



Article

Resemblance-Ranking Peptide Library to Screen for Binders to Antibodies on a Peptidomic Scale

Felix Jenne ^{1,†}, Sergey Biniaminov ^{2,†} , Nathalie Biniaminov ², Philipp Marquardt ², Clemens von Bojničić-Kninski ³, Roman Popov ³, Anja Seckinger ⁴ , Dirk Hose ^{4,*} and Alexander Nesterov-Mueller ^{1,*}

¹ Institute of Microstructure Technology, Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany; felix.jenne@t-online.de

² HS Analysis, Haid-und-Neu-Straße 7, 76131 Karlsruhe, Germany; sergey.biniaminov@hs-analysis.com (S.B.); nathalie.biniaminov@hs-analysis.com (N.B.); philipp.marquardt@hs-analysis.com (P.M.)

³ Axxelera, Karl-Floesser-Str. 14, 76189 Karlsruhe, Germany; clemens.kninski@axxelera.com (C.v.B.-K.); roman.popov@axxelera.com (R.P.)

⁴ Department of Hematology and Immunology, Myeloma Center Brussels, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, 1090 Jette, Belgium; as@lamkapbio.com

* Correspondence: dirk.hose@vub.be (D.H.); alexander.nesterov-mueller@kit.edu (A.N.-M.)

† These authors contributed equally to this work.

Abstract: A novel resemblance-ranking peptide library with 160,000 10-meric peptides was designed to search for selective binders to antibodies. The resemblance-ranking principle enabled the selection of sequences that are most similar to the human peptidome. The library was synthesized with ultra-high-density peptide arrays. As proof of principle, screens for selective binders were performed for the therapeutic anti-CD20 antibody rituximab. Several features in the amino acid composition of antibody-binding peptides were identified. The selective affinity of rituximab increased with an increase in the number of hydrophobic amino acids in a peptide, mainly tryptophan and phenylalanine, while a total charge of the peptide remained relatively small. Peptides with a higher affinity exhibited a lower sum helix propensity. For the 30 strongest peptide binders, a substitutional analysis was performed to determine dissociation constants and the invariant amino acids for binding to rituximab. The strongest selective peptides had a dissociation constant in the hundreds of the nano-molar range. The substitutional analysis revealed a specific hydrophobic epitope for rituximab. To show that conformational binders can, in principle, be detected in array format, cyclic peptide substitutions that are similar to the target of rituximab were investigated. Since the specific binders selected via the resemblance-ranking peptide library were based on the hydrophobic interactions that are widespread in the world of biomolecules, the library can be used to screen for potential linear epitopes that may provide information about the cross-reactivity of antibodies.

Keywords: rituximab; off targeting; ultra-high-density peptide microarrays



Citation: Jenne, F.; Biniaminov, S.; Biniaminov, N.; Marquardt, P.; von Bojničić-Kninski, C.; Popov, R.; Seckinger, A.; Hose, D.; Nesterov-Mueller, A. Resemblance-Ranking Peptide Library to Screen for Binders to Antibodies on a Peptidomic Scale. *Int. J. Mol. Sci.* **2022**, *23*, 3515. <https://doi.org/10.3390/ijms23073515>

Academic Editor: Antonio Carrieri

Received: 16 February 2022

Accepted: 21 March 2022

Published: 23 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Screening for specific antibody binders at the peptidomic scale is of practical value. They can provide important clues about pathogens in the case of circulating antibodies in the patient's blood or about potential off-targets for therapeutic antibodies.

There are different approaches to determine antibody binders using a screening approach. For example, tissue cross-reactivity assays based on immunohistochemistry are traditionally performed to identify potential off-target epitopes at the tissue level [1]. Data from such studies do not provide comprehensive information about off-target epitopes but are mandatory in preclinical safety assessments [2,3]. Novel approaches comprise recombinant protein arrays generated either by pipetting them onto a solid support [4] or by direct translation and transcription from DNA arrays [5–7]. In this way, the binding

of the antibody to the conformational fragments of the proteins, as well as the linear fragments available in the configuration that the proteins take on the surface of the chip, can be identified. Because proteins can change their configuration, for example, during various post-translational modifications [8] or an increase in temperature [9], new linear fragments from the proteome can be available for interactions. They can be identified using short (<10 amino acids) peptide libraries for screening at a higher resolution.

H.B. Larman et al. constructed a synthetic representation of the complete human proteome for the discovery of unreported autoantibodies, which were engineered as peptides on the surface of T7 phage [10]. However, phage display screenings are notorious for the identification of false-positive hits [11]. These emerge for two important reasons: binding to non-target-related materials used during selection and propagation advantages [12]. Often, hits from phage display screening are validated, for example, using peptide microarrays to find invariant amino acids via substitutional analysis [13,14] or to determine the affinity and the serological number of peptide–antibody interactions in peptide array format [15]. Knowledge of invariant amino acids in epitopes contributes to a more targeted search for antigens in proteomic databanks.

To fully display the human peptidome, tens of millions of short peptides are needed. This is even more potentiated if, in future, not only linear but also non-linear targets are to be displayed. Today, such possibilities are provided by the maskless lithographic technique [16,17]. However, a density of 1 million different peptides per square centimeter means focusing the modulated light from the beamer into $10 \times 10 \mu\text{m}$ pixels, including the distance between spots, which leads to the partial radiation overlap of neighboring pixels and reduces the quality of synthesis. Such a problem does not arise in mask lithographic peptide synthesis since the mask adheres tightly to the synthesis surface and the diffraction exposure of neighboring pixels is excluded [18]. However, lithographic peptide synthesis must use and precisely position several hundred masks (one mask per amino acid multiplied by the number of peptide residues: 20×10 masks and positioning steps for a ten-mer peptide library) to gain complete combinatorics for a library of interest, which makes such chips very expensive. The pseudo-randomization of the peptide content was proposed by Legutki et al. to minimize the number of masks required for super high-density peptide synthesis [19]. However, this procedure significantly reduces the similarity of the pseudo-random peptide library with the human peptidome. Today, displaying the entire peptidome on a single chip is impossible. However, a smart library design for the reduction mapping of a peptidome can be applied.

The goal of this work is to generate a novel resemblance-ranking library of 160,000 10-meric peptides with the highest resemblance to the whole peptidome and the study of its features. As proof of principle, we investigated the interaction of the resemblance-ranking library with the therapeutic anti-CD20 antibody rituximab (RTX) [20]. The selective binders to rituximab were validated by measuring the corresponding dissociation constants and substitutional analysis.

2. Results

2.1. Combinatorial Diversity versus Human Peptidome

The theoretical number N of unique k -mers exponentially grows according to M^k , where $M = 20$ is the number of proteinogenic amino acids. In the human peptidome, however, the diversity of the longer chains with $k \geq 7$ is significantly reduced and asymptotically approaches 11 million fragments (Figure 1a). This fact determines the optimal length of potential linear epitope fragments, which the adaptive immune system can address.

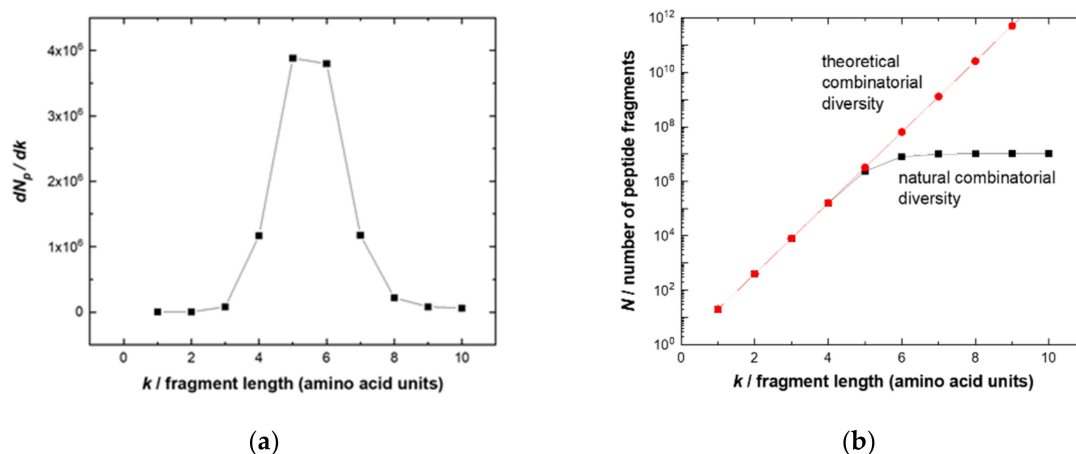


Figure 1. (a) Comparison of the number N of unique fragments of the complete combinatorial library (red line with circles) with the number N_p of unique fragments of the human peptidome (black line with squares) depending on the length of the peptide fragment. Starting from the 6-mer peptide, N_p tends to saturate; (b) entropy gain dN_p/dk versus the length k of the peptide sequence.

The likely explanation of such optimization is that, on the one hand, short fragments ($k \leq 4$) are all present in the peptidome and thus specificity would be difficult to obtain, i.e., resulting in potential autoantigens. On the other hand, targeting longer fragments ($k > 8$) would be more “expensive” (due to the combinatorial complexity of 20^k), necessitating a higher diversity (and therefore number) of naïve B cells. The optimum fragments have a length of five to six amino acids that provide sufficient uniqueness and distinction for a specific sequence between a full-combinatorial library of potential pathogens and a human peptidome. These epitope lengths account for the highest entropy gain of a growing peptide chain (Figure 1b). Thus, the library of 10-mer peptides would be a good approximation to cover the potential linear epitopes of physiologically occurring or therapeutic antibodies relying on the same repertoire. Peptides of this length can also exhibit secondary structures such as alpha helixes (3.4 amino acids per turn) [21].

2.2. Design and Properties of the Resemblance-Ranking Peptide Library

In total, 20,350 manually annotated and reviewed proteins of *Homo sapiens sapiens* were collected with the analysis software HSA KIT (HS Analysis GmbH) [22] using the Swiss-Prot section of the UniProt database [23] (access date 27 March 2020). The protein sequences were preprocessed to replace all 37 selenocysteines (U) with cysteines (C). The proteins were computationally sliced into continuous 10-mer fragments with an overlap of nine amino acids. All duplicate fragments were eliminated, resulting in 10,438,489 unique peptides. These peptides, in turn, consisted of 51,475,217 unique k -length sub-fragments ($1 \leq k \leq 10$), which comprised a basis of a high-dimensional vector space.

Each unique peptide was one-hot (binary) encoded [24] as a 51,475,217-dimensional binary vector accounting for all the k -length sub-fragments being included in the respective amino acid sequence. Altogether, the human peptidome was represented as a sparse matrix SP_{nm} of shape $10,438,489 \times 51,475,217$, where n is the number of unique peptides and m the number of unique k -length fragments.

Each encoding k -length sub-fragment was assigned a weighted score according to the number of its occurrences in the whole proteome (see Section 4 and supporting files with codes). These weight scores were represented as vector W_m .

The vector F_n in Equation (1)

$$F_n = \sum_{i=1}^m SP_{ni} W_i \quad (1)$$

results as a sum of the weighted scores over all k -length sub-fragments that are included in the corresponding peptide. The peptides with the maximum total weight in the F_n were added to the resemblance-ranking library.

Table 1 presents the ten highest-scored peptides. The difference in scores cannot be used for cross-comparison because they are obtained in different iterations of the sequential scoring and selection algorithm. Each next peptide added to the library has a different basis for evaluation: we reset the scores of those subfragments that are already included in the library with previously selected peptides. Each subsequent peptide is evaluated only for those fragments of k -mers that are not yet available in the library.

Table 1. Ten highly-scored peptides.

Peptide Sequence	Scores
IHTGEKPYKC	4726
KCEECGKAFS	2006
HQRIHTGERP	1493
THTGEKPYEC	1404
EDEEEEEED	1299
LPPPPPPPLP	1123
PAAAAAAGG	1059
EKPYKCEECG	959
PYECKEKGKA	956
KPYKCNECGK	906

The distribution of the peptides of the resemblance-ranking peptide library according to the scores does not correlate with the peptide sum hydrophobicity, sum molecular weight, and sum helix propensity. Interestingly, the resemblance-ranking library has an integral negative charge considering that the occurrence of positively charged amino acids (Arg, His, Lys) is higher than those negatively charged (Asp, Glu) in the peptidome (Figure 2). According to the negative charge, the entire resemblance-ranking library is divided into two groups: the first 30,000 peptides have greater negativity (greater slope of the red dotted line in Figure 2) than the rest.

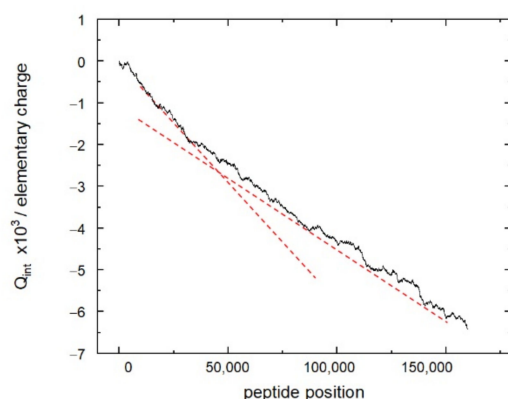


Figure 2. Integral charge of the resemblance-ranking peptide library depending on the interval of integration. Peptides are arranged according to their weight in the library.

2.3. Interaction of the Resemblance-Ranking Peptide Library with RTX

In the first screen, an RTX concentration of 30 $\mu\text{g}/\text{mL}$ was used. This value correlates with a presumptive “active” level of up 20 $\mu\text{g}/\text{mL}$ in anti-lymphoma treatment [25–27]. We

assumed that the strength of the interaction is correlated with the intensity of fluorescent signals from secondary antibodies.

Even though the library itself is integrally negatively charged, during the screen the most intense interaction of RTX was observed with peptides, the total charge of which is no more than two (Figure 3). Moreover, the sign of this charge does not matter. This trend can be seen in the example of the positively charged amino acids arginine (R) and lysine (K) and negatively charged amino acids glutamate (E) and aspartate (D) (Figure 4). In the region of the strongest signals, where off-target sequences are expected, strong charge oscillations are observed, which may reflect the specificity of the rituximab paratope.

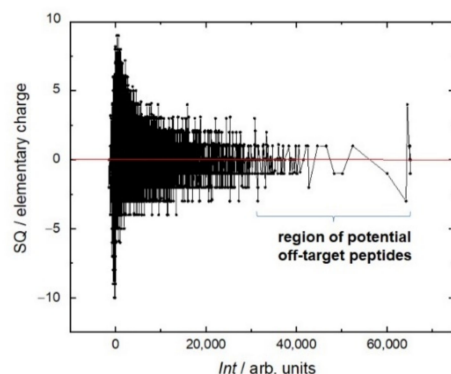


Figure 3. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the sum charge SQ of the corresponding peptides. Here and in further graphs, the black line indicates the neighboring peptides with the growing signal intensity.

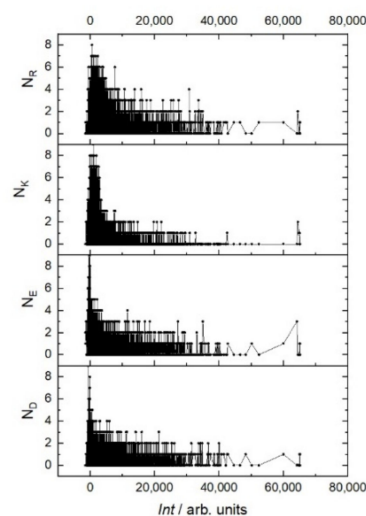


Figure 4. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the number N of positively charged amino acids R and K and negatively charged amino acids E and D of the corresponding peptides.

Figure 5 shows that the antibody affinity increases with the molecular weight of peptides. This is due to the appearance of large amino acids, the structures of which are more capable of providing stronger interactions. This is the main mechanism to increase the affinity in the case of linear sequences in contrast to conformational epitopes. As shown in Figure 6, this trend is caused mostly by hydrophobic amino acids. A more detailed analysis of individual amino acids showed that the affinity increases with the number of hydrophobic amino acids, such as tryptophan (W), phenylalanine (F), leucine (L), isoleucine (I), and the polar amino acid tyrosine (Y). These amino acids eliminate the amino acids with low molecular weights, such as alanine (A), threonine (T), and serine (S). Other amino

acids, such as methionine (M), valine (V), proline (P), glycine (G), as well as glutamine (Q), and asparagine (N), correlate very weakly with affinity. The increase in the affinity correlates most strongly with the sum of W and F in the peptide (Figure 7).

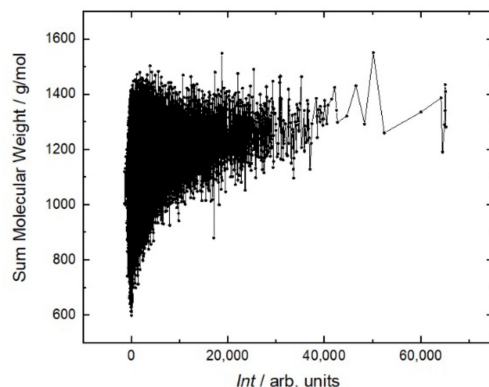


Figure 5. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the sum molecular weight of the corresponding peptides.

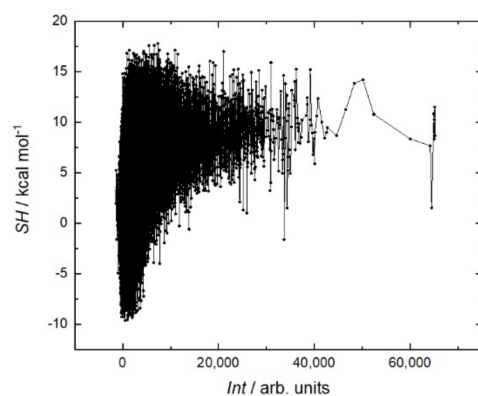


Figure 6. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the sum hydrophobicity SH of the corresponding peptides [28]. Wimley–White whole-residue hydrophobicity scales were used for calculations.

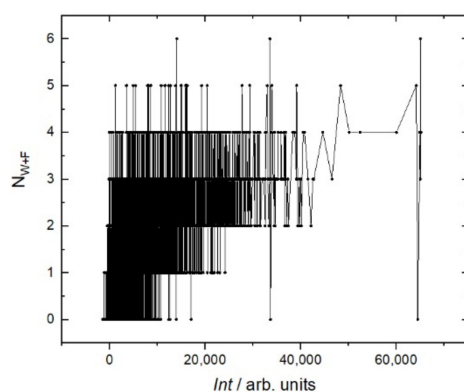


Figure 7. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the number of sum of hydrophobic amino acids W and F of the corresponding peptides.

It is generally believed that a stiffer structure allows for stronger binding in peptide–protein interactions. Various methods are used to reduce entropy; for example, peptide cyclization [29] or the clips method [30]. In the case of rituximab, a sum helix propensity

was not a criterion for strong binding to the peptide (Figure 8). On the contrary, the peptides with the strongest affinity are located in the region with a low sum helix propensity of about 5 kcal mol.

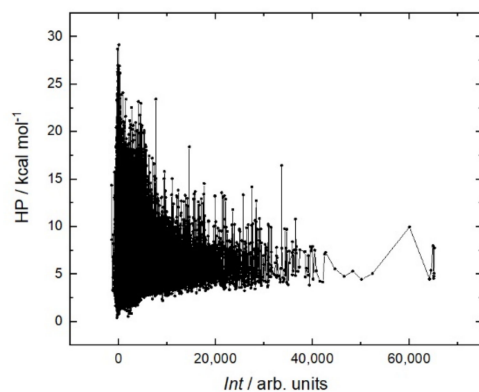


Figure 8. Fluorescent intensity Int of RTX interaction with the resemblance-ranking peptide library versus the sum helix propensity HP of the corresponding peptides [31].

2.4. Validation of the Resemblance-Ranking Library Hits via Substitutional Microarrays and K_D Measurements in Array Format

The 30 peptides that gave the highest signals in the first screen were selected to identify dissociation constants K_D to RTX and invariant amino acids. Thanks to the ultra-dense format of peptide microarrays, the library with all substitutions for 30 peptides fit into one of eight windows (Figure 9). Accordingly, eight identical sublibraries were synthesized and incubated with RTX in the form of a diluted series with concentrations of 200, 100, 50, 25, 12.5, 6.25, 3.125, and 1.5625 $\mu\text{g}/\text{mL}$. This allowed the generation of both a substitutional analysis for each selected peptide and to identify individual K_D values for the selected peptides and all 200 substitutional variations for each of the selected peptides. K_D and the maximum fluorescent signal at saturation I_{sat} were calculated via the approximation of the measured fluorescent signals by the Formula (2):

$$I_{obs} = I_{sat} \frac{n}{K_D + n} \quad (2)$$

where I_{obs} is the measured signal intensity for the defined spot, n is the concentration of the analyte (RTX in our case) in the solution, and K_D is the equilibrium dissociation constant [32].

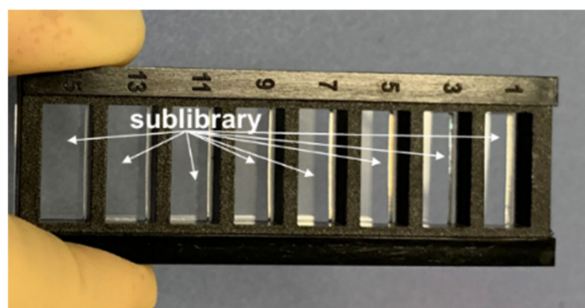


Figure 9. Sublibrary consisting of the substitutions to the peptides selected from the first screens. Eight windows were used to incubate the same sublibrary with different RTX concentrations.

In our case, only six peptides had K_D values ≤ 1000 nM (Table 2). These values are significantly higher than the reported K_D for the RTX target, which are between 5 and 20 nM [33]. However, these values are in the range of protein–protein interactions representing important biological functions [34,35]. The selected peptides with $K_D \leq 1000$ nM

have a significant content of hydrophobic amino acids W and F, which corresponds to the trend shown in Figure 8. Usually, hydrophobic amino acids are associated with nonspecific interactions. Nevertheless, the substitutional analysis showed that epitopes with hydrophobic amino acids have a clear selective profile and may be responsible for binding with high specificity (Figure 10).

Table 2. Off-target peptides for RTX with a dissociation constant ≤ 1000 nM with their containing proteins.

Off-Target Peptide	K_D , nM	Protein, Submitted Names	ID
WFAEFWEENF	243	Single-stranded DNA-binding protein 3	Q9BWW4
		Metabotropic glutamate receptor 6	O15303
		Metabotropic glutamate receptor 8	O00222
		Seven transmembrane helix receptor	Q8NHA9
		cDNA FLJ75348, highly similar to Homo sapiens metabotropic glutamate receptor 8b	A8K2D2
RDGDRFWWEN	577	Myeloperoxidase	P05164
		Lactoperoxidase	P22079
LHSWWCVFWD	655	Single-stranded DNA-binding protein 2	P81877
		Single-stranded DNA-binding protein 3	Q9BWW4
		Single-stranded DNA-binding protein 4	Q9BWG4
		HSPC116	Q9P038
		Single-stranded DNA binding protein 4, isoform CRA_d LisH domain-containing protein	A0A024R7K9 A1L192
SYSLEIQWWY	676	V-set and transmembrane domain-containing protein 2-like protein	Q96N03
		V-set and transmembrane domain-containing protein 2B	A6NLU5
FTGWFLAWDP	853	Villin-1	P09327
		Advillin	O75366
YFPRARWYDY	1110	Probable maltase-glucoamylase 2	Q2M2H8
		Maltase-glucoamylase	Q8TE24
		Maltase-glucoamylase, intestinal	E7ER45

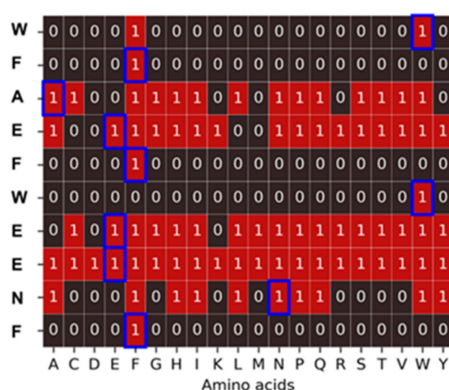


Figure 10. Substitution analysis for the peptide WFAEFWEENF. The blue frames correspond to the amino acids of the peptide under study. Squares show the corresponding fluorescent signal for all substitution sequences. In the case of red squares, the signal is equal to or higher than the signal from the original sequence. The epitope is composed of a combination of the invariant amino acids, W and F: W/FFxxFWxxxF.

3. Discussion

The recent analysis of the maturation of the affinity of the capsid protein VP1 of the poliovirus epitope under alanine substitutions showed that there is no clear boundary

between the nonspecific and medically relevant sequences [36]. When replacing polyalanine with the corresponding amino acids, the affinity increased in a logarithmic manner up to a known specific signal. The resemblance-ranking library provides a unique opportunity to trace the trends in the formation of selective peptide binders for antibodies, from very weak ones to epitopes of potential significance with K_D values in the submicromolar range.

Earlier research with phage display suggested that the binding of RTX to the targeted discontinuous epitope EPANPSEKNSSTQY occurs within the extracellular fragment of CD20, cyclized via the disulfide bridge between Cys(167) and Cys(183) residues, which do not contain hydrophobic epitopes [37]. The sequence EPANPSEK was identified as the key region of this interaction using X-ray diffraction [38]. The 10-mer peptide EPANPSEKNS that was presented on the same peptide chip along with the resemblance library gave the signal at the noise level when incubated with RTX (126,756 peptides from the resemblance-ranking library delivered stronger signals). This confirmed the non-linear structure of the RTX target binding. We performed a substitution analysis of the EPANPSEKNSSTQY sequence, which was enclosed in a macrocycle, as shown in Figure 11, to mimic the cyclic fragment of CD20. In this case, the macrocycle was formed using head-to-tail thioether macrocyclization with the orthogonal side group deprotection of cysteine [39].

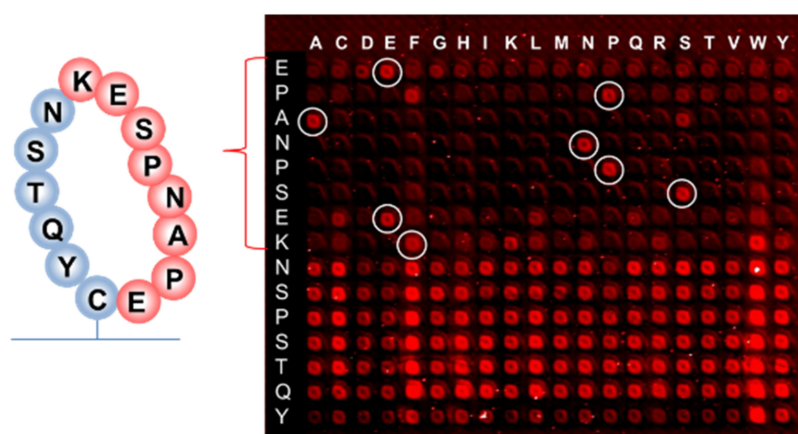


Figure 11. (right) Fluorescence scan of the substitutional array for the cyclized peptide EPANPSEKNSSTQY after incubation with rituximab and subsequent immunostaining. The white circles show amino acids from the original sequence EPANPSEK. The cyclic peptides were formed via the thioether macrocyclization. (left) The RTX target sequence and the scheme of the peptide cycle.

Although hydrophobic amino acids W and F are not present in the RTX target sequence, selective binding remains when the proline P at the second position of the EPANPSEK epitope is replaced by phenylalanine F, and lysine K at the end of EPANPSEK is replaced by tryptophan W. Other substitutions at these positions resulted in the disruption of the selective binding with RTX.

RTX hydrophobic linear epitopes WWWS/T [40] and WPXWLE [41] were previously reported. Interestingly, both forward and reverse sequences, WPKWLE and ELWKPW, bind rituximab, but not other anti-CD20 mAbs [42]. Fornoni et al. provided data to support a nonimmune mechanism by which RTX may exert its therapeutic effects in recurrent FSGS after renal transplantation [43]. These data were based on the off-target binding of RTX to the reverse sequence ELWKPW that is part of the acid sphingomyelinase-like phosphodiesterase 3b protein expressed on podocytes. The peptide LWKPWLQPCC from the resemblance-ranking peptide library sharing the 5-meric fragment with this podocyte epitope was identified at the 3156th place in terms of fluorescent signal intensity with a ratio of 0.13 to the maximum signal value among 160,000 peptides. It is likely that medically relevant signals can also be located with the weaker resemblance-ranking library signals within 10% of the maximum fluorescent intensity.

To the best of our knowledge, the molecular reason for the specific epitopes with hydrophobic amino acids is not well understood. At the same time, it is known that water competes with the ligand for binding to the target and critically affects its dissociation [44]. Due to water, binding affinity is often difficult to predict even with structural information [45,46]. It is likely that the observed hydrophobic epitopes carry information about the specific displacement of water molecules from the RTX paratope during ligand association [47].

The resemblance-ranking library does not take into account the concentration of proteins that may be available for interaction with the therapeutic antibody. For the specific target- or off-target binding of monoclonal antibodies, it would indeed be excellent to have data on tissue distribution and an abundance of all potential targets a priori at hand. However, these data do not exist. Even for frequently targeted and best characterized hematological proteins, such as B-cell maturation antigen (BCMA), only semiquantitative data regarding expression height exist [48]. As it is therefore not possible to include target abundances for all potential target sequences across all tissues a priori, we suggest using our approach to first identify potential linear target sequences, and then further assess these regarding target expression and distribution (patho-)physiologically.

4. Materials and Methods

4.1. Design of the Resemblance-Ranking Peptide Library

To maximize the representing power of the library with 160,000 peptides, only such amino acid sequences were chosen which had the most in common with the whole human peptidome, while being largely different from each other. In each iteration, all 10,438,489 peptides were dynamically scored accounting for the weighted scores of their k-length sub-fragments, as well as for the sub-fragments of the peptides selected in the previous iterations. The initial protein sequences, the resemblance-ranking algorithm as the python codes with comments for every processing step, and the resulting 160,000 peptides are available in the Zenodo repository, <https://zenodo.org/record/6046581#.YgaDBJYo82w> (access date 20 March 2022).

4.2. Staining of the Resemblance-Ranking Peptide Library with RTX

Peptide microarrays: A peptide library of 160,000 peptides as well other peptide microarrays for the substitutions and the K_D analysis, were fabricated via the ultra-high-density peptide microarray technology of AXCELERA (Karlsruhe, Germany) [49]. The AXCELERA technology allows for the 30 μm -sized square peptide spots with a pitch of 60 μm that corresponds to ca. 28,000 spots/ cm^2 .

Immunostaining: Peptide microarrays were incubated with the monoclonal anti-CD20 antibody rituximab (Sigma-Aldrich, Steinheim, Germany) with concentrations depending on the type of experiment, diluted in PBS-T (Phosphate Buffered Saline with Tween) overnight at 4 °C. A rituximab concentration of 30 $\mu\text{g}/\text{mL}$ was used for the first screen of the entire resemblance-ranking peptide library. Then, the peptide chips were stained with the fluorescently labeled secondary anti-human-IgG antibodies (Jackson ImmunoResearch, Philadelphia, PA, USA). The signals from the spots were read out with the confocal scanner Innoscan 1100 AL. The scans were performed in the red channel (635 nm), with a resolution of 2 $\mu\text{m}/\text{pixel}$, a speed of 35 $\mu\text{m}/\text{s}$, and a PMT gain of 9.

5. Conclusions and Outlook

The entropy analysis of the peptidome showed that the five and six amino acid fragments should play an essential functional biological role (constructive, metabolic, and signaling) since they account for the maximum entropy increase. A resemblance-ranking peptide library with 160,000 10-meric peptides that represent the entire human peptidome was designed. The library is based on the vector representation of peptides through their shorter fragments. This approach enables the encoding of the entire peptidome as one large sparse matrix, the dimensions of which are determined by the number of peptides and

the number of smaller amino acid fragments. The resemblance-ranking peptide library was constructed from those peptides whose fragments are most frequently present in the peptidome.

The resemblance-ranking library was synthesized using ultra-high-density peptide arrays to search for selective binders to the therapeutic antibody rituximab. The peptide library enabled us to highlight trends in the composition of peptides that enhance their affinity to rituximab. The affinity weakly depended on the charge of the peptide, but it grew with an increase in its mass. This was due to an increase in the number of hydrophobic amino acids, especially tryptophan and phenylalanine. The peptides with the maximum affinity were found in the region with low sum helix propensity.

Dissociation constants K_D were measured for the selected 30 peptides with the highest fluorescent signals. Six peptides with $K_D \leq 1000$ nM were identified among them. A substitutional analysis revealed specific epitopes with invariant hydrophobic amino acids W and F.

To show that conformational binders can, in principle, be detected in array format, cyclic peptide substitutions that are similar to the target of rituximab were investigated. Specific substitutions of proline for phenylalanine and lysine for tryptophan in the target rituximab cyclic peptide did not interfere with the selective binding of the antibody. The reason for the selective hydrophobic epitopes was discussed.

The 10-meric peptide binders to rituximab selected in the screens are part of the proteins belonging to the human proteome. This does not mean that the corresponding proteins are automatically off-targeted by the therapeutic antibody. However, information on specific peptide binders can supplement information on off-targets obtained, for example, using other high-throughput methods such as protein arrays, to determine linear epitopes and possible reasons for the cross-reactivity of antibodies.

Author Contributions: Conceptualization, R.P., D.H. and A.N.-M.; methodology, N.B., P.M., D.H., A.S., F.J., S.B. and C.v.B.-K.; software, R.P. and S.B.; validation, D.H., A.S., F.J., S.B., C.v.B.-K., R.P. and A.N.-M.; writing—original draft preparation, D.H. and A.N.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Förderprogramm „Zentrales Innovationsprogramm Mittelstand“ des Bundesministeriums für Wirtschaft und Energie (BMWi), grant number ZF4407505CR9 and DFG, grant number AOBJ655892.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The initial protein sequences, the resemblance-ranking algorithm as the python codes with comments for every processing step, and the resulted 160,000 peptides are available in the Zenodo repository, <https://zenodo.org/record/6046581#.YgaDBJYo82w> (access date 20 March 2022).

Acknowledgments: We acknowledge support from the KIT-Publication Fund of the Karlsruhe Institute of Technology.

Conflicts of Interest: ANM is a shareholder of AXCELERA. Other authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Leach, M.W.; Halpern, W.G.; Johnson, C.W.; Rojko, J.L.; MacLachlan, T.K.; Chan, C.M.; Galbreath, E.; Ndifor, A.M.; Blanset, D.L.; Polack, E.; et al. Use of Tissue Cross-reactivity Studies in the Development of Antibody-based Biopharmaceuticals: History, Experience, Methodology, and Future Directions. *Toxicol. Pathol.* **2010**, *38*, 1138–1166. [[CrossRef](#)] [[PubMed](#)]
2. Almutairi, A.; Sun, Z.; Al Safran, Z.; Poovathumkadavi, A.; Albader, S.; Ifdailat, H. Optimal Scanning Protocols for Dual-Energy CT Angiography in Peripheral Arterial Stents: An in Vitro Phantom Study. *Int. J. Mol. Sci.* **2015**, *16*, 11531–11549. [[CrossRef](#)] [[PubMed](#)]

3. FDA. Guidances (Drugs). 2021. Available online: <https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs> (accessed on 15 February 2022).
4. Michaud, G.A.; Salcius, M.; Zhou, F.; Bangham, R.; Bonin, J.; Guo, H.; Snyder, M.; Predki, P.F.; Schweitzer, B.I. Analyzing antibody specificity with whole proteome microarrays. *Nat. Biotechnol.* **2003**, *21*, 1509–1512. [[CrossRef](#)] [[PubMed](#)]
5. He, M.; Stoevesandt, O.; Palmer, E.A.; Khan, F.; Ericsson, O.; Taussig, M.J. Printing protein arrays from DNA arrays. *Nat. Methods* **2008**, *5*, 175–177. [[CrossRef](#)] [[PubMed](#)]
6. Ramachandran, N.; Hainsworth, E.; Bhullar, B.; Eisenstein, S.; Rosen, B.; Lau, A.Y.; Walter, J.C.; LaBaer, J. Self-Assembling Protein Microarrays. *Science* **2004**, *305*, 86–90. [[CrossRef](#)] [[PubMed](#)]
7. Angenendt, P.; Kreuzberger, J.; Glökler, J.; Hoheisel, J.D. Generation of High Density Protein Microarrays by Cell-free in Situ Expression of Unpurified PCR Products. *Mol. Cell. Proteom.* **2006**, *5*, 1658–1666. [[CrossRef](#)] [[PubMed](#)]
8. Deribe, Y.L.; Pawson, T.; Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **2010**, *17*, 666–672. [[CrossRef](#)]
9. Biggar, K.K.; Dawson, N.J.; Storey, K.B. Real-time protein unfolding: A method for determining the kinetics of native protein denaturation using a quantitative real-time thermocycler. *BioTechniques* **2012**, *53*, 231. [[CrossRef](#)]
10. Larman, H.B.; Zhao, Z.; Laserson, U.; Li, M.Z.; Ciccia, A.; Gakidis, M.A.M.; Church, G.; Kesari, S.; LeProust, E.M.; Solimini, N.L.; et al. Autoantigen discovery with a synthetic human peptidome. *Nat. Biotechnol.* **2011**, *29*, 535–541. [[CrossRef](#)]
11. Hoen, P.A.; Jirka, S.M.; Broeke, B.R.T.; Schultes, E.A.; Aguilera, B.; Pang, K.H.; Heemskerk, H.; Aartsma-Rus, A.; van Ommen, G.J.; Dunnen, J.T.D. Phage display screening without repetitious selection rounds. *Anal. Biochem.* **2012**, *421*, 622–631. [[CrossRef](#)]
12. Vodnik, M.; Zager, U.; Strukelj, B.; Lunder, M. Phage Display: Selecting Straws Instead of a Needle from a Haystack. *Molecules* **2011**, *16*, 790–817. [[CrossRef](#)] [[PubMed](#)]
13. Weber, L.K.; Palermo, A.; Kügler, J.; Armant, O.; Isse, A.; Rentschler, S.; Jaenisch, T.; Hubbuch, J.; Dübel, S.; Nesterov-Mueller, A.; et al. Single amino acid fingerprinting of the human antibody repertoire with high density peptide arrays. *J. Immunol. Methods* **2017**, *443*, 45–54. [[CrossRef](#)] [[PubMed](#)]
14. Weber, L.K.; Isse, A.; Rentschler, S.; Kneusel, R.E.; Palermo, A.; Hubbuch, J.; Nesterov-Mueller, A.; Breitling, F.; Loeffler, F.F. Antibody fingerprints in lyme disease deciphered with high density peptide arrays. *Eng. Life Sci.* **2017**, *17*, 1078–1087. [[CrossRef](#)] [[PubMed](#)]
15. Palermo, A.; Nesterov-Mueller, A. Serological Number for Characterization of Circulating Antibodies. *Int. J. Mol. Sci.* **2019**, *20*, 604. [[CrossRef](#)] [[PubMed](#)]
16. Shin, D.-S.; Lee, K.-N.; Yoo, B.-W.; Kim, J.; Kim, M.; Kim, Y.-K.; Lee, Y.-S. Automated Maskless Photolithography System for Peptide Microarray Synthesis on a Chip. *J. Comb. Chem.* **2010**, *12*, 463–471. [[CrossRef](#)]
17. Buus, S.; Rockberg, J.; Forsström, B.; Nilsson, P.; Uhlen, M.; Schafer-Nielsen, C. High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays. *Mol. Cell. Proteom.* **2012**, *11*, 1790–1800. [[CrossRef](#)] [[PubMed](#)]
18. Fodor, S.P.A.; Read, J.L.; Pirrung, M.C.; Stryer, L.; Lu, A.T.; Solas, D. Light-Directed, Spatially Addressable Parallel Chemical Synthesis. *Science* **1991**, *251*, 767–773. [[CrossRef](#)]
19. Legutki, J.B.; Zhao, Z.-G.; Greving, M.; Woodbury, N.; Johnston, S.A.; Stafford, P. Scalable high-density peptide arrays for comprehensive health monitoring. *Nat. Commun.* **2014**, *5*, 4785. [[CrossRef](#)]
20. Smith, M.R. Rituximab (monoclonal anti-CD20 antibody): Mechanisms of action and resistance. *Oncogene* **2003**, *22*, 7359–7368. [[CrossRef](#)]
21. Gimpelev, M.; Forrest, L.; Murray, D.; Honig, B. Helical Packing Patterns in Membrane and Soluble Proteins. *Biophys. J.* **2004**, *87*, 4075–4086. [[CrossRef](#)]
22. Person, F.; Petschull, T.; Wulf, S.; Buescheck, F.; Biniaminov, S.; Fehrle, W.; Oh, J.; Skerka, C.; Zipfel, P.F.; Wiech, T. In situ Visualization of C3/C5 Convertases to Differentiate Complement Activation. *Kidney Int. Rep.* **2020**, *5*, 927–930. [[CrossRef](#)] [[PubMed](#)]
23. UniProt. Available online: <http://www.uniprot.org> (accessed on 15 February 2022).
24. Okada, S.; Ohzeki, M.; Taguchi, S. Efficient partition of integer optimization problems with one-hot encoding. *Sci. Rep.* **2019**, *9*, 13036. [[CrossRef](#)] [[PubMed](#)]
25. Cartron, G.; Blasco, H.; Paintaud, G.; Watier, H.; Le Guellec, C. Pharmacokinetics of rituximab and its clinical use: Thought for the best use? *Crit. Rev. Oncol.* **2007**, *62*, 43–52. [[CrossRef](#)]
26. Rodriguez, J.; Gutierrez, A. Pharmacokinetic properties of rituximab. *Rev. Recent Clin. Trials* **2008**, *3*, 22–30. [[CrossRef](#)] [[PubMed](#)]
27. Berinstein, N.L.; Grillo-Lopez, A.J.; White, C.A.; Bence-Bruckler, I.; Maloney, D.; Czuczman, M.; Green, D.; Rosenberg, J.; McLaughlin, P.; Shen, D. Association of serum Rituximab (IDEC-C2B8) concentration and anti-tumor response in the treatment of recurrent low-grade or follicular non-Hodgkin's lymphoma. *Ann. Oncol.* **1998**, *9*, 995–1001. [[CrossRef](#)] [[PubMed](#)]
28. Fauchere, J.L.; Pliska, V. Hydrophobic Parameters- Π of Amino-Acid Side-Chains from the Partitioning of N-Acetyl-Amino-Acid Amides. *Eur. J. Med. Chem.* **1983**, *18*, 369–375.
29. White, C.J.; Yudin, A.K. Contemporary strategies for peptide macrocyclization. *Nat. Chem.* **2011**, *3*, 509–524. [[CrossRef](#)] [[PubMed](#)]
30. Streefkerk, D.E.; Schmidt, M.; Ippel, J.H.; Hackeng, T.M.; Nuijens, T.; Timmerman, P.; van Maarseveen, J.H. Synthesis of Constrained Tetracyclic Peptides by Consecutive CEPS, CLIPS, and Oxime Ligation. *Org. Lett.* **2019**, *21*, 2095–2100. [[CrossRef](#)]
31. Pace, C.N.; Scholtz, J.M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J.* **1998**, *75*, 422–427. [[CrossRef](#)]

32. Jones, R.B.; Gordus, A.; Krall, J.A.; MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **2006**, *439*, 168–174. [[CrossRef](#)]
33. Bar, L.; Nguyen, C.; Galibert, M.; Santos-Schneider, F.; Aldrian, G.; Dejeu, J.; Lartia, R.; Coche-Guérente, L.; Molina, F.; Boturyn, D. Determination of the Rituximab Binding Site to the CD20 Epitope Using SPOT Synthesis and Surface Plasmon Resonance Analyses. *Anal. Chem.* **2021**, *93*, 6865–6872. [[CrossRef](#)] [[PubMed](#)]
34. Sukenik, S.; Ren, P.; Gruebele, M. Weak protein-protein interactions in live cells are quantified by cell-volume modulation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6776–6781. [[CrossRef](#)] [[PubMed](#)]
35. Seo, M.-H.; Park, J.; Kim, E.; Hohng, S.; Kim, H.-S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **2014**, *5*, 3724. [[CrossRef](#)] [[PubMed](#)]
36. Popov, R.; Shankara, G.K.; von Bojničić-Kninski, C.; Nesterov-Mueller, A. Renaissance Distribution for Statistically Failed Experiments. *Int. J. Mol. Sci.* **2019**, *20*, 3250. [[CrossRef](#)]
37. Binder, M.; Otto, F.; Mertelsmann, R.; Veelken, H.; Trepel, M. The epitope recognized by rituximab. *Blood* **2006**, *108*, 1975–1978. [[CrossRef](#)]
38. Du, J.; Wang, H.; Zhong, C.; Peng, B.; Zhang, M.; Li, B.; Huo, S.; Guo, Y.; Ding, J. Structural Basis for Recognition of CD20 by Therapeutic Antibody Rituximab. *J. Biol. Chem.* **2007**, *282*, 15073–15080. [[CrossRef](#)]
39. Roberts, K.D.; Lambert, J.N.; Ede, N.J.; Bray, A.M. Efficient synthesis of thioether-based cyclic peptide libraries. *Tetrahedron Lett.* **1998**, *39*, 8357–8360. [[CrossRef](#)]
40. Polyak, M.J.; Deans, J.P. Alanine-170 and proline-172 are critical determinants for extracellular CD20 epitopes; heterogeneity in the fine specificity of CD20 monoclonal antibodies is defined by additional requirements imposed by both amino acid sequence and quaternary structure. *Blood* **2002**, *99*, 3256–3262. [[CrossRef](#)]
41. Perosa, F.; Favoino, E.; Caragnano, M.A.; Dammacco, F. Generation of biologically active linear and cyclic peptides has revealed a unique fine specificity of rituximab and its possible cross-reactivity with acid sphingomyelinase-like phosphodiesterase 3b precursor. *Blood* **2006**, *107*, 1070–1077. [[CrossRef](#)]
42. Chan, A.C. Rituximab's New Therapeutic Target: The Podocyte Actin Cytoskeleton. *Sci. Transl. Med.* **2011**, *3*, 85ps21. [[CrossRef](#)]
43. Fornoni, A.; Sageshima, J.; Wei, C.; Merscher-Gomez, S.; Aguillon-Prada, R.; Jauregui, A.N.; Li, J.; Mattiazzi, A.; Ciancio, G.; Chen, L.; et al. Rituximab Targets Podocytes in Recurrent Focal Segmental Glomerulosclerosis. *Sci. Transl. Med.* **2011**, *3*, 85ra46. [[CrossRef](#)] [[PubMed](#)]
44. Homans, S.W. Water, water everywhere—Except where it matters? *Drug Discov. Today* **2007**, *12*, 534–539. [[CrossRef](#)] [[PubMed](#)]
45. Ladbury, J.E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973–980. [[CrossRef](#)]
46. Schiebel, J.; Gaspari, R.; Wulsdorf, T.; Ngo, K.; Sohn, C.; Schrader, T.E.; Cavalli, A.; Ostermann, A.; Heine, A.; Klebe, G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Commun.* **2018**, *9*, 1–15. [[CrossRef](#)] [[PubMed](#)]
47. Pantsar, T.; Kaiser, P.D.; Kudolo, M.; Forster, M.; Rothbauer, U.; Laufer, S.A. Decisive role of water and protein dynamics in residence time of p38 α MAP kinase inhibitors. *Nat. Commun.* **2022**, *13*, 569. [[CrossRef](#)] [[PubMed](#)]
48. Seckinger, A.; Delgado, J.A.; Moser, S.; Moreno, L.; Neuber, B.; Grab, A.; Lipp, S.; Merino, J.; Prosper, F.; Emde, M.; et al. Target Expression, Generation, Preclinical Activity, and Pharmacokinetics of the BCMA-T Cell Bispecific Antibody EM801 for Multiple Myeloma Treatment. *Cancer Cell* **2017**, *31*, 396–410. [[CrossRef](#)] [[PubMed](#)]
49. AXXELERA UG. Available online: <https://axxelera.com/> (accessed on 15 February 2022).