Klebsiella pneumoniae* Clinical Strain by Immunoproteome Analysis. *Proteomics* **2006**, *6*, 836–844. [CrossRef]

18. Lundberg, U.; Senn, B.M.; Schüler, W.; Meinke, A.; Hanner, M. Identification and Characterization of Antigens as Vaccine Candidates against *Klebsiella pneumoniae*. *Hum. Vaccin. Immunother.* **2013**, *9*, 497–505. [CrossRef]

19. Magalhães, S.; Aroso, M.; Roxo, I.; Ferreira, S.; Cerveira, F.; Ramalheira, E.; Ferreira, R.; Vitorino, R. Proteomic Profile of Susceptible and Multidrug-Resistant Clinical Isolates of *Escherichia coli* and *Klebsiella pneumoniae* Using Label-Free and Immunoproteomic Strategies. *Res. Microbiol.* **2017**, *168*, 222–233. [CrossRef]

20. Rappuoli, R. Reverse Vaccinology. *Curr. Opin. Microbiol.* **2000**, *3*, 445–450. [CrossRef]

21. Rappuoli, R.; Bottomley, M.J.; D'Oro, U.; Finco, O.; De Gregorio, E. Reverse Vaccinology 2.0: Human Immunology Instructs Vaccine Antigen Design. *J. Exp. Med.* **2016**, *213*, 469–481. [CrossRef] [PubMed]

22. Bidmos, F.A.; Siris, S.; Gladstone, C.A.; Langford, P.R. Bacterial Vaccine Antigen Discovery in the Reverse Vaccinology 2.0 Era: Progress and Challenges. *Front. Immunol.* **2018**, *9*, 2315. [CrossRef] [PubMed]

23. Fenoy, E.; Edera, A.A.; Stegmayer, G. Transfer Learning in Proteins: Evaluating Novel Protein Learned Representations for Bioinformatics Tasks. *Brief. Bioinform.* **2022**, *23*, bbac232. [CrossRef] [PubMed]

24. Clifford, J.N.; Høie, M.H.; Deleuran, S.; Peters, B.; Nielsen, M.; Marcatili, P. BepiPred-3.0: Improved B-Cell Epitope Prediction Using Protein Language Models. *Protein Sci.* **2022**, *31*, e4497. [CrossRef]

25. Shashkova, T.I.; Umerenkov, D.; Salnikov, M.; Strashnov, P.V.; Konstantinova, A.V.; Lebed, I.; Shcherbinin, D.N.; Asatryan, M.N.; Kardymon, O.L.; Ivanisenko, N.V. SEMA: Antigen B-Cell Conformational Epitope Prediction Using Deep Transfer Learning. *Front. Immunol.* **2022**, *13*, 960985. [CrossRef]

26. Galanis, K.A.; Nastou, K.C.; Papandreou, N.C.; Petichakis, G.N.; Pigis, D.G.; Iconomidou, V.A. Linear B-Cell Epitope Prediction for In Silico Vaccine Design: A Performance Review of Methods Available via Command-Line Interface. *Int. J. Mol. Sci.* **2021**, *22*, 3210. [CrossRef]

27. McConnell, M.J.; Martín-Galiano, A.J. Designing Multi-Antigen Vaccines Against *Acinetobacter baumannii* Using Systemic Approaches. *Front. Immunol.* **2021**, *12*, 666742. [CrossRef]

28. Dar, H.A.; Zaheer, T.; Shehroz, M.; Ullah, N.; Naz, K.; Muhammad, S.A.; Zhang, T.; Ali, A. Immunoinformatics-Aided Design and Evaluation of a Potential Multi-Epitope Vaccine against *Klebsiella pneumoniae*. *Vaccines* **2019**, *7*, 88. [CrossRef]

29. Mehmood, A.; Naseer, S.; Ali, A.; Fatimah, H.; Rehman, S.; Kiani, A.K. Identification of Novel Vaccine Candidates against Carbapenem Resistant *Klebsiella pneumoniae*: A Systematic Reverse Proteomic Approach. *Comput. Biol. Chem.* **2020**, *89*, 107380. [CrossRef]

30. Cuscino, N.; Fatima, A.; Di Pilato, V.; Bulati, M.; Alfano, C.; Monaca, E.; Di Mento, G.; Di Carlo, D.; Cardinale, F.; Monaco, F.; et al. Computational Design and Characterization of a Multiepitope Vaccine against Carbapenemase-Producing *Klebsiella pneumoniae* Strains, Derived from Antigens Identified through Reverse Vaccinology. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4446–4463. [CrossRef]

31. Wang, Z.; Guo, G.; Li, Q.; Li, P.; Li, M.; Zhou, L.; Tan, Z.; Zhang, W. Combing Immunoinformatics with Pangenome Analysis To Design a Multiepitope Subunit Vaccine against *Klebsiella pneumoniae* K1, K2, K47, and K64. *Microbiol. Spectr.* **2022**, *10*, e0114822. [CrossRef] [PubMed]

32. Allemailem, K.S. A Comprehensive Computer Aided Vaccine Design Approach to Propose a Multi-Epitopes Subunit Vacciner against Genus Klebsiella Using Pan-Genomics, Revese Vaccinology, and Biophysical Techniques. *Vaccines* **2021**, *9*, 1087. [CrossRef] [PubMed]

33. Cañada-García, J.E.; Moure, Z.; Sola-Campoy, P.J.; Delgado-Valverde, M.; Cano, M.E.; Gijón, D.; González, M.; Gracia-Ahufinger, I.; Larrosa, N.; Mulet, X.; et al. CARB-ES-19 Multicenter Study of Carbapenemase-Producing *Klebsiella pneumoniae* and *Escherichia coli* From All Spanish Provinces Reveals Interregional Spread of High-Risk Clones Such as ST307/OXA-48 and ST512/KPC-3. *Front. Microbiol.* **2022**, *13*, 918362. [CrossRef]

34. Kumar, C.K.; Sands, K.; Walsh, T.R.; O'Brien, S.; Sharland, M.; Lewnard, J.A.; Hu, H.; Srikantiah, P.; Laxminarayan, R. Global, Regional, and National Estimates of the Impact of a Maternal *Klebsiella pneumoniae* Vaccine: A Bayesian Modeling Analysis. *PLoS Med.* **2023**, *20*, e1004239. [CrossRef]

35. Kringelum, J.V.; Nielsen, M.; Padkjær, S.B.; Lund, O. Structural Analysis of B-Cell Epitopes in Antibody: Protein Complexes. *Mol. Immunol.* **2013**, *53*, 24–34. [CrossRef]

36. Potocnakova, L.; Bhide, M.; Pulzova, L.B. An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J. Immunol. Res.* **2016**, *2016*, 6760830. [CrossRef]

37. Doud, M.B.; Lee, J.M.; Bloom, J.D. How Single Mutations Affect Viral Escape from Broad and Narrow Antibodies to H1 Influenza Hemagglutinin. *Nat. Commun.* **2018**, *9*, 1386. [CrossRef]

38. Cao, Y.; Wang, J.; Jian, F.; Xiao, T.; Song, W.; Yisimayi, A.; Huang, W.; Li, Q.; Wang, P.; An, R.; et al. Omicron Escapes the Majority of Existing SARS-CoV-2 Neutralizing Antibodies. *Nature* **2022**, *602*, 657–663. [CrossRef]

39. Guilhen, C.; Charbonnel, N.; Parisot, N.; Gueguen, N.; Iltis, A.; Forestier, C.; Balestrino, D. Transcriptional Profiling of *Klebsiella pneumoniae* Defines Signatures for Planktonic, Sessile and Biofilm-Dispersed Cells. *BMC Genom.* **2016**, *17*, 237. [CrossRef]

40. Jeong, H.; Mason, S.P.; Barabási, A.L.; Oltvai, Z.N. Lethality and Centrality in Protein Networks. *Nature* **2001**, *411*, 41–42. [CrossRef]

41. Eckburg, P.B.; Bik, E.M.; Bernstein, C.N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S.R.; Nelson, K.E.; Relman, D.A. Diversity of the Human Intestinal Microbial Flora. *Science* **2005**, *308*, 1635–1638. [CrossRef] [PubMed]

42. Overly Cottom, C.; Stephenson, R.; Wilson, L.; Noinaj, N. Targeting BAM for Novel Therapeutics against Pathogenic Gram-Negative Bacteria. *Antibiotics* **2023**, *12*, 679. [CrossRef]

43. Slifka, M.K.; Amanna, I. How Advances in Immunology Provide Insight into Improving Vaccine Efficacy. *Vaccine* **2014**, *32*, 2948–2957. [CrossRef]

44. Harro, J.M.; Achermann, Y.; Freiberg, J.A.; Allison, D.L.; Brao, K.J.; Marinos, D.P.; Sanjari, S.; Leid, J.G.; Shirtliff, M.E. Clearance of *Staphylococcus aureus* from In Vivo Models of Chronic Infection by Immunization Requires Both Planktonic and Biofilm Antigens. *Infect. Immun.* **2019**, *88*, e00586-19. [CrossRef]

45. Lu, L.L.; Suscovich, T.J.; Fortune, S.M.; Alter, G. Beyond Binding: Antibody Effector Functions in Infectious Diseases. *Nat. Rev. Immunol.* **2018**, *18*, 46–61. [CrossRef]

46. Juraja, S.M.; Mulhern, T.D.; Hudson, P.J.; Hattarki, M.K.; Carmichael, J.A.; Nuttall, S.D. Engineering of the *Escherichia coli* Im7 Immunity Protein as a Loop Display Scaffold. *Protein Eng. Des. Sel.* **2006**, *19*, 231–244. [CrossRef]

47. Martinson, J.N.V.; Walk, S.T. *Escherichia coli* Residency in the Gut of Healthy Human Adults. *EcoSal Plus* **2020**, *9*. [CrossRef]

48. Gato, E.; Rodiño-Janeiro, B.K.; Gude, M.J.; Fernández-Cuenca, F.; Pascual, Á.; Fernández, A.; Pérez, A.; Bou, G. Diagnostic Tool for Surveillance, Detection and Monitoring of the High-Risk Clone. *K. pneumoniae* ST15. *J. Hosp. Infect.* **2023**, *142*, 18–25. [CrossRef]

49. Ahn, D.; Peñaloza, H.; Wang, Z.; Wickersham, M.; Parker, D.; Patel, P.; Koller, A.; Chen, E.I.; Bueno, S.M.; Uhlemann, A.-C.; et al. Acquired Resistance to Innate Immune Clearance Promotes *Klebsiella pneumoniae* ST258 Pulmonary Infection. *JCI Insight* **2016**, *1*, e89704. [CrossRef]

50. Kitts, P.A.; Church, D.M.; Thibaud-Nissen, F.; Choi, J.; Hem, V.; Sapojnikov, V.; Smith, R.G.; Tatusova, T.; Xiang, C.; Zherikov, A.; et al. Assembly: A Resource for Assembled Genomes at NCBI. *Nucleic Acids Res.* **2016**, *44*, D73–D80. [CrossRef]

51. Maiden, M.C.; Bygraves, J.A.; Feil, E.; Morelli, G.; Russell, J.E.; Urwin, R.; Zhang, Q.; Zhou, J.; Zurth, K.; Caugant, D.A.; et al. Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3140–3145. [CrossRef] [PubMed]

52. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]

53. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved Protein Subcellular Localization Prediction with Refined Localization Subcategories and Predictive Capabilities for All Prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. [CrossRef] [PubMed]

54. Sievers, F.; Higgins, D.G. Clustal Omega for Making Accurate Alignments of Many Protein Sequences. *Protein Sci.* **2018**, *27*, 135–145. [CrossRef] [PubMed]

55. Mirdita, M.; Steinegger, M.; Söding, J. MMseqs2 Desktop and Local Web Server App for Fast, Interactive Sequence Searches. *Bioinformatics* **2019**, *35*, 2856–2858. [CrossRef]

56. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-Mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [CrossRef]

57. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res* **2019**, *47*, D309–D314. [CrossRef]

58. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res* **2021**, *49*, D412–D419. [CrossRef]

59. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic Acids Res.* **2013**, *41*, e121. [CrossRef]

60. Zou, Y.; Xue, W.; Luo, G.; Deng, Z.; Qin, P.; Guo, R.; Sun, H.; Xia, Y.; Liang, S.; Dai, Y.; et al. 1,520 Reference Genomes from Cultivated Human Gut Bacteria Enable Functional Microbiome Analyses. *Nat. Biotechnol.* **2019**, *37*, 179–185. [CrossRef]

61. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef] [PubMed]

62. Hagberg, A.; Swart, P.; Chult, D. Exploring Network Structure, Dynamics, and Function Using NetworkX. In Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 19–24 August 2008.

63. Hon, J.; Marusiak, M.; Martinek, T.; Kunka, A.; Zendulka, J.; Bednar, D.; Damborsky, J. SoluProt: Prediction of Soluble Protein Expression in *Escherichia coli*. *Bioinformatics* **2021**, *37*, 23–28. [CrossRef] [PubMed]

64. Sharma, N.; Naorem, L.D.; Jain, S.; Raghava, G.P.S. ToxinPred2: An Improved Method for Predicting Toxicity of Proteins. *Brief. Bioinform.* **2022**, *23*, bbac174. [CrossRef]

65. Nguyen, M.N.; Krutz, N.L.; Limviphuvadh, V.; Lopata, A.L.; Gerberick, G.F.; Maurer-Stroh, S. AllerCatPro 2.0: A Web Server for Predicting Protein Allergenicity Potential. *Nucleic Acids Res.* **2022**, *50*, W36–W43. [CrossRef]

66. Käll, L.; Krogh, A.; Sonnhammer, E.L.L. Advantages of Combined Transmembrane Topology and Signal Peptide Prediction--the Phobius Web Server. *Nucleic Acids Res.* **2007**, *35*, W429–W432. [CrossRef]

67. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* **2022**, *40*, 1023–1025. [CrossRef]