MAbTope: A Method for Improved Epitope Mapping

Thomas Bourquard, Astrid Musnier, Vincent Puard, Shifa Tahir, Mohammed Akli Ayoub, Yann Jullian, Thomas Boulo, Nathalie Langonné Gallay, Herve Watier, Gilles Bruneau, et al.

MAbTope: a Method for Improved Epitope Mapping[1]

Running title: MAbTope: a New Method for Improved Epitope Mapping

Thomas Bourquard[*,†,2], Astrid Musnier[*,‡,2], Vincent Puard[‡], Shifa Tahir[*], Mohammed Akli Ayoub[*,§] , Yann Jullian[¶], Thomas Boulo[*], Nathalie Langonné[*,‖] , Hervé Watier[‖], Gilles Bruneau[*], Eric Reiter[*], Pascale Crépieux[*], Anne Poupon[*]

*: PRC, INRA, CNRS, Université François Rabelais-Tours, 37380, Nouzilly, France.

†: Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

‡: MAbSilico SAS, Domaine de l'Orfrasière, 37380 Nouzilly, France.

§: Biology Department, College of Science, United Arab Emirates University, PO Box 15551 Al Ain, United Arab Emirates

¶: CaSciModOT, UFR de Sciences et Techniques, Université François-Rabelais, F-37041 Tours, France

[1] This publication was funded with support from the French National Research Agency under the program "Investissements d'avenir" Grant Agreement LabEx MabImprove: ANR-10-LABX-53 ; ANR (Contract n° ANR-2011–1619 01) ; ARTE2, MODUPHAC, MABSILICO, "ARD2020 Biomédicament" grants from Région Centre.

[2] Equal contributions

|| : CNRS; Université François-Rabelais de Tours; CHRU de Tours, UMR 7292, 37032 Tours, France

Corresponding author:

Anne Poupon

Tel: +33 2 45 47 75 05

Fax: +33 2 45 47 77 43

email: Anne.poupon@inra.fr

Abstract

Antibodies are very efficient drugs, about 70 of them are already approved for medical use, over 500 are in clinical development, and many more are in preclinical development. One important step in the characterization and protection of a therapeutic antibody is the determination of its cognate epitope. The gold standard is the 3D structure of the antibody-antigen complex by crystallography or NMR. However, it remains a tedious task and its outcome is uncertain. We have developed MAbTope, a docking-based prediction method of the epitope associated with straightforward experimental validation procedures. We show that MAbTope predicts the correct epitope for each of 129 tested examples of antibody-antigen complexes of known structure. We further validated this method through the successful determination of the epitopes recognized by two therapeutic antibodies targeting tumor necrosis factor α (TNF-α): certolizumab and golimumab.

Introduction

The use of antibodies (Abs) as drugs against a large number of diseases has dramatically increased in the last decade, and this tendency should still intensify in the near future (1). Because many antibodies are often developed against a same target, it has become essential to determine the epitope of an antibody early in its development. Moreover, the identification of the epitope is an important element in the understanding of Ab mechanism of action (2).

Aside from 3D structures, most experimental methods available for epitope determination are based either on (i) site-directed mutagenesis; (ii) peptide arrays (3–5) or (iii) mass spectrometry (6). Most peptide-based methods use 15-30 amino acids overlapping peptides of the target arrayed on solid support, which are then exposed to the antibody (4, 5). This identification of interacting peptides can then be completed by alanine-scanning in order to define the epitope more precisely (3). In the mass spectrometry-based approach, the antibody-antigen complex is subjected either to hydrogen/deuterium exchange (HDX-MS) (7), or to enzymatic digestion, which allows differentiating target peptides that are "protected" by the presence of the antibody. These peptides can then be identified using mass spectrometry (see (6) for a review). It should be noted that even when successful, these different approaches are likely to provide non-identical definitions of the epitope. Indeed, because of the crystallisation step that freezes the complex structure in one out of many possible conformations, X-ray structure identifies only the most stable interactions. Alanine scanning does not allow identifying all the interacting residues for different reasons: the mutated amino acid might interact with the antibody through its main chain or the mutation to alanine might not be drastic enough to give rise to a measurable difference in

affinity. Still, there is usually a large overlap between the epitopes identified by each method, which corresponds to the core of the interface.

However, these approaches are expensive, time-consuming and, except crystallography, remain error-prone. Indeed, the results obtained through HDX-MS are sometimes very difficult to interpret, for example when there is a conformational change in the target between the free and complexed forms (7). Peptide arrays performance at identifying epitopes are limited by different factors (8): immobilization methods, affinity of the peptides and conformational constraints induced by the immobilization. For these reasons, many efforts have been put in developing *in silico* methods capable of predicting antibody-antigen interactions. This endeavour has taken two main directions: (i) B-cell epitope prediction, which aims at predicting the regions of a protein that are the most amenable of being targeted by an antibody; and (ii) partner-specific approaches, which aim at predicting the epitope for a single antibody-target pair (see (9, 10) for reviews). Only the second type of method leads to the prediction of the epitope for a given antibody, though B-cell epitope prediction can be a useful first step in this process. Amongst the partner-specific approaches, three main categories can be distinguished: predictors based on the intrinsic properties of the partners, predictors based on co-evolution of the partners, and predictors based on docking. However, few of these methods are dedicated to the special case of antibody-antigen interaction.

The aim of docking methods is originally the prediction of the conformation of the assembly between two interacting proteins. From a correct prediction of this conformation, the interaction regions can be straightforwardly defined. For this reason, docking methods have been applied to the prediction of interaction interfaces, and in some cases the specific issue of predicting the epitope and the paratope. Some methods provide accurate results, such as

Rosetta (11) and Z-dock (12), but in local docking only, meaning that they require a partial knowledge of the epitope. The introduction of sDARS, a pairwise statistic potential specific of antibody-antigen interactions, allows PIPER/Cluspro (which is the algorithm used for docking within the Bioluminate suite) to achieve satisfactory results (13), placing at least one near-native solution in the top-10 predicted conformations. The particularity of this statistic potential as compared to previously used ones is that it accounts for the asymmetry of the antibody-antigen interaction. Another example of a web-server specific for antibody-antigen docking is Frodock (14, 15). Frodock uses spherical harmonics for conformation generation (as opposed to fast-Fourier transform for most other algorithms, including PIPER), and a combination of energetic (van der Waals, electrostatics and desolvation) and knowledge-based potentials, optimized for the different categories of complexes (enzyme, antibodies and others). However, the goal of these methods is predicting the conformation of the assembly, meaning predicting the interaction region, but also the precise relative orientations of the two partners, and not predicting the epitope. Even though they perform better at this task than the other types of epitope prediction methods, they are not optimized for it.

We have developed a new method for epitope determination, named MAbTope, which integrates both a docking-based prediction method and experimental steps. Indeed, the software part of the method automatically outputs peptides, without any human intervention, that can be readily used for experimental validation. We also show how these peptides can be used to design point mutations in the target, allowing a more precise definition of the epitope. Thus, this method, although in part computational, is not just a prediction method, but also includes the experimental validation of the epitope.

Material and methods

Overview of the method

The 3D structures of the antibody and of the target are used as input of the Hex software (16) (Figure 1). Hex generates more than 10^8 docking poses and ranks them according to energetic criteria (H-ranking). Each of the Hex top-500 docking poses is evaluated using 30 specific and 30 non-specific scoring functions. The non-specific scoring function is identical to the one used in PRIOR (17), the specific scoring function has been re-optimised, using the learning dataset described hereafter, and the same machine-learning procedure: genetic algorithm and CMA-ES (18), and in both cases the area under the ROC-curve is used as fitness function. A consensus score is then computed using the formula:

$$CONS(i) = \frac{k_b}{2} \cdot \left[ln\left(\frac{\sum_{j=1}^{30} rank_j^{NS}(i)}{\sum_{j=1}^{30} rank_j^{NS}(0)}\right) + ln\left(\frac{\sum_{j=1}^{30} rank_j^{S}(i)}{\sum_{j=1}^{30} rank_j^{S}(0)}\right)\right]$$

Where $rank_j^{NS}(i)$ is the ranking of pose i according to the non-specific function j, and $\sum_{j=1}^{30} rank_j^{S}(i)$ is the ranking of pose i according to the specific function j. The rankings of pose 0 (the best ranked according to Hex) are used for normalization.

For each pose, the algorithm also computes the A-score, C-score and P-score (see hereafter). For each residue *r* of the target, we compute a value, *Vr*, which is the sum of the Hex ranks of the poses in which *r* belongs to the interface. For a given pose, the A-score is the sum of the *Vr* of the residues that belong to the epitope in this particular pose. For each pose, the C-score is the sum of the ranks of the other poses that have a RMSD value lower than 5 Å with this particular pose.

The consensus, Hex-rank, A-score, C-score and P-score are used to generate 5 different rankings. For each pose, the sum of its ranks in the different rankings is computed. These numbers are used to generate the final ranking. The top 30 solutions are then used to

compute the interface frequency (IF) of each residue of the target, which is equal to the number of poses within these 30 in which the residue belongs to the interface. This IF is used to design the interface peptides (see hereafter).

P-score

A new post-processing function has been introduced: the P-score. For a given docking pose, we count the number of CDR amino acids that are closer than 4Å to an atom of the target, and normalize by the total number of CDR residues. The docking poses are then ranked by decreasing values of this ratio. This rank is the P-score of the pose.

Specific learning dataset

The learning dataset is composed of 393 non-redundant antibody-target complexes manually extracted from the PDB in January 2015. Only the complexes in which the target is larger than 40 residues were considered. These complexes contain 392 distinct Abs and the targets belong to 165 distinct Pfam families. The definition of non-redundancy we use is weaker than what is usually used, since antibodies are very special proteins, and overall sequence identity, even restricted to the variable domain, is not indicative of the antibody specificity, and consequently on its ability to form a complex with its target. The criteria retained for considering two Ab-Ag complexes as non-redundant were: (i) targets are not related (they belong to different Pfam families); or (ii) targets are related but the epitopes recognized by the antibodies have less than 20% overlap; or (iii) targets are related and epitopes are overlapping but the CDRs of the considered antibody differ in 10 or more positions. This third criterion is justified by the fact that most pairs of antibodies differing by 10 or more residues within the CDRs, even when they present a very high overall sequence identity, do not share the same target.

Test dataset

To evaluate the performance of the method, a test dataset has been designed. It consists in the 82 complexes of the learning dataset for which the 3D structures of the individual partners are known. For the evaluation, the learning has been done in leave-one-out, meaning that the epitope of a given antibody is predicted using a scoring function learnt on a dataset not containing the 3D structure of the complex it forms with its target. Forty-seven new complexes whose 3D structure has been determined after January 2015, and which were non-redundant with those already present in the learning dataset have been added to this test set.

We distinguished “small” targets (40 to 300 amino acid long) from “large” targets (more than 300 amino acid long). However, the results obtained for the two categories only slightly differ.

Negative controls

In order to better evaluate the method performance, we have included negative controls. To this aim, we have compared, for each target of the test set, the epitope predicted by docking each of the non-cognate Abs of the test set to the actual epitope.

Epitope definition

In this work, an amino acid of a protein targeted by an antibody will be considered as belonging to the epitope if at least one of its atoms is at less than 4 Å of an atom belonging to an amino acid of the antibody. These distances are computed on the crystallographic structure of the complex.

Definition of epitope peptides

Each amino acid of the target is attributed a value, which is the number of poses within the 30 top-ranked ones in which this amino acid belongs to the predicted epitope. Different sizes of pose sets have been tested, and 30 is a satisfactory compromise (data not shown). Each 15 amino acid peptide of the sequence is then given a score equal to the sum of these values for each amino acid in the peptide. The peptides are then ranked along this score. Peptides overlapping by at least 8 amino acids with a better-ranked peptide are ignored. For benchmarking, the relevance of a given peptide is evaluated by the number of residues that belong to the crystallographic epitope. This definition of epitope peptide was also used for the testing of Cluspro and FRODOCK. In EpiPred predictions, amino acids present in the first predicted epitope were given a score of 3, a score of 2 for the amino acids of the second epitope and a score of 1 for the amino acids of the third epitope. In PPiPP predictions, the scores given in the program output were considered. Epitope peptides were then built as explained above.

The choice of 15 mers is a compromise between two empirical observations we have made along the development of this method. (i) Shorter peptides tend to give a poor signal. Our hypotheses are that they are too flexible (short peptides present less long-range interactions and are thus more flexible), which decreases their binding to the antibody. Moreover, the secondary structure is important for binding, and very short peptides have no chance to adopt hairpin or strand conformations. (ii) Longer peptides tend to span over more than one loop, and interpretation of experimental results is then more difficult. A second aspect is that longer peptides have a higher tendency to precipitate.

Evaluation criteria

MAbTope first output is a ranked list of docking poses. To evaluate the distance between these poses and the native solution, CAPRI criteria were used (19):

- High quality (***): fnat>0.5 AND (Irmsd<1 OR Lrmsd<1);
- Medium quality (**): [fnat ε[0.3, 0.5] AND (Irmsd<2 OR Lrmsd<5)] OR [fnat>0.5 AND Irmsd>1 AND Lrmsd>1];
- Acceptable (*): [fnat>0.3 AND Irmsd>2 AND Lrmsd>5] OR [fnat ε[0.1, 0.3] AND (Irmsd<4 OR Lrmsd<10)];

where fnat is the fraction of correctly predicted contacts, Lrmsd (Ligand RMSD) is the RMSD between the predicted position of the ligand and its position in the crystal structure, Irmsd (Interface RMSD) is the same but reduced to the interface residues.

Since our epitope predictions are based on the evaluation of a set of conformations, and not on a single conformation, it was necessary for us to also evaluate the number of "indicative" conformations:

- Indicative (+): [fnat>0.1] OR [Lrmsd< 10] OR [Irmsd < 5]

Introducing this new category is very useful for evaluating docking performance in the perspective of epitope determination. Indeed, the docking poses falling in this category, even though their geometry is too distant from the crystal structure to be considered as acceptable by the Capri criteria, still define an interaction area on the target that overlaps with the actual epitope, and thus give valuable information on the epitope.

To evaluate the docking performances of our algorithms, for each complex in the test set we calculate the rank of the first near-native pose with the CAPRI criteria and with our own criteria (CAPRI + indicative).

The second output of MAbTope is a list of peptides, ranked on the predicted probability they match with the epitope. To evaluate the epitope prediction accuracy, we calculate the number of residues in each peptide that belong to the actual epitope (and do not belong to better ranked peptides), normalized by the total number of residues in the epitope.

Binding kinetics of certolizumab to biotinylated peptides using biolayer interferometry (BLI)

All measurements were performed with the Octet RED96 System (Pall Forte Bio, Fremont, CA, USA), in the manufacturer kinetics buffer, at 30 °C, with shaking at 1000rpm. Biotinylated peptides were immobilized during 200 seconds on streptavidin-coated sensors (SA) at 0.5, 1 and 5 µg/mL for P1-3, P1-1 and P1-2 respectively and left for equilibration for 120 seconds in kinetics buffer. Typical capture variability within a row of eight tips did not exceed 0.1 nm. Binding was assessed at 100, 200, 400, 600, 800, 1000 and 1200 µg/mL certolizumab for 300 seconds. Two parallel corrections were carried out by subtracting the association of certolizumab on an immobilized non-relevant biotinylated peptide, and by subtracting the loading baseline drift on non-associated sensors. Data were analyzed using Octet Software 9.0 version. Since certolizumab is a Fab', experimental data were fitted with the binding equation describing a 1:1 interaction. Considering the weak affinity of peptides for the antibody and the fact that the dissociation is almost immediate, we restrained the dissociation analysis to the 20 first seconds. Global analyses of the datasets assuming that binding was reversible (full dissociation) were carried out using nonlinear least-squares fitting, allowing a single set of binding parameters to be obtained simultaneously for all concentrations used in each experiment.

HTRF-based competition assay

The competition between Golimumab and either the Certolizumab or the peptides for the TNFα was assessed *in vitro* using an HTRF®-based assay in 384-well plate. The Golimumab and the Certolizumab were kindly provided by Denis Mulleman (CHRU Bretonneau, Tours, France). The Golimumab was incubated at 0.1, 0.33 and 1 nM with 8 ng of TNFα (NP_000585.2, Val77-Leu233) N-terminally fused to the AviTag® (Avidity LLC, Aurora, CO, USA) purchased from ACROBiosystems (Newark, DE, USA) in 10 µl of PPI - Terbium detection buffer (CisBio Bioassays, Condolet, France). Five microliters containing either 4 mM of non-biotinylated peptides (GeneCust, Dudelange, Luxembourg) or 4 µM of Certolizumab were added. The HTRF-compatible fluorophore Terbium cryptate and d2 conjugated to either an anti-Fc Ab or the streptavidin (from CisBio Bioassays, Condolet, France) were finally added in 5 µl. After 1h incubation at room temperature, the fluorescence at 620 nm and 665 nm were measured on the TriStar² LB 942 microplate reader (Berthold Technologies GmbH & Co, Wildbad, Germany). Data were expressed as the emission ratio 665 nm / 620 nm subtracted by the non-specific signal obtained without Ab nor peptide.

Interaction measurement by peptide array

Peptide array:

The interaction between the different biotinylated peptides (GeneCust, Dudelange, Luxembourg) and golimumab was assessed *in vitro* using peptide array. Biotinylated peptides are first diluted in printing buffer (20 % Glycerol and 1 M DMSO) for a final concentration of 0.8 nM and 1.6 mM. Peptides spotted in two replicates in 16 identical sub-arrays on a nitrocellulose coated glass slide (ONCYTE® Film slides, Grace Bio-Labs, USA) using a Nano-Plotter (GeSIM, Germany). Slides are dried overnight at room temperature.

Preparation of antibodies:

Golimumab is fluorescently labelled with iFluor™ 680 amine dye (AAT Bioquest, USA) following the protocol of the provider. Excess of dye are eliminated by centrifugation on Amicon Ultra filter (Merck Millipore, Darmstadt, Germany). Antibodies are prepared fresh for the incubation by diluting into PBS-T (PBS 1 X, 0.1 % Tween 20) supplemented with 1% of BSA (Sigma) for a final concentration of 2 ng/ml.

Incubation:

Slides are mount with Pro-Plate® chamber (Grace Bio-Labs, USA) for the following steps. Slides are hydrated with 150 µl per well of PBS-T solution for 15 min under agitation on a seesaw rocker. PBS-T is removed and 100 µl of Super G blocking buffer (Grace Bio-Labs, USA) is added for 1h incubation on a seesaw rocker. After removing the blocking buffer, 100 µl per well of antibodies diluted in PBS-t supplemented with 1 % BSA (corresponding to 200 ng) are added for 2 h incubation on a seesaw rocker. Then, antibodies are removed and slides are washed two times with PBS-T for 5 min and once with PBS (150 µl/well). Finally, slides are rinsed with filtered water for one minute and air-dried.

Detection and analysis:

Slides are scanned with an InnoScan 710-IR scanner (Innopsys, France) at 670 nm wavelength, 3 µm resolution, PMT of 1 and low intensity of the laser. Image analysis is performed using the circular feature alignment of Mapix software (Innopsys, France). Relative Fluorescence Unit (RFU) is obtained by retrieving the median fluorescence signal intensity of surrounding each feature to the median fluorescent signal of the feature. RFU is used to measure the interaction between the different peptides and the antibody. Graphs are generated using GraphPad Software (GraphPad Prism 5 Software, San Diego, CA, USA).

In vitro FRET binding measurement

The interaction between the different biotinylated peptides (GeneCust, Dudelange, Luxembourg) and Certolizumab, or Golimumab, or Eculizumab used as a negative control, was assessed by HTRF®. All experiments were performed in PPI-Terbium or -Europium detection buffers (CisBio Bioassays, Condolet, France). For this, 5 µL of biotinylated peptides (4 mM) were first incubated with 5 µL of either of the mAbs (1.6 µg/mL) for 1 hr at room temperature. Then, 5 µL of streptavidin and 5 µL of anti-Fab (for Certolizumab) or anti-Fc (for Golimumab and Eculizumab) antibodies conjugated with HTRF compatible fluorophores, Terbium or Europium cryptate and d2, were added in quantities recommended by the manufacturer. After an overnight incubation at 4°C, the fluorescence emissions at 620 nm and 665 nm were measured using the appropriate HTRF program on a TriStar² LB 942 Modular microplate reader (Berthold Technologies GmbH & Co. Wildbad, Germany). Data are represented as specific FRET signals calculated as the 665 nm/620 nm emission ratio subtracted of the binding on the non-relevant Ab.

Golimumab binding on mutant TNFα by flow cytometry

Three TNFα mutants were designed starting from the sequence NP_000585.2 by incorporating the mutations predicted to alter the interaction with Golimumab according to our docking solution. The mutant TNFα constructions contain the following mutations: TNFα_P1-1m6 (N222A, R223A, D225A, F229A, E231A, Q234A), TNFα_P3-1m7 (R167A, Y172A, Q173A, T174A, K175A, N177A), and TNFα_P4-1m6 (Q106A, E108A, Q110A, Q112A, R116A). The cDNA of the 3 mutants and the wild-type TNFα fused to a Flag tag on their N-terminus and depleted of the 77 first residues which contain the transmembrane part of the protein targeted by proteases, were synthesized and subcloned in pcDNA3.1 by GenScript

(Piscataway, NJ 08854, USA). HEK293N cells were transiently transfected with the TNFα constructions or a mock vector using Metafectene (Biontex Laboratories GmbH, München, Germany) according to manufacturer's instructions. Thirty hours after transfection, the cells were fixed and permeabilized according to the BD Cytofix/Cytoperm kit instructions (BD Biosciences, San Jose, CA, USA). All the following hybridations were performed in the kit's perm/wash buffer. Five hundred thousand cells of each transfected population were incubated with 5 µg of Golimumab for 1 hr at room temperature and washed once in 2 ml buffer. The binding of Golimumab was assessed with the allophycocyanin (APC)-labelled anti-IgG1 antibody from Miltenyi Biotech (Bergisch Gladbach, Germany) diluted to 1:100. The expression level of each of the constructions was evaluated with and anti-Flag Ab coupled to phycoerythrin (PE) also from Miltenyi Biotech (Bergisch Gladbach, Germany). After staining, all the cells were washed once in 2 ml working buffer and once in 2 ml PBS-EDTA 2 mM and finally suspended in 200 µl of PBS-EDTA 2 mM. The fluorescence was assessed with the MACSQuant Analyzer 10 (Miltenyi Biotec, Bergisch Glabach, Germany) and the data analyzed with FlowJo software (FlowJo LLC, Ashland, OR, USA).

Binding kinetics of Certolizumab to biotinylated peptides using biolayer interferometry (BLI)

All measurements were performed with the Octet RED96 System (Pall Forte Bio, Fremont, CA, USA), in the manufacturer kinetics buffer, at 30 °C, shaking at 1000 rpm. Biotinylated peptides were immobilized during 200 seconds on streptavidin-coated sensors (SA) at 0.5, 1 and 5 µg/mL for C1-3, C1-1 and C1-2 respectively and left for equilibration for 120 seconds in kinetics buffer. Typical capture variability within a row of eight tips did not exceed 0.1 nm. Binding was assessed at 100, 200, 400, 600, 800, 1000 and 1200 µg/mL Certolizumab for 300

seconds. Two parallel corrections were carried out by subtracting the association of Certolizumab on an immobilized non-relevant biotinylated peptide, and by subtracting the loading baseline drift on non-associated sensors. Data were analyzed using Octet Software 9.0 version. Since Certolizumab is a Fab', experimental data were fitted with the binding equation describing a 1:1 interaction. Considering the weak affinity of peptides for the antibody and the fact that the dissociation is almost immediate, we restrained the dissociation analysis to the 20 first seconds. Global analyses of the datasets assuming that binding was reversible (full dissociation) were carried out using nonlinear least-squares fitting, allowing a single set of binding parameters to be obtained simultaneously for all concentrations used in each experiment.

Statistical analysis

Experimental data were analysed under Prism 6 software (GraphPad Software, La Jolla, CA, USA). Data were expressed as mean ± sem and ANOVA statistical analysis was applied.

Results

Principle and Benchmarking

MAbTope involves three successive steps. The first step is the docking of the antibody on its target, which results in the generation of docking poses (possible conformations of the antibody-antigen complex), through a method related to PRIOR, a general protein-protein docking method we had previously developed (17, 21, 22). The second step is the ranking of these docking poses in order to extract 30 poses that tile the epitope, and the design of four so-called *interacting peptides*, that is, peptides predicted to be part of the epitope. The third step is the experimental validation based on the interacting peptides. Different methods can be used: measurement of the binding of each of these four peptides with the antibody, competition for antibody binding between the peptides and the target, or measurement of the binding of target mutated on residues belonging to these peptides.

The design of the interacting peptides from the docking poses is crucial for the success of the method. At this step, all the possible 15 amino acid-long peptides of the target are ranked according to the frequency at which their amino acids are found within the epitope in the 30 top-ranked docking poses. MAbTope predicts a correct peptide, that is, a peptide that contains residues belonging to the crystallographic interface, within the 4 best-ranked ones for all of the 129 complexes tested. On average the 4 best-ranked peptides contain more than 80% of the epitope residues, and the minimum is 30%, meaning that the epitope is at least partly found for all complexes in the test set (Figure 1A, Table S1). As a control, each antibody of the test set was docked to all the targets of the other antibodies. In this test, on average only 36 % of the residues belonging to the epitope of the specific antibody are found within the 4 best-ranked peptides.

MAbTope performs much better than Cluspro or FRODOCK at predicting the epitopes, as they identify, within the 4 best-ranked peptides, 36 % and 35 % of the epitope residues respectively. One reason is that, in MAbTope, the 30 top docking poses are centred on the correct epitope, and not distributed on the whole surface of the target. This is illustrated in Figure 2 (see also Figures S1, S2) by the example of the complex between the HIV gp120 glycoprotein and the VRC03 antibody (PDB 3SE8) (23). This particularity, which can be found for all of the tested examples, arises for two main reasons. First, for conformation generation, we use Hex with very restrictive angle parameters, the obtained poses are consequently already well focused. Second, the A-score (as defined in materials and methods) favours over-represented poses, and consequently decreases the diversity of the top-ranked poses. As a result, the amino acids constituting the epitope are almost all found in more than half of the 30 selected docking poses. Consequently, the four best ranking peptides all contain amino acids belonging to the interface (Figure S3). In addition, peptides 1 and 2 contain 7 and 6 amino acids belonging to the epitope, respectively. It should be noted that peptides 3 and 4 also contain 8 and 6 residues, respectively, and can also be considered as good predictions. Finally, the 6 best-ranking peptides contain all the amino acids belonging to the epitope.

We also compared the performance of MAbTope to that of two non-docking-based epitope prediction methods: PPiPP (24) and EpiPred (25). The results show that MAbTope clearly outperforms these two methods, confirming that a detailed consideration of shape and electrostatic complementarity, which results from the docking procedure, is necessary for high quality predictions (Figure 1B, Table S1).

The last step of the method consists in the experimental validation. Our first approach consists in measuring the binding of the antibody to the peptides. For each designed peptide,

three peptides are synthesized, all of the same length but sliding three amino acids along the sequence. The first one starts and ends 3 amino acids upstream of the designed epitope peptide, the second one corresponds to the designed one, and the third one starts and ends 3 amino acids downstream. This choice was made to overcome the issue of some peptides being insoluble. A second approach is to measure the competition between these peptides and the target for the binding of the antibody. Finally, as the residues present within these peptides are those predicted to belong to the epitope, they can be used to predict point mutations of the target reducing the binding of the antibody.

It should be highlighted that MAbTope is able to find the epitope of each antibody, and not only the most antigenic sites on the target protein as defined by B-cell epitope prediction methods. This is well illustrated by the example of gp120, to which 25 antibodies of the benchmark bind. Whereas some regions of gp120 are targeted by a large number of antibodies, including some that do not belong to the benchmark since the structure of the isolated antibody is not known, other regions are also targeted. Accordingly, the interaction peptides designed through MAbTope are spread on the whole target sequence (Figure 2). MAbTope correctly builds at least one correct peptide for each of these 25 antibodies, and two peptides for 19.

Validation on golimumab and certolizumab

To validate the method, we next predicted the epitopes of two therapeutic antibodies targeting Tumor Necrosis Factor α (TNF-α): golimumab and certolizumab. These two antibodies are already widely used in clinic, but their respective epitope is still unknown. We built homology models of the two antibodies and used MAbTope to predict the epitopes they bind. On the basis of the predicted epitope-antibody interface, four different sets of peptides have been selected and synthesized (G1 to G4 for golimumab and C1 to C4 for

421 certolizumab, Figure 4). The P1 family overlaps with G3 and C4, and corresponds to the
422 region containing the highest overlap between both predictions. The P4 family overlaps with
423 C2 and G4. The P3 family overlaps with G1 and C1. Finally, the P2 family does not overlap
424 with a 4-top predicted peptide, but lies in a region well exposed and predicted by MAbTope
425 to belong to certolizumab epitope but not golimumab one. Peptides G2 and C3 were ignored
426 since they are partly buried and have consequently low chances to interact efficiently with
427 the antibody.
428 After the initial submission of this paper, the structure of the complex between certolizumab
429 and TNF-α has been published (26). Comparison with our prediction shows that out of the 20
430 residues constituting the epitope, 17 belong to peptides C1 to C4 (Figure 3A). This shows
431 that certolizumab epitope can be considered as conformational since it involves residues
432 belonging to five different peptides. Nevertheless, we are still able to show the specific
433 binding of some of these peptides to the antibody through HTRF and interferometry (Figure
434 S4).
435 To validate the epitope of golimumab, we first have shown that it competes with
436 certolizumab for the binding to TNF-α, using HTRF (Figure 4A and S5). We thus performed
437 further experimental validations on golimumab solely. We have also shown, using both HTRF
438 and peptide array (RPPA), that golimumab specifically binds the P3-1, P3-2 and P3-3 peptides
439 (Figures 4B and 4C). Finally, we have shown, using HTRF, that peptides P1-1, P1-2, P1-3, P3-1
440 and P3-3 decrease the binding of golimumab to TNF-α in a dose-dependent manner. Note
441 that we observe a strong competition with the P1 series peptides in this last experiment,
442 whereas we could not observe the binding of these peptides in the direct binding
443 experiments. One hypothesis is that the biotin, which is attached at the N-ter of the peptide
444 in the direct binding experiments, could prevent the binding to the antibody. The specificity

of the binding of the P1 series peptides is confirmed by the flow cytometry experiments presented hereafter.

To further validate, we mutated in TNF-α the residues belonging to peptides of series 1, 3 and 4 to alanines, and observed the binding of golimumab using flow-cytometry (Figure 5A and S6). We observed that each TNF-α construct expressed well in cells by detecting flag epitope that was added to all constructs. Interestingly, we found that the binding of golimumab its target was almost abolished when the TNF-α was mutated at positions indicated within P1 and P3 series, and reduced by 50% for mutations within the P4 series peptides. Finally, for peptides P3-1 and P3-3, which gave the best signals in HTRF, we mutated individually the residues belonging to these peptides and whose side-chains are exposed, and measured the binding to golimumab using HTRF (Figure 5C and 5D). These results show that, as predicted, residues Y172, T174 and K175 are essential for golimumab binding to TNF-α.

Discussion

The results obtained on the 129 antibody-target complexes of the benchmark show that the *in silico* prediction is robust, since within the benchmark, the predicted peptides contain on average 80 % of the epitopes residues. This number is not much affected by the type of epitope: 79 % for conformational epitopes (105 out of 129), 89% for linear epitopes (14 out of 129). Neither is it much affected by the size of target: 88 % for targets up to 300 residues long, 70 % for larger targets. The main limitation of the *in silico* step is that the 3D structure of the target is needed. We have already tested the approach using homology models of the target when the 3D structure is not available. Although good results could be obtained is the few tested cases, this requires further investigations.

Based on the designed peptides, we present three different experimental validations of the predicted epitope. Our first approach consists in measuring the direct binding of the designed peptides, either through HTRF, peptide array or through interferometry. Good results could be obtained for the golimumab peptides of series 3. However, no signal is observed for series 1 peptides, although we later demonstrate that these peptides belong to the epitope. The second approach consists in making a competition between the peptides and the target for the binding of the antibody. Using this method we were able to validate the peptides of series 1, and confirm the peptides of series 3. Nevertheless, both approaches are limited by the fact that some peptides tend to be “sticky”. Another limit to these approaches is the solubility of peptides, which is not always sufficient.

Importantly, the interaction peptides can also be used to design point-mutations in the target potentially decreasing the affinity of the antibody. In the TNF-α we mutated to alanines the residues belonging to peptide series 1, 3 and 4 whose side-chains point towards the solvent. We show using flow cytometry that these mutations indeed abolish (series 1 and 3) or decrease (series 4) the binding of the antibody. However, this approach also has its limitations: the difficulty of expressing some target or their mutated forms, especially if they are toxic for the cells. The endogenous expression of the native target could also raise some issues.

Despite the known limitations of each experimental approach proposed, it is reasonable to assume that their combined use will convey more robustness to the overall validation process.

Further demonstration of MAbTope ability to determine the epitope is given through the examples of certolizumab and golimumab. For these two antibodies, although their 3D structure was not known at the beginning of this study, we were able to predict and

492 experimentally validate the epitopes. A good example is given by peptide 1.3, which contains
493 only one residue belonging to the epitope, but for which we were able to measure the
494 specific binding with certolizumab (Figure S4). Using mutated peptides we were also able to
495 refine these results, and show the importance of individual residues in the epitope.
496 Two other therapeutic antibodies are used in clinic for their ability to bind TNF-α: infliximab
497 and adalimumab, and the 3D structures of the corresponding complexes with the target are
498 known [4G3Y for infliximab (27) and 3WD5 for adalimumab (28)]. A recent meta-analysis has
499 compared the efficacy of different TNF-α-blocking agents, including the four antibodies cited
500 above. It concludes that infliximab and golimumab are less efficient in the treatment of
501 rheumatoid arthritis than adalimumab and certolizumab (29). By contrast, a meta-analysis
502 performed in ulcerative colitis indicated that infliximab is better than adalimumab and
503 probably golimumab (30). Their affinities for TNF-α (4.5×10^{-10} M for infliximab (28), $7.05 \times$
504 10^{-11} M for adalimumab (28), 1.8×10^{-11} M for golimumab (31) and 1.32×10^{-10} M for
505 certolizumab (US patent US20050042219 A1) do not explain these differences. Hu *et al.* (28)
506 hypothesized that the difference of efficacy between infliximab and adalimumab could be
507 partly due to the fact that adalimumab binds in the groove between two monomers, and has
508 consequently a higher overlap with the TNF-α receptor binding interface and a better
509 neutralizing activity, than infliximab, which binds to a monomer. By contrast, the ability to
510 target inflammatory cells expressing membrane TNF-α, which could be monomeric, and to
511 induce apoptotic signals seems important determinants of therapeutic activity of anti- TNF-α
512 agents in inflammatory bowel diseases (32). These reasons could also account for the
513 difference of efficacy between certolizumab and golimumab, as certolizumab binds in the
514 groove (like adalimumab), whereas golimumab binds to the monomer (Figure S7), knowing
515 that certolizumab differs from the three others by its monovalency and the absence of an Fc

region. However, the fact that the structure of the four anti-TNF-α therapeutic antibodies is now known will help at understanding the subtle differences in their clinical activities.

Conclusion

In conclusion, MAbTope initial prediction of the epitope is very robust. On a benchmark of 129 antibody-antigen complexes, MAbTope correctly defines the epitope in each case. In addition, MAbTope allows defining four 15 amino acid peptides, among which at least one belongs to the epitope, which in turn allows experimental validation. These peptides also allow the design of point mutations that can be used to validate and refine the predicted epitope. Although the information obtained through MAbTope does not allow defining the precise interactions taking place between the antibody and the target, it allows defining with good precision the region of the target involved in the interaction. This information is sufficient for understanding the mechanism of action of the antibody, a crucial step in the development of therapeutics, but also diagnostic or biotechnological tools. Taken together, MAbTope is not just a prediction method, but constitutes an integrated workflow allowing identification of the epitope. With the example of two therapeutic antibodies, certolizumab and golimumab, we show that it can be successfully applied to antibodies whose 3D structure is unknown.

Acknowledgements

We thank Prof. Denis Mulleman for providing certolizumab and golimumab. We thank Dr. Olivier Lichtarge for reading and advices.

539

Bibliography

1. Watier, H., and Reichert, Janice. 2017. Evolution of antibody therapeutics (Chapter 2). In *Protein Therapeutics*, Jallal B, Vaughan T, Osbur J. Wiley. 25–50.

2. Van Regenmortel, M. H. V. 2014. Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition. *J. Mol. Recognit.* 27: 627–639.

3. Cunningham, B. C., and J. A. Wells. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244: 1081–1085.

4. Forsström, B., B. B. Axnäs, K.-P. Stengele, J. Bühler, T. J. Albert, T. A. Richmond, F. J. Hu, P. Nilsson, E. P. Hudson, J. Rockberg, and M. Uhlen. 2014. Proteome-wide Epitope Mapping of Antibodies Using Ultra-dense Peptide Arrays. *Mol. Cell. Proteomics* 13: 1585–1597.

5. Hansen, C. S., T. Østerbye, P. Marcatili, O. Lund, S. Buus, and M. Nielsen. 2017. ArrayPitope: Automated Analysis of Amino Acid Substitutions for Peptide Microarray-Based Antibody Epitope Mapping. *PLOS ONE* 12: e0168453.

6. Opuni, K. F., M. Al-Majdoub, Y. Yefremova, R. F. El-Kased, C. Koy, and M. O. Glocker. 2016. Mass spectrometric epitope mapping. *Mass Spectrom. Rev.* 9999: 1–13.

7. Wei, H., J. Mo, L. Tao, R. J. Russell, A. A. Tymiak, G. Chen, R. E. Iacob, and J. R. Engen. 2014. Hydrogen/Deuterium Exchange Mass Spectrometry for Probing Higher Order Structure of Protein Therapeutics: Methodology and Applications. *Drug Discov. Today* 19: 95–102.

8. Szymczak, L. C., H.-Y. Kuo, and M. Mrksich. 2017. Peptide Arrays: Development and Application. .

9. Esmaielbeiki, R., K. Krawczyk, B. Knapp, J.-C. Nebel, and C. M. Deane. 2016. Progress and challenges in predicting protein interfaces. *Brief. Bioinform.* 17: 117–131.

10. Yao, B., D. Zheng, S. Liang, and C. Zhang. 2013. Conformational B-Cell Epitope Prediction on Antigen Protein Structures: A Review of Current Algorithms and Comparison with

564 Common Binding Site Prediction Methods. *PLoS ONE* 8: e62249.

565 11. Weitzner, B. D., J. Jeliazkov, S. Lyskov, N. A. Marze, D. Kuroda, R. Frick, N. Biswas, and J. J.
566 Gray. 2016. Modeling and docking antibody structures with Rosetta. *bioRxiv* 069930.

567 12. Krawczyk, K., T. Baker, J. Shi, and C. M. Deane. 2013. Antibody i-Patch prediction of the
568 antibody binding site improves rigid local antibody–antigen docking. *Protein Eng. Des. Sel.*
569 26: 621–629.

570 13. Brenke, R., D. R. Hall, G.-Y. Chuang, S. R. Comeau, T. Bohnuud, D. Beglov, O. Schueler-
571 Furman, S. Vajda, and D. Kozakov. 2012. Application of asymmetric statistical potentials to
572 antibody-protein docking. *Bioinformatics* 28: 2608–2614.

573 14. Garzon, J. I., J. R. Lopéz-Blanco, C. Pons, J. Kovacs, R. Abagyan, J. Fernandez-Recio, and P.
574 Chacon. 2009. FRODOCK: a new approach for fast rotational protein–protein docking.
575 *Bioinformatics* 25: 2544–2551.

576 15. Ramírez-Aportela, E., J. R. López-Blanco, and P. Chacón. 2016. FRODOCK 2.0: fast
577 protein–protein docking server. *Bioinformatics* 32: 2386–2388.

578 16. Ghoorah, A. W., M.-D. Devignes, M. Smaïl-Tabbone, and D. W. Ritchie. 2013. Protein
579 docking using case-based reasoning: Protein Docking Using Case-Based Reasoning. *Proteins*
580 *Struct. Funct. Bioinforma.* 81: 2150–2158.

581 17. Bourquard, T., F. Landomiel, E. Reiter, P. Crépieux, D. W. Ritchie, J. Azé, and A. Poupon.
582 2015. Unraveling the molecular architecture of a G protein-coupled receptor/β-arrestin/Erk
583 module complex. *Sci. Rep.* 5.

584 18. Igel, C., T. Suttorp, and N. Hansen. 2006. A computational efficient covariance matrix
585 update and a (1+ 1)-CMA for evolution strategies. In *Proceedings of the 8th annual*
586 *conference on Genetic and evolutionary computation* ACM. 453–460.

587 19. Méndez, R., R. Leplae, M. F. Lensink, and S. J. Wodak. 2005. Assessment of CAPRI

588 predictions in rounds 3–5 shows progress in docking procedures. *Proteins Struct. Funct.*
589 *Bioinforma.* 60: 150–169.
590 20. Wiedemann, A., L. Mijouin, M. A. Ayoub, E. Barilleau, S. Canepa, A. P. Teixeira-Gomes, Y.
591 L. Vern, M. Rosselin, E. Reiter, and P. Velge. 2016. Identification of the epidermal growth
592 factor receptor as the receptor for Salmonella Rck–dependent invasion. *FASEB J.* 30: 4180–
593 4191.
594 21. Bourquard, T., J. Bernauer, J. Azé, and A. Poupon. 2009. Comparing Voronoi and Laguerre
595 tessellations in the protein-protein docking context. In IEEE. 225–232.
596 22. Bourquard, T., J. Bernauer, J. Azé, and A. Poupon. 2011. A collaborative filtering
597 approach for protein-protein docking scoring functions. *PLoS One* 6: e18541.
598 23. Wu, X., T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. S. Longo, M. Louder, K.
599 McKee, S. O'Dell, S. Perfetto, S. D. Schmidt, W. Shi, L. Wu, Y. Yang, Z.-Y. Yang, Z. Yang, Z.
600 Zhang, M. Bonsignori, J. A. Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, M. Simek, D. R.
601 Burton, W. C. Koff, N. A. Doria-Rose, M. Connors, J. C. Mullikin, G. J. Nabel, M. Roederer, L.
602 Shapiro, P. D. Kwong, and J. R. Mascola. 2011. Focused Evolution of HIV-1 Neutralizing
603 Antibodies Revealed by Structures and Deep Sequencing. *Science* 333: 1593–1602.
604 24. Ahmad, S., and K. Mizuguchi. 2011. Partner-Aware Prediction of Interacting Residues in
605 Protein-Protein Complexes from Sequence Data. *PLOS ONE* 6: e29104.
606 25. Dunbar, J., K. Krawczyk, J. Leem, C. Marks, J. Nowak, C. Regep, G. Georges, S. Kelm, B.
607 Popovic, and C. M. Deane. 2016. SAbPred: a structure-based antibody prediction server.
608 *Nucleic Acids Res.* 44: W474–W478.
609 26. Lee, J. U., W. Shin, J. Y. Son, K.-Y. Yoo, and Y.-S. Heo. 2017. Molecular Basis for the
610 Neutralization of Tumor Necrosis Factor α by Certolizumab Pegol in the Treatment of
611 Inflammatory Autoimmune Diseases. *Int. J. Mol. Sci.* 18.

612 27. Liang, S., J. Dai, S. Hou, L. Su, D. Zhang, H. Guo, S. Hu, H. Wang, Z. Rao, Y. Guo, and Z. Lou.
613 2013. Structural Basis for Treating Tumor Necrosis Factor α (TNFα)-associated Diseases with
614 the Therapeutic Antibody Infliximab. *J. Biol. Chem.* 288: 13799–13807.
615 28. Hu, S., S. Liang, H. Guo, D. Zhang, H. Li, X. Wang, W. Yang, W. Qian, S. Hou, H. Wang, Y.
616 Guo, and Z. Lou. 2013. Comparison of the Inhibition Mechanisms of Adalimumab and
617 Infliximab in Treating Tumor Necrosis Factor α-Associated Diseases from a Molecular View. *J.*
618 *Biol. Chem.* 288: 27059–27067.
619 29. Aaltonen, K. J., L. M. Virkki, A. Malmivaara, Y. T. Konttinen, D. C. Nordström, and M. Blom.
620 2012. Systematic Review and Meta-Analysis of the Efficacy and Safety of Existing TNF
621 Blocking Agents in Treatment of Rheumatoid Arthritis. *PLOS ONE* 7: e30275.
622 30. Danese, S., G. Fiorino, L. Peyrin-Biroulet, E. Lucenteforte, G. Virgili, L. Moja, and S.
623 Bonovas. 2014. Biological agents for moderately to severely active ulcerative colitis: a
624 systematic review and network meta-analysis. *Ann. Intern. Med.* 160: 704–711.
625 31. Shealy, D. J., A. Cai, K. Staquet, A. Baker, E. R. Lacy, L. Johns, O. Vafa, G. Gunn, S. Tam, S.
626 Sague, D. Wang, M. Brigham-Burke, P. Dalmonte, E. Emmell, B. Pikounis, P. J. Bugelski, H.
627 Zhou, B. J. Scallon, and J. M. Giles-Komar. 2010. Characterization of golimumab, a human
628 monoclonal antibody specific for human tumor necrosis factor α. *mAbs* 2: 428–439.
629 32. Mulleman, D., M. Ohresser, and H. Watier. 2013. *Introduction to Biological and Small*
630 *Molecule Drug Research and Development: Chapter 14. The case of anti-TNF agents*,. Elsevier
631 Inc. Chapters.
632
633

Figure legends

Figure 1: Principle and performance of the method

A: From the 3D structures of the antibody and the target, HEX generates docking poses and ranks them according to energetic criteria (H-rank). Each of the HEX top-500 docking poses is evaluated using both non-specific and specific scoring functions. A consensus score is computed, and the poses are ranked according to this score. In parallel, for each docking pose, the A-score, C-score and P-score (see hereafter) are computed, and the poses are ranked. The final ranking of poses is a consensus of the 5 different rankings (HEX, consensus, A-ranking, C-ranking, P-ranking). The top 30 solutions are used to compute, for each residue of the target, the frequency at which it appears within the epitope in these top 30 poses, and to design the epitope peptides. B: Ratio of residues of the epitope within the designed peptides identified using MAbTope on the complete test set (blue), cluspro (orange), frodock (green), epipred (violet) and PPiPP (purple). The values obtained when docking the false positives is shown in red. C: Ratio of epitope residues with the designed peptides identified using MAbTope on the complete test set (blue), on unique targets (green), on small targets (orange) and large targets (purple). The values obtained when docking the false positives on unique targets is shown in red.

Figure 2: Epitope peptides designed for 25 antibodies targeting HIV gp120.

A: Top-4 epitope peptides designed for the 25 antibodies of the benchmark targeting gp120. The designed peptides are all mapped on the sequence of 4ZMJ (chains B and G of GP120), although some of the antibodies target gp120 proteins of different clades. However, the epitopes recognized by the different antibodies have a homologous region in 4ZMJ. Each

coloured region represents one designed peptide; a black star indicates that the peptide belongs to the epitope. Red stars outside of these coloured regions indicate residues of the epitope that do not belong to a designed peptide. B: 3D structures of the complexes between the 25 antibodies (cartoon) and gp120 (surface). All the structures have been superimposed on 4ZMJ, the colour code is given in A.

Figure 3: Residues present at the interface of docking poses of the Golimumab-TNF and Certolizumab-TNF complexes.

A: Golimumab: dark blue: residues present at the interface of more than 20 poses, medium blue: 10 to 20 poses, light blue: 1 to 10 poses. Certolizumab: dark, medium and light violet. Validation peptides used in experiments are shown below the sequences. First best-ranking peptides predicted are boxed for each sequence. Red stars indicate the residues of certolizumab epitope in the crystal structure of the complex.

B and C: selected docking poses for the assembly of TNF-α with golimumab (B) and certolizumab (C). The three TNF-α monomers are shown in different shades of grey (light, medium and dark grey). The peptides selected for experimental validation are shown (P1 in red, P2 in orange, P3 in dark green, P4 in light green).

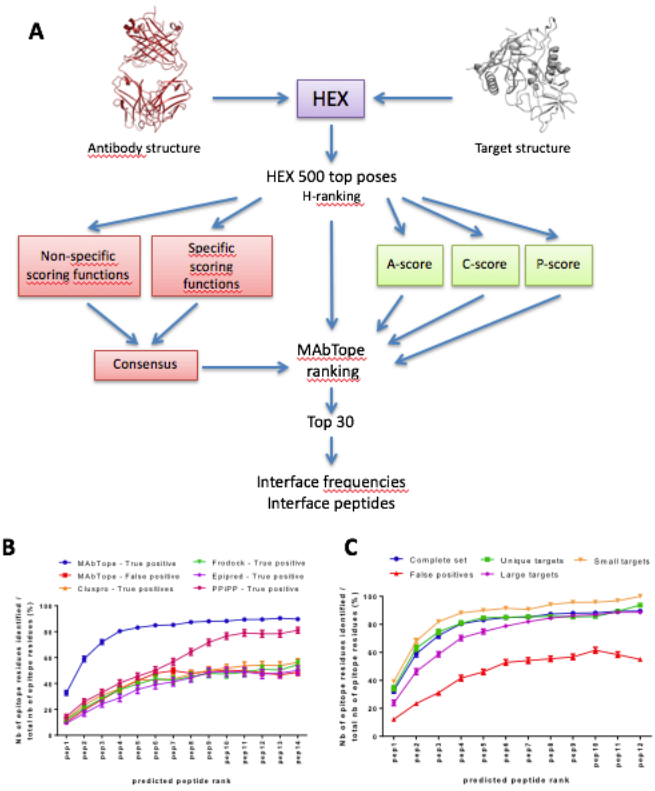
Figure 4: Validation of the predicted epitope of golimumab.

A. Certolizumab-induced displacement of Golimumab from AviTag-TNFα and thus bind the same epitope. The initial binding of Golimumab on AviTag-TNFα was measured at 0.1, 0.33 and 1 nM (brown, orange and red full symbol, respectively) using the HTRF mix anti-IgG-Tb /

679 streptavidin-d2. Increasing doses of Certolizumab were added, from 10^{-15} to 10^{-5} M (empty
680 symbols). The IC50 of the displacements are indicated above the graph.
681 B.C.D. Peptide-based validation assays of Golimumab epitope. B. The 11 peptides predicted
682 to belong to the epitope at 1 mM, biotinylated at their N-terminus, were incubated with 8 ng
683 of Golimumab or a non-relevant mAb and the HTRF mix anti-IgG-Tb / streptavidin-d2. The
684 HTRF signals obtained with Golimumab were corrected by the non-specific binding on the
685 irrelevant mAb considered as a baseline. C. The 11 predicted peptides and one control
686 peptide were spotted on a nitrocellulose-coated glass slide. After a blocking step, the slides
687 were incubated with 200 ng of fluorescently labeled Golimumab. RFU is calculated and used
688 to compare the interaction between the peptides and the Golimumab. D. Displacement of
689 the Golimumab from the AviTag-TNFα by the peptides. Selected non-biotinylated peptides at
690 1, 0.1 or 0.01 mM were incubated with golimumab and biotinylated AviTag-TNFα. The
691 complex formed by golimumab and TNF-α was detected by the HTRF mix anti-IgG-Tb /
692 streptavidin-d2. As a displacement control, Certolizumab was incubated with Golimumab /
693 TNFα complex. HTRF signals were subtracted of the signal obtained with the control
694 antibody (eculizumab).
695

696 **Figure 5: Predicted mutations abolish the binding of Golimumab on complete TNFα and**
697 **peptides.**
698 A. HEK293 cells were transfected with either a mock vector, the wild-type TNFα (Val 77 - Leu
699 233) or with 3 mutated TNFα constructions. The mutations were selected among the amino
700 acids whose side-chain is solvent-exposed within peptides P1-1, P3-1 and P4-1. Cells were
701 fixed, permeabilized and incubated with golimumab (0.33 µM). They were then stained by
702 detection of the Flag epitope fused to the different TNF-α constructs (PE-conjugated anti-

703 Flag antibody) (upper panels), or of golimumab using an anti-IgG coupled to APC as the
704 secondary antibody (lower panels). The plots show the side-scatter *vs* light intensity for both
705 channels.
706 B.C.D. Validation of the residues predicted to be implicated in the interaction between
707 Golimumab and TNFα. Mutated variants of the P3-1 (B) and P3-3 (C) peptides were designed
708 as indicated. Biotinylated peptides were incubated at 1 mM with 8 ng of either Golimumab
709 or a non-relevant mAb and the HTRF mix anti-IgG-Tb / streptavidin-d2. The HTRF signals
710 obtained with Golimumab were corrected by the non-specific binding on the irrelevant mAb
711 considered as a baseline. D. Summary of the residues found critical in the interaction
712 between Golimumab and TNFα. In red are indicated the residues obtained from the P3-1
713 mutants series and in blue the ones from the P3-3 mutants series.
714
715



716 Figure 1

717

718

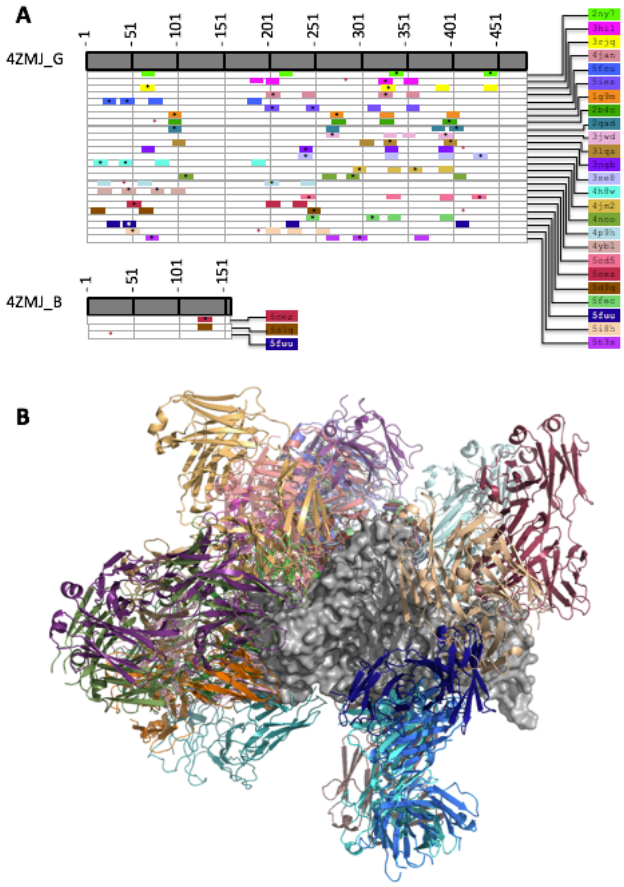


Figure 2

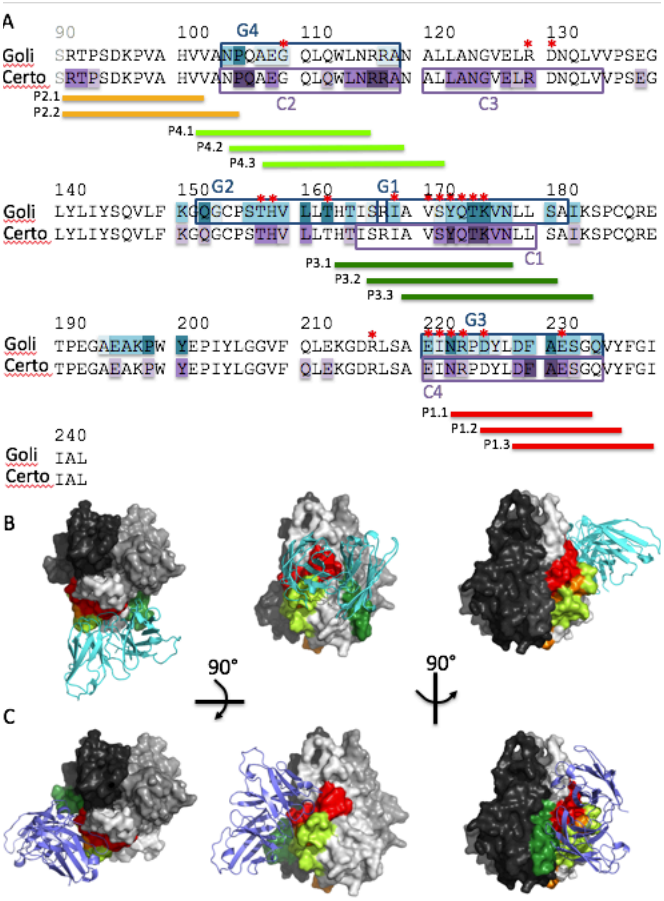

A
Goli
Certo
G4
C2
C3
G2
G1
C1
G3
C4
P2.1
P2.2
P4.1
P4.2
P4.3
P3.1
P3.2
P3.3
P1.1
P1.2
P1.3
B
90°
90°
C

Figure 3

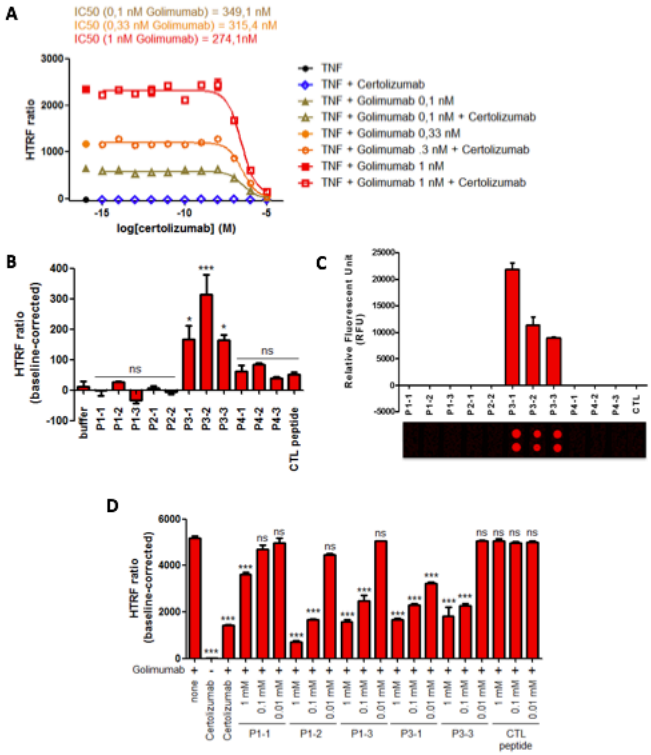

A
IC50 (0,1 nM Golimumab) = 349,1 nM
IC50 (0,33 nM Golimumab) = 315,4 nM
IC50 (1 nM Golimumab) = 274,1nM
HTRF ratio
log[certolizumab] (M)
TNF
TNF + Certolizumab
TNF + Golimumab 0,1 nM
TNF + Golimumab 0,1 nM + Certolizumab
TNF + Golimumab 0,33 nM
TNF + Golimumab .3 nM + Certolizumab
TNF + Golimumab 1 nM
TNF + Golimumab 1 nM + Certolizumab
B
HTRF ratio (baseline-corrected)
buffer
CTL peptide
C
Relative Fluorescent Unit (RFU)
D
HTRF ratio (baseline-corrected)
Golimumab
Certolizumab
P1-1
P1-2
P1-3
P3-1
P3-3
CTL peptide

Figure 4

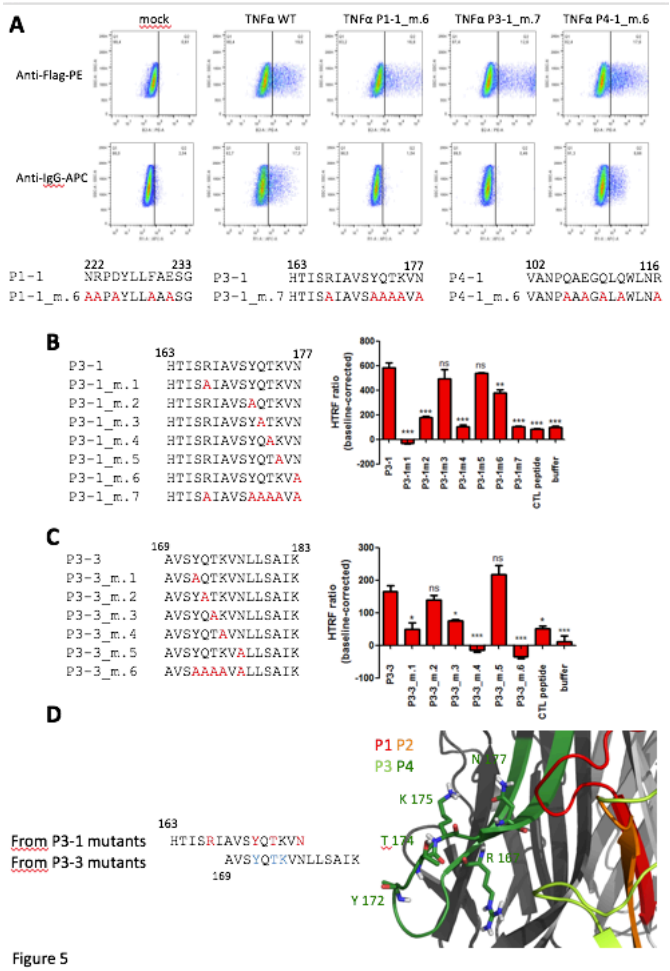

A
mock
TNFα WT
TNFα P1-1_m.6
TNFα P3-1_m.7
TNFα P4-1_m.6
Anti-Flag-PE
Anti-IgG-APC
P1-1 NRPDYLLFAESG
P1-1_m.6 AAPAYLLAAASG
P3-1 HTISRIAVSYQTKVN
P3-1_m.7 HTISAIAVSAAAAVA
P4-1 VANPQAEGQLQWLNR
P4-1_m.6 VANPAAAGALAWLNA
B
P3-1 HTISRIAVSYQTKVN
P3-1_m.1 HTISAIAVSYQTKVN
P3-1_m.2 HTISRIAVSAQTKVN
P3-1_m.3 HTISRIAVSYATKVN
P3-1_m.4 HTISRIAVSYQAKVN
P3-1_m.5 HTISRIAVSYQTAVN
P3-1_m.6 HTISRIAVSYQTKVA
P3-1_m.7 HTISAIAVSAAAAVA
HTRF ratio (baseline-corrected)
C
P3-3 AVSYQTKVNLLSAIK
P3-3_m.1 AVSAQTKVNLLSAIK
P3-3_m.2 AVSYATKVNLLSAIK
P3-3_m.3 AVSYQAKVNLLSAIK
P3-3_m.4 AVSYQTAVNLLSAIK
P3-3_m.5 AVSYQTKVALLSAIK
P3-3_m.6 AVSAAAAVALLSAIK
CTL peptide
buffer
D
From P3-1 mutants HTISRIAVSYQTKVN
From P3-3 mutants AVSYQTKVNLLSAIK
P1 P2
P3 P4
K 175
T 174
Y 172

Figure 5

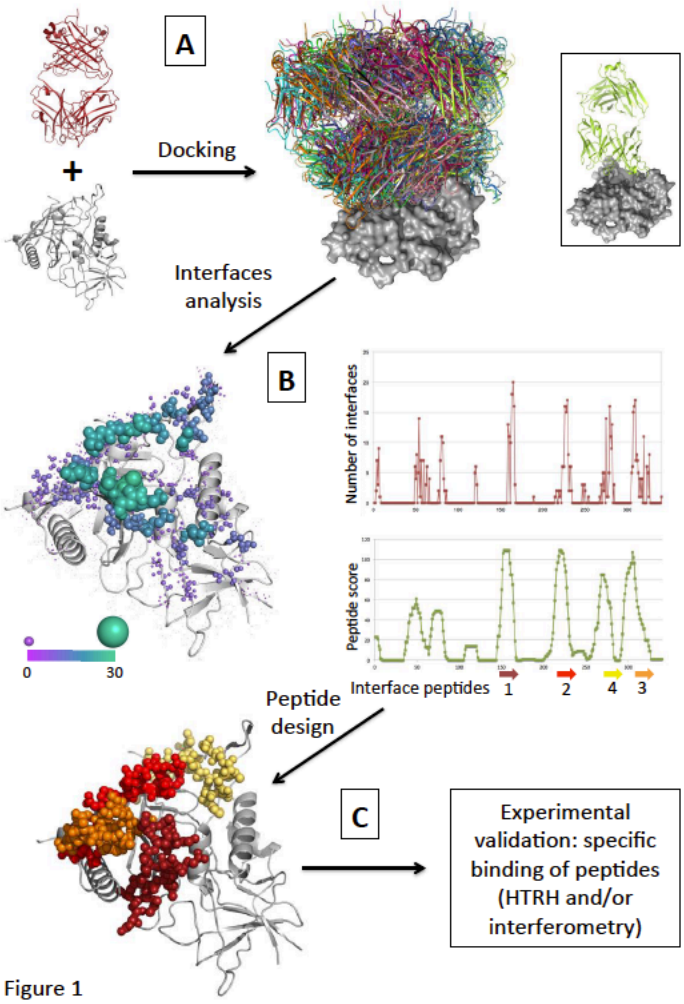


Figure 1

731

Figure S1

A: The isolated structure of the VRC03 antibody (PDB: 5JXA) is docked on the isolated structure of GP120 (PDB: 3TGT), the 30 poses best ranked by MAbTope are shown. For comparison, the structure of the crystallographic complex (PDB:3SE8 (23)) is shown in insert. B: For each amino acid of the target, we count the number of the top-30 poses in which this amino acid is found at the interface. This count is figured as spheres on the left, on the structure of the target, and as a graph on the top-right. For each 15-mer of the target, these counts are summed resulting in a score, shown as a graph down-right. C: This score is used to rank the peptides from highest to lowest score, and the 4 best-ranked peptides are defined as the interface peptides. Their specific binding to the antibody is then experimentally evaluated using HTRF and/or interferometry. The peptides can also be used to design point-mutations in the target.

732

733

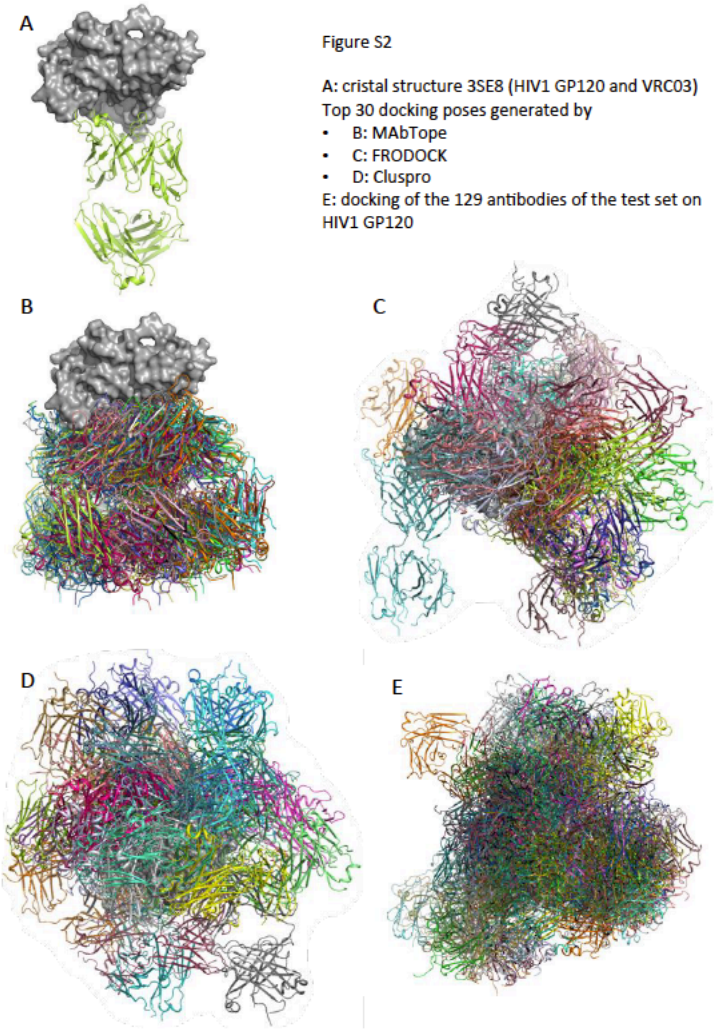


Figure S2

A: cristal structure 3SE8 (HIV1 GP120 and VRC03)
Top 30 docking poses generated by

- B: MAbTope
- C: FRODOCK
- D: Cluspro

E: docking of the 129 antibodies of the test set on HIV1 GP120

734

735

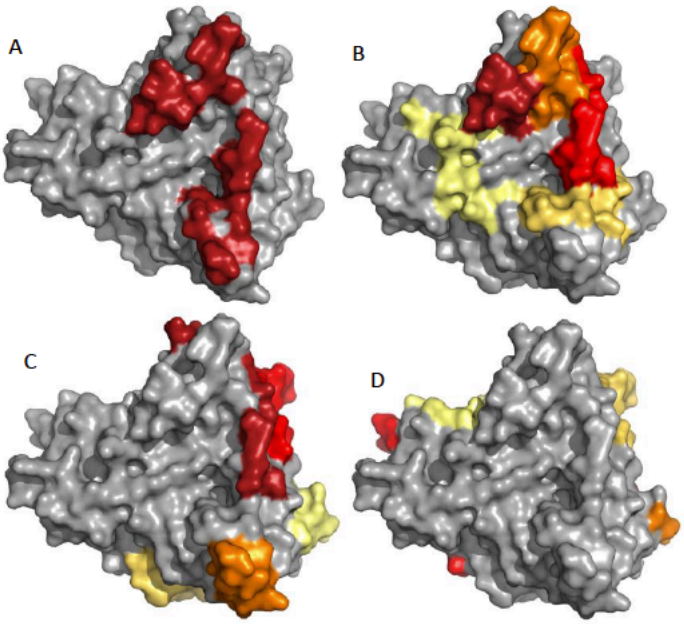
736

737

Figure S3

A: Epitope in cristal structure
Five first peptides predicted by:

- B: MAbTope
- C: FRODOCK
- D: Cluspro



738

739

740

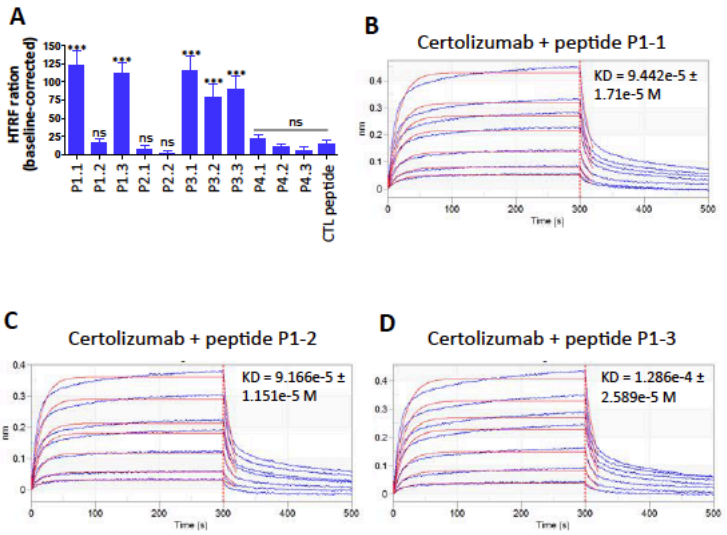


Figure S4: Predicted peptides bind the Certolizumab.
A. Biotinylated peptides were incubated at 1 mM with 8 ng of Certolizumab and the HTRF sensors mix anti-Fab-d2 / streptavidin-Tb. HTRF signal were corrected by the buffer.
B.C.D. BLI analysis of Certolizumab binding to P1-1 (B), P1-2 (C) and P1-3 (D) peptide. The analyte Fab was diluted at 100, 200, 400, 600, 800, 1000 and 1200 µg/ml. Raw data are shown in blue, and fits according to the 1:1 model in red.

741

742

743

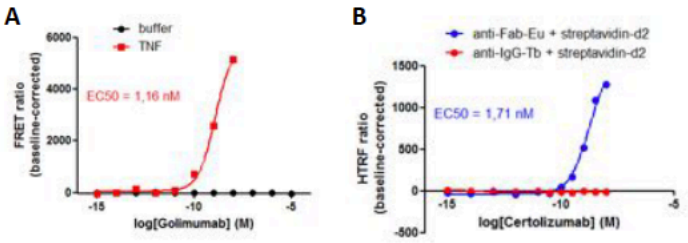


Figure S5: Golimumab and Certolizumab binding to TNFα are observed using HTRF. Golimumab (A) and Certolizumab (B) were incubated at indicated concentrations with 8 ng of biotinylated AviTag-TNFα and the HTRF sensors. The TNFα is sensored by the d2-coupled streptavidin, Golimumab by an anti-IgG1 coupled to Tb (A) and the Certolizumab by an anti-Fab coupled to Eu (B, blue curve). The Certolizumab is not detected by the anti-IgG-Tb sensor (B, red curve).

744

745

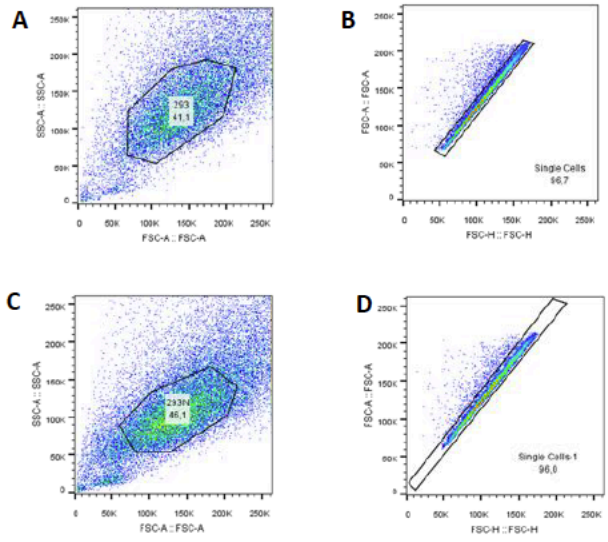


Figure S6: FACS gating strategy
A.B. Flag-PE-stained HEK293 cells
C.D. Golimumab and IgG-APC-stained HEK293 cells
A and C. The cells were selected in the core of the population in the foward vs side scatter plots. B and D. Single cells were selected form the FSC-A vs FSC-H plot.

746

747

748

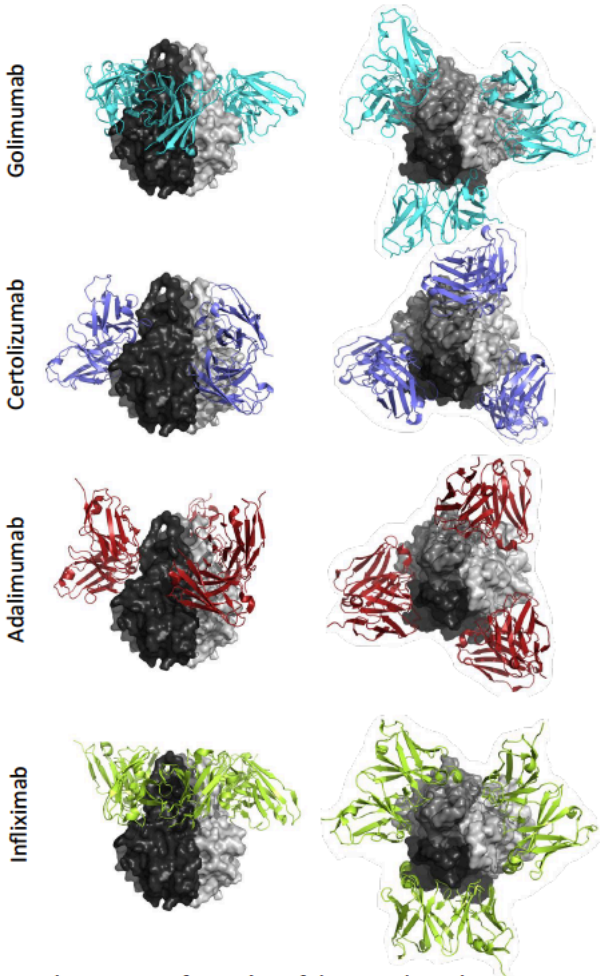


Figure S7: conformation of the complexes between TNF-α and four different antibodies.

749

750

751

Table S1: Comparision of the accuracies of epitope predictions using MAbTope, Clusprot, Frodock, EpiPred and PPiPP.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
MAbTope complete testset	0,33 (± 0.21)	0,59 (± 0.24)	0,72 (± 0.21)	0,80 (± 0.18)	0,83 (± 0.17)	0,85 (± 0.16)	0,85 (± 0.15)	0,87 (± 0.14)	0,88 (± 0.13)	0,88 (± 0.13)	0,89 (± 0.12)	0,89 (0.12)
MAbTope negative controls	0,10 (± 0.09)	0,19 (± 0.14)	0,28 (± 0.17)	0,36 (± 0.20)	0,42 (± 0.22)	0,48 (± 0.24)	0,50 (± 0.26)	0,48 (± 0.24)	0,49 (± 0.26)	0,50 (± 0.27)	0,50 (± 0.27)	0,48 (± 0.26)
MAbTope unique targets	0,34 (± 0.20)	0,63 (± 0.22)	0,75 (± 0.20)	0,81 (± 0.19)	0,85 (± 0.18)	0,85 (± 0.17)	0,85 (± 0.15)	0,86 (± 0.15)	0,85 (± 0.15)	0,86 (± 0.07)	0,89 (± 0.1)	0,94 (± 0.09)
MAbTope small targets	0,39 (± 0.19)	0,68 (± 0.20)	0,82 (± 0.14)	0,88 (± 0.11)	0,90 (± 0.11)	0,91 (± 0.08)	0,91 (± 0.08)	0,94 (± 0.06)	0,96 (± 0.06)	0,96 (± 0.05)	0,97 (± 0.04)	1,00 (NA)
MAbTope large targets	0,24 (± 0.21)	0,46 (± 0.24)	0,59 (± 0.21)	0,70 (± 0.21)	0,75 (± 0.19)	0,79 (± 0.18)	0,82 (± 0.17)	0,84 (± 0.15)	0,86 (± 0.14)	0,87 (± 0.14)	0,89 (± 0.12)	0,89 (± 0.11)
Cluspro complete test set	0,12 (± 0.20)	0,23 (± 0.30)	0,31 (± 0.33)	0,36 (± 0.34)	0,43 (± 0.35)	0,43 (± 0.35)	0,43 (± 0.34)	0,47 (± 0.33)	0,50 (± 0.33)	0,52 (± 0.32)	0,54 (± 0.34)	0,54 (± 0.33)
Frodock complete test set	0,12 (± 0.20)	0,21 (± 0.27)	0,28 (± 0.32)	0,35 (± 0.34)	0,39 (± 0.36)	0,44 (± 0.37)	0,43 (± 0.37)	0,45 (± 0.36)	0,48 (± 0.36)	0,47 (± 0.37)	0,49 (± 0.36)	0,51 (± 0.36)
EpiPred complete test set	0,09 (± 0.17)	0,17 (± 0.25)	0,24 (± 0.29)	0,29 (± 0.32)	0,35 (± 0.35)	0,39 (± 0.36)	0,41 (± 0.36)	0,44 (± 0.37)	0,49 (± 0.39)	0,49 (± 0.39)	0,49 (± 0.39)	0,47 (± 0.39)
PPiPP small targets	0,14 (± 0.20)	0,26 (± 0.25)	0,33 (± 0.28)	0,40 (± 0.28)	0,46 (± 0.29)	0,50 (± 0.31)	0,57 (± 0.29)	0,64 (± 0.29)	0,72 (± 0.28)	0,77 (± 0.28)	0,79 (± 0.29)	0,78 (± 0.30)

752