

## 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.

### ► To cite this version:

Thomas Bourquard, Astrid Musnier, Vincent Puard, Shifa Tahir, Mohammed Akli Ayoub, et al.. MAbTope: A Method for Improved Epitope Mapping. Journal of Immunology, 2018, 201 (10), pp.3096-3105. 10.4049/jimmunol.1701722 . hal-02373708

## HAL Id: hal-02373708 https://hal.science/hal-02373708v1

Submitted on 21 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# <sup>1</sup> MAbTope: a Method for Improved

# <sup>2</sup> Epitope Mapping<sup>1</sup>

| 3  | Running title: MAbTope: a New Method for Improved Epitope Mapping   |
|----|---|
| 4  | Thomas Bourquard <sup>*,†,2</sup> , Astrid Musnier <sup>*,‡,2</sup> , Vincent Puard <sup>‡</sup> , Shifa Tahir <sup>*</sup> , Mohammed Akli               |
| 5  | Ayoub $^{*,\$}$ , Yann Jullian $^{\P}$ , Thomas Boulo $^{*}$ , Nathalie Langonné $^{*,\parallel}$ , Hervé Watier $^{\parallel}$ , Gilles Bruneau $^{*}$ , |
| 6  | Eric Reiter <sup>*</sup> , Pascale Crépieux <sup>*</sup> , Anne Poupon <sup>*</sup>   |
| 7  |   |
| 8  | *: PRC, INRA, CNRS, Université François Rabelais-Tours, 37380, Nouzilly, France.  |
| 9  | $\dagger$ : Department of Human and Molecular Genetics, Baylor College of Medicine, Houston,  |
| 10 | Texas 77030, USA  |
| 11 | ‡: MAbSilico SAS, Domaine de l'Orfrasière, 37380 Nouzilly, France.  |
| 12 | $\S$ : Biology Department, College of Science, United Arab Emirates University, PO Box 15551 Al   |
| 13 | Ain, United Arab Emirates   |
| 14 | ¶: CaSciModOT, UFR de Sciences et Techniques, Université François-Rabelais, F-37041 Tours,  |
| 15 | France  |
|    |   |

<sup>1</sup> 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.

<sup>&</sup>lt;sup>2</sup> Equal contributions

- 16 || : CNRS; Université François-Rabelais de Tours; CHRU de Tours, UMR 7292, 37032 Tours,
- 17 France
- 18 Corresponding author:
- 19 Anne Poupon
- 20 Tel: +33 2 45 47 75 05
- 21 Fax: +33 2 45 47 77 43
- 22 email: Anne.poupon@inra.fr
- 23

#### 24 Abstract

25 Antibodies are very efficient drugs, about 70 of them are already approved for medical use, 26 over 500 are in clinical development, and many more are in preclinical development. One 27 important step in the characterization and protection of a therapeutic antibody is the 28 determination of its cognate epitope. The gold standard is the 3D structure of the antibody-29 antigen complex by crystallography or NMR. However, it remains a tedious task and its 30 outcome is uncertain. We have developed MAbTope, a docking-based prediction method of the epitope associated with straightforward experimental validation procedures. We show 31 32 that MAbTope predicts the correct epitope for each of 129 tested examples of antibody-33 antigen complexes of known structure. We further validated this method through the 34 successful determination of the epitopes recognized by two therapeutic antibodies targeting 35 tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ): certolizumab and golimumab.

#### 37 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 44 45 are based either on (i) site-directed mutagenesis; (ii) peptide arrays (3-5) or (iii) mass 46 spectrometry (6). Most peptide-based methods use 15-30 amino acids overlapping peptides 47 of the target arrayed on solid support, which are then exposed to the antibody (4, 5). This 48 identification of interacting peptides can then be completed by alanine-scanning in order to 49 define the epitope more precisely (3). In the mass spectrometry-based approach, the 50 antibody-antigen complex is subjected either to hydrogen/deuterium exchange (HDX-MS) 51 (7), or to enzymatic digestion, which allows differentiating target peptides that are 52 "protected" by the presence of the antibody. These peptides can then be identified using 53 mass spectrometry (see (6) for a review). It should be noted that even when successful, 54 these different approaches are likely to provide non-identical definitions of the epitope. 55 Indeed, because of the crystallisation step that freezes the complex structure in one out of 56 many possible conformations, X-ray structure identifies only the most stable interactions. 57 Alanine scanning does not allow identifying all the interacting residues for different reasons: 58 the mutated amino acid might interact with the antibody through its main chain or the 59 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.

62 However, these approaches are expensive, time-consuming and, except crystallography, 63 remain error-prone. Indeed, the results obtained through HDX-MS are sometimes very 64 difficult to interpret, for example when there is a conformational change in the target 65 between the free and complexed forms (7). Peptide arrays performance at identifying 66 epitopes are limited by different factors (8): immobilization methods, affinity of the peptides 67 and conformational constraints induced by the immobilization. For these reasons, many 68 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, 69 70 which aims at predicting the regions of a protein that are the most amenable of being 71 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 72 73 method leads to the prediction of the epitope for a given antibody, though B-cell epitope 74 prediction can be a useful first step in this process. Amongst the partner-specific approaches, 75 three main categories can be distinguished: predictors based on the intrinsic properties of 76 the partners, predictors based on co-evolution of the partners, and predictors based on 77 docking. However, few of these methods are dedicated to the special case of antibody-78 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

84 Rosetta (11) and Z-dock (12), but in local docking only, meaning that they require a partial 85 knowledge of the epitope. The introduction of sDARS, a pairwise statistic potential specific of 86 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-87 88 native solution in the top-10 predicted conformations. The particularity of this statistic 89 potential as compared to previously used ones is that it accounts for the asymmetry of the 90 antibody-antigen interaction. Another example of a web-server specific for antibody-antigen 91 docking is Frodock (14, 15). Frodock uses spherical harmonics for conformation generation 92 (as opposed to fast-Fourier transform for most other algorithms, including PIPER), and a 93 combination of energetic (van der Waals, electrostatics and desolvation) and knowledge-94 based potentials, optimized for the different categories of complexes (enzyme, antibodies 95 and others). However, the goal of these methods is predicting the conformation of the 96 assembly, meaning predicting the interaction region, but also the precise relative 97 orientations of the two partners, and not predicting the epitope. Even though they perform 98 better at this task than the other types of epitope prediction methods, they are not 99 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.

#### 108 Material and methods

#### 109 Overview of the method

The 3D structures of the antibody and of the target are used as input of the Hex software 110 (16) (Figure 1). Hex generates more than 10<sup>8</sup> docking poses and ranks them according to 111 112 energetic criteria (H-ranking). Each of the Hex top-500 docking poses is evaluated using 30 113 specific and 30 non-specific scoring functions. The non-specific scoring function is identical 114 to the one used in PRIOR (17), the specific scoring function has been re-optimised, using the 115 learning dataset described hereafter, and the same machine-learning procedure: genetic 116 algorithm and CMA-ES (18), and in both cases the area under the ROC-curve is used as 117 fitness function. A consensus score is then computed using the formula:

118 
$$CONS(i) = \frac{k_b}{2} \cdot \left[ ln \left( \frac{\sum_{j=1}^{30} rank_j^{NS}(i)}{\sum_{j=1}^{30} rank_j^{NS}(0)} \right) + ln \left( \frac{\sum_{j=1}^{30} rank_j^{S}(i)}{\sum_{j=1}^{30} rank_j^{S}(0)} \right) \right]$$

119 Where  $rank_j^{NS}(i)$  is the ranking of pose *i* according to the non-specific function *j*, and 120  $\sum_{j=1}^{30} rank_j^S(i)$  is the ranking of pose *i* according to the specific function *j*. The rankings of 121 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 Cscore is the sum of the ranks of the other poses that have a RMSD value lower than 5 Å with this particular pose.

128 The consensus, Hex-rank, A-score, C-score and P-score are used to generate 5 different 129 rankings. For each pose, the sum of its ranks in the different rankings is computed. These 130 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).

134 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.

**139** Specific learning dataset

140 The learning dataset is composed of 393 non-redundant antibody-target complexes 141 manually extracted from the PDB in January 2015. Only the complexes in which the target is 142 larger than 40 residues were considered. These complexes contain 392 distinct Abs and the 143 targets belong to 165 distinct Pfam families. The definition of non-redundancy we use is 144 weaker than what is usually used, since antibodies are very special proteins, and overall 145 sequence identity, even restricted to the variable domain, is not indicative of the antibody 146 specificity, and consequently on its ability to form a complex with its target. The criteria 147 retained for considering two Ab-Ag complexes as non-redundant were: (i) targets are not 148 related (they belong to different Pfam families); or (ii) targets are related but the epitopes 149 recognized by the antibodies have less than 20% overlap; or (iii) targets are related and 150 epitopes are overlapping but the CDRs of the considered antibody differ in 10 or more 151 positions. This third criterion is justified by the fact that most pairs of antibodies differing by 152 10 or more residues within the CDRs, even when they present a very high overall sequence 153 identity, do not share the same target.

#### 154 **Test dataset**

155 To evaluate the performance of the method, a test dataset has been designed. It consists in 156 the 82 complexes of the learning dataset for which the 3D structures of the individual 157 partners are known. For the evaluation, the learning has been done in leave-one-out, 158 meaning that the epitope of a given antibody is predicted using a scoring function learnt on 159 a dataset not containing the 3D structure of the complex it forms with its target. Forty-seven 160 new complexes whose 3D structure has been determined after January 2015, and which 161 were non-redundant with those already present in the learning dataset have been added to 162 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.

#### 166 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.

#### 170 Epitope definition

171 In this work, an amino acid of a protein targeted by an antibody will be considered as 172 belonging to the epitope if at least one of its atoms is at less than 4 Å of an atom belonging 173 to an amino acid of the antibody. These distances are computed on the crystallographic 174 structure of the complex.

#### **175 Definition of epitope peptides**

176 Each amino acid of the target is attributed a value, which is the number of poses within the 177 30 top-ranked ones in which this amino acid belongs to the predicted epitope. Different sizes 178 of pose sets have been tested, and 30 is a satisfactory compromise (data not shown). Each 179 15 amino acid peptide of the sequence is then given a score equal to the sum of these values 180 for each amino acid in the peptide. The peptides are then ranked along this score. Peptides 181 overlapping by at least 8 amino acids with a better-ranked peptide are ignored. For 182 benchmarking, the relevance of a given peptide is evaluated by the number of residues that 183 belong to the crystallographic epitope. This definition of epitope peptide was also used for 184 the testing of Cluspro and FRODOCK. In EpiPred predictions, amino acids present in the first 185 predicted epitope were given a score of 3, a score of 2 for the amino acids of the second 186 epitope and a score of 1 for the amino acids of the third epitope. In PPiPP predictions, the 187 scores given in the program output were considered. Epitope peptides were then built as 188 explained above.

189 The choice of 15 mers is a compromise between two empirical observations we have made 190 along the development of this method. (i) Shorter peptides tend to give a poor signal. Our 191 hypotheses are that they are too flexible (short peptides present less long-range interactions 192 and are thus more flexible), which decreases their binding to the antibody. Moreover, the 193 secondary structure is important for binding, and very short peptides have no chance to 194 adopt hairpin or strand conformations. (ii) Longer peptides tend to span over more than one 195 loop, and interpretation of experimental results is then more difficult. A second aspect is 196 that longer peptides have a higher tendency to precipitate.

#### **197** Evaluation criteria

198 MAbTope first output is a ranked list of docking poses. To evaluate the distance between 199 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)];</li>

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]

212 Introducing this new category is very useful for evaluating docking performance in the 213 perspective of epitope determination. Indeed, the docking poses falling in this category, 214 even though their geometry is too distant from the crystal structure to be considered as 215 acceptable by the Capri criteria, still define an interaction area on the target that overlaps 216 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)

226 All measurements were performed with the Octet RED96 System (Pall Forte Bio, Fremont, 227 CA, USA), in the manufacturer kinetics buffer, at 30 °C, with shaking at 1000rpm. 228 Biotinylated peptides were immobilized during 200 seconds on streptavidin-coated sensors 229 (SA) at 0.5, 1 and 5 µg/mL for P1-3, P1-1 and P1-2 respectively and left for equilibration for 230 120 seconds in kinetics buffer. Typical capture variability within a row of eight tips did not 231 exceed 0.1 nm. Binding was assessed at 100, 200, 400, 600, 800, 1000 and 1200 µg/mL 232 certolizumab for 300 seconds. Two parallel corrections were carried out by subtracting the 233 association of certolizumab on an immobilized non-relevant biotinylated peptide, and by 234 subtracting the loading baseline drift on non-associated sensors. Data were analyzed using 235 Octet Software 9.0 version. Since certolizumab is a Fab', experimental data were fitted with 236 the binding equation describing a 1:1 interaction. Considering the weak affinity of peptides 237 for the antibody and the fact that the dissociation is almost immediate, we restrained the 238 dissociation analysis to the 20 first seconds. Global analyses of the datasets assuming that 239 binding was reversible (full dissociation) were carried out using nonlinear least-squares 240 fitting, allowing a single set of binding parameters to be obtained simultaneously for all 241 concentrations used in each experiment.

#### 242 HTRF-based competition assay

243 The competition between Golimumab and either the Certolizumab or the peptides for the TNFα was assessed *in vitro* using an HTRF<sup>®</sup>-based assay in 384-well plate. The Golimumab 244 245 and the Certolizumab were kindly provided by Denis Mulleman (CHRU Bretonneau, Tours, 246 France). The Golimumab was incubated at 0.1, 0.33 and 1 nM with 8 ng of TNFa 247 (NP 000585.2, Val77-Leu233) N-terminally fused to the AviTag® (Avidity LLC, Aurora, CO, 248 USA) purchased from ACROBiosystems (Newark, DE, USA) in 10 µl of PPI - Terbium detection 249 buffer (CisBio Bioassays, Condolet, France). Five microliters containing either 4 mM of non-250 biotinylated peptides (GeneCust, Dudelange, Luxembourg) or 4 µM of Certolizumab were 251 added. The HTRF-compatible fluorophore Terbium cryptate and d2 conjugated to either an 252 anti-Fc Ab or the streptavidin (from CisBio Bioassays, Condolet, France) were finally added in 253 5  $\mu$ l. After 1h incubation at room temperature, the fluorescence at 620 nm and 665 nm were 254 measured on the TriStar<sup>2</sup> LB 942 microplate reader (Berthold Technologies GmbH & Co, 255 Wildbad, Germany). Data were expressed as the emission ratio 665 nm / 620 nm subtracted 256 by the non-specific signal obtained without Ab nor peptide.

#### 257 Interaction measurement by peptide array

#### 258 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 subarrays on a nitrocellulose coated glass slide (ONCYTE<sup>®</sup> Film slides, Grace Bio-Labs, USA) using a Nano-Plotter (GeSIM, Germany). Slides are dried overnight at room temperature.

265 *Preparation of antibodies:* 

Golimumab is fluorescently labelled with iFluor<sup>™</sup> 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.

271 Incubation:

272 Slides are mount with Pro-Plate<sup>®</sup> chamber (Grace Bio-Labs, USA) for the following steps. 273 Slides are hydrated with 150 µl per well of PBS-T solution for 15 min under agitation on a 274 seesaw rocker. PBS-T is removed and 100 µl of Super G blocking buffer (Grace Bio-Labs, USA) 275 is added for 1h incubation on a seesaw rocker. After removing the blocking buffer, 100 µl per 276 well of antibodies diluted in PBS-t supplemented with 1 % BSA (corresponding to 200 ng) are 277 added for 2 h incubation on a seesaw rocker. Then, antibodies are removed and slides are 278 washed two times with PBS-T for 5 min and once with PBS (150  $\mu$ l/well). Finally, slides are 279 rinsed with filtered water for one minute and air-dried.

280 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).

#### 288 In vitro FRET binding measurement

The interaction between the different biotinylated peptides (GeneCust, Dudelange, 289 290 Luxembourg) and Certolizumab, or Golimumab, or Eculizumab used as a negative control, 291 was assessed by HTRF<sup>®</sup>. All experiments were performed in PPI-Terbium or -Europium 292 detection buffers (CisBio Bioassays, Condolet, France). For this, 5 µL of biotinylated peptides 293 (4 mM) were first incubated with 5  $\mu$ L of either of the mAbs (1.6  $\mu$ g/mL) for 1 hr at room 294 temperature. Then, 5 µL of streptavidin and 5 µL of anti-Fab (for Certolizumab) or anti-Fc 295 (for Golimumab and Eculizumab) antibodies conjugated with HTRF compatible fluorophores, 296 Terbium or Europium cryptate and d2, were added in quantities recommended by the 297 manufacturer. After an overnight incubation at 4°C, the fluorescence emissions at 620 nm 298 and 665 nm were measured using the appropriate HTRF program on a TriStar<sup>2</sup> LB 942 299 Modular microplate reader (Berthold Technologies GmbH & Co. Wildbad, Germany). Data 300 are represented as specific FRET signals calculated as the 665 nm/620 nm emission ratio 301 subtracted of the binding on the non-relevant Ab.

302

#### **303** Golimumab binding on mutant TNFα by flow cytometry

304 Three TNF $\alpha$  mutants were designed starting from the sequence NP\_000585.2 by 305 incorporating the mutations predicted to alter the interaction with Golimumab according to 306 our docking solution. The mutant TNFa constructions contain the following mutations: 307 TNFα\_P1-1m6 (N222A, R223A, D225A, F229A, E231A, Q234A), TNFα\_P3-1m7 (R167A, Y172A, 308 Q173A, T174A, K175A, N177A), and TNFα P4-1m6 (Q106A, E108A, Q110A, Q112A, R116A). 309 The cDNA of the 3 mutants and the wild-type TNF $\alpha$  fused to a Flag tag on their N-terminus 310 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 311

312 (Piscataway, NJ 08854, USA). HEK293N cells were transiently transfected with the TNF $\alpha$ 313 constructions or a mock vector using Metafectene (Biontex Laboratories GmbH, München, 314 Germany) according to manufacturer's instructions. Thirty hours after transfection, the cells 315 were fixed and permeabilized according to the BD Cytofix/Cytoperm kit instructions (BD 316 Biosciences, San Jose, CA, USA). All the following hybridations were performed in the kit's 317 perm/wash buffer. Five hundred thousand cells of each transfected population were 318 incubated with 5  $\mu$ g of Golimumab for 1 hr at room temperature and washed once in 2 ml 319 buffer. The binding of Golimumab was assessed with the allophycocyanin (APC)-labelled 320 anti-IgG1 antibody from Miltenyi Biotech (Bergisch Gladbach, Germany) diluted to 1:100. 321 The expression level of each of the constructions was evaluated with and anti-Flag Ab 322 coupled to phycoerythrin (PE) also from Miltenyi Biotech (Bergisch Gladbach, Germany). 323 After staining, all the cells were washed once in 2 ml working buffer and once in 2 ml PBS-324 EDTA 2 mM and finally suspended in 200 µl of PBS-EDTA 2 mM. The fluorescence was 325 assessed with the MACSQuant Analyzer 10 (Miltenyi Biotec, Bergisch Glabach, Germany) and 326 the data analyzed with FlowJo software (FlowJo LLC, Ashland, OR, USA).

327

328 Binding kinetics of Certolizumab to biotinylated peptides using biolayer interferometry 329 (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 300 336 seconds. Two parallel corrections were carried out by subtracting the association of Certolizumab on an immobilized non-relevant biotinylated peptide, and by subtracting the 337 338 loading baseline drift on non-associated sensors. Data were analyzed using Octet Software 339 9.0 version. Since Certolizumab is a Fab', experimental data were fitted with the binding 340 equation describing a 1:1 interaction. Considering the weak affinity of peptides for the 341 antibody and the fact that the dissociation is almost immediate, we restrained the 342 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 343 344 fitting, allowing a single set of binding parameters to be obtained simultaneously for all 345 concentrations used in each experiment.

346

#### 347 Statistical analysis

348 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.

#### 350 **Results**

#### 351 Principle and Benchmarking

352 MAbTope involves three successive steps. The first step is the docking of the antibody on its 353 target, which results in the generation of docking poses (possible conformations of the 354 antibody-antigen complex), through a method related to PRIOR, a general protein-protein 355 docking method we had previously developed (17, 21, 22). The second step is the ranking of 356 these docking poses in order to extract 30 poses that tile the epitope, and the design of four 357 so-called interacting peptides, that is, peptides predicted to be part of the epitope. The third 358 step is the experimental validation based on the interacting peptides. Different methods can 359 be used: measurement of the binding of each of these four peptides with the antibody, 360 competition for antibody binding between the peptides and the target, or measurement of 361 the binding of target mutated on residues belonging to these peptides.

362 The design of the interacting peptides from the docking poses is crucial for the success of the 363 method. At this step, all the possible 15 amino acid-long peptides of the target are ranked 364 according to the frequency at which their amino acids are found within the epitope in the 30 365 top-ranked docking poses. MAbTope predicts a correct peptide, that is, a peptide that 366 contains residues belonging to the crystallographic interface, within the 4 best-ranked ones 367 for all of the 129 complexes tested. On average the 4 best-ranked peptides contain more 368 than 80% of the epitope residues, and the minimum is 30%, meaning that the epitope is at 369 least partly found for all complexes in the test set (Figure 1A, Table S1). As a control, each 370 antibody of the test set was docked to all the targets of the other antibodies. In this test, on 371 average only 36 % of the residues belonging to the epitope of the specific antibody are 372 found within the 4 best-ranked peptides.

373 MAbTope performs much better than Cluspro or FRODOCK at predicting the epitopes, as 374 they identify, within the 4 best-ranked peptides, 36 % and 35 % of the epitope residues 375 respectively. One reason is that, in MAbTope, the 30 top docking poses are centred on the 376 correct epitope, and not distributed on the whole surface of the target. This is illustrated in 377 Figure 2 (see also Figures S1, S2) by the example of the complex between the HIV gp120 378 glycoprotein and the VRC03 antibody (PDB 3SE8) (23). This particularity, which can be found 379 for all of the tested examples, arises for two main reasons. First, for conformation 380 generation, we use Hex with very restrictive angle parameters, the obtained poses are 381 consequently already well focused. Second, the A-score (as defined in materials and 382 methods) favours over-represented poses, and consequently decreases the diversity of the 383 top-ranked poses. As a result, the amino acids constituting the epitope are almost all found 384 in more than half of the 30 selected docking poses. Consequently, the four best ranking 385 peptides all contain amino acids belonging to the interface (Figure S3). In addition, peptides 386 1 and 2 contain 7 and 6 amino acids belonging to the epitope, respectively. It should be 387 noted that peptides 3 and 4 also contain 8 and 6 residues, respectively, and can also be 388 considered as good predictions. Finally, the 6 best-ranking peptides contain all the amino 389 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,

397 three peptides are synthesized, all of the same length but sliding three amino acids along the 398 sequence. The first one starts and ends 3 amino acids upstream of the designed epitope 399 peptide, the second one corresponds to the designed one, and the third one starts and ends 400 3 amino acids downstream. This choice was made to overcome the issue of some peptides 401 being insoluble. A second approach is to measure the competition between these peptides 402 and the target for the binding of the antibody. Finally, as the residues present within these 403 peptides are those predicted to belong to the epitope, they can be used to predict point 404 mutations of the target reducing the binding of the antibody.

405 It should be highlighted that MAbTope is able to find the epitope of each antibody, and not 406 only the most antigenic sites on the target protein as defined by B-cell epitope prediction 407 methods. This is well illustrated by the example of gp120, to which 25 antibodies of the 408 benchmark bind. Whereas some regions of gp120 are targeted by a large number of 409 antibodies, including some that do not belong to the benchmark since the structure of the 410 isolated antibody is not known, other regions are also targeted. Accordingly, the interaction 411 peptides designed through MAbTope are spread on the whole target sequence (Figure 2). 412 MAbTope correctly builds at least one correct peptide for each of these 25 antibodies, and 413 two peptides for 19.

#### 414 Validation on golimumab and certolizumab

To validate the method, we next predicted the epitopes of two therapeutic antibodies targeting Tumor Necrosis Factor  $\alpha$  (TNF- $\alpha$ ): 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.

After the initial submission of this paper, the structure of the complex between certolizumab and TNF- $\alpha$  has been published (26). Comparison with our prediction shows that out of the 20 residues constituting the epitope, 17 belong to peptides C1 to C4 (Figure 3A). This shows that certolizumab epitope can be considered as conformational since it involves residues belonging to five different peptides. Nevertheless, we are still able to show the specific binding of some of these peptides to the antibody through HTRF and interferometry (Figure S4).

435 To validate the epitope of golimumab, we first have shown that it competes with 436 certolizumab for the binding to TNF- $\alpha$ , 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- $\alpha$  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 experimentspresented hereafter.

447 To further validate, we mutated in TNF- $\alpha$  the residues belonging to peptides of series 1, 3 and 4 to alanines, and observed the binding of golimumab using flow-cytometry (Figure 5A 448 and S6). We observed that each TNF- $\alpha$  construct expressed well in cells by detecting flag 449 450 epitope that was added to all constructs. Interestingly, we found that the binding of 451 golimumab its target was almost abolished when the TNF- $\alpha$  was mutated at positions 452 indicated within P1 and P3 series, and reduced by 50% for mutations within the P4 series 453 peptides. Finally, for peptides P3-1 and P3-3, which gave the best signals in HTRF, we 454 mutated individually the residues belonging to these peptides and whose side-chains are 455 exposed, and measured the binding to golimumab using HTRF (Figure 5C and 5D). These 456 results show that, as predicted, residues Y172, T174 and K175 are essential for golimumab 457 binding to TNF- $\alpha$ .

#### 458 **Discussion**

459 The results obtained on the 129 antibody-target complexes of the benchmark show that the 460 in silico prediction is robust, since within the benchmark, the predicted peptides contain on 461 average 80 % of the epitopes residues. This number is not much affected by the type of 462 epitope: 79 % for conformational epitopes (105 out of 129), 89% for linear epitopes (14 out 463 of 129). Neither is it much affected by the size of target: 88 % for targets up to 300 residues 464 long, 70 % for larger targets. The main limitation of the *in silico* step is that the 3D structure 465 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 466 467 few tested cases, this requires further investigations.

468 Based on the designed peptides, we present three different experimental validations of the 469 predicted epitope. Our first approach consists in measuring the direct binding of the 470 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 471 472 observed for series 1 peptides, although we later demonstrate that these peptides belong to 473 the epitope. The second approach consists in making a competition between the peptides 474 and the target for the binding of the antibody. Using this method we were able to validate 475 the peptides of series 1, and confirm the peptides of series 3. Nevertheless, both approaches 476 are limited by the fact that some peptides tend to be "sticky". Another limit to these 477 approaches is the solubility of peptides, which is not always sufficient.

478 Importantly, the interaction peptides can also be used to design point-mutations in the 479 target potentially decreasing the affinity of the antibody. In the TNF- $\alpha$  we mutated to 480 alanines the residues belonging to peptide series 1, 3 and 4 whose side-chains point towards 481 the solvent. We show using flow cytometry that these mutations indeed abolish (series 1 482 and 3) or decrease (series 4) the binding of the antibody. However, this approach also has its 483 limitations: the difficulty of expressing some target or their mutated forms, especially if they 484 are toxic for the cells. The endogenous expression of the native target could also raise some 485 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- $\alpha$ : 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- $\alpha$ -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 performed in ulcerative colitis indicated that infliximab is better than adalimumab and 502 probably golimumab (30). Their affinities for TNF- $\alpha$  (4.5 x 10<sup>-10</sup> M for infliximab (28), 7.05 x 503  $10^{-11}$  M for adalimumab (28), 1.8 x  $10^{-11}$  M for golimumab (31) and 1.32 x  $10^{-10}$  M for 504 505 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 consequently a higher overlap with the TNF- $\!\alpha$  receptor binding interface and a better 508 509 neutralizing activity, than infliximab, which binds to a monomer. By contrast, the ability to 510 target inflammatory cells expressing membrane TNF- $\alpha$ , which could be monomeric, and to 511 induce apoptotic signals seems important determinants of therapeutic activity of anti- TNF- $\alpha$ 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- $\alpha$  therapeutic antibodies is now known will help at understanding the subtle differences in their clinical activities.

518

519 Conclusion

520 In conclusion, MAbTope initial prediction of the epitope is very robust. On a benchmark of 521 129 antibody-antigen complexes, MAbTope correctly defines the epitope in each case. In 522 addition, MAbTope allows defining four 15 amino acid peptides, among which at least one 523 belongs to the epitope, which in turn allows experimental validation. These peptides also 524 allow the design of point mutations that can be used to validate and refine the predicted 525 epitope. Although the information obtained through MAbTope does not allow defining the 526 precise interactions taking place between the antibody and the target, it allows defining with 527 good precision the region of the target involved in the interaction. This information is 528 sufficient for understanding the mechanism of action of the antibody, a crucial step in the 529 development of therapeutics, but also diagnostic or biotechnological tools. Taken together, 530 MAbTope is not just a prediction method, but constitutes an integrated workflow allowing 531 identification of the epitope. With the example of two therapeutic antibodies, certolizumab 532 and golimumab, we show that it can be successfully applied to antibodies whose 3D 533 structure is unknown.

534

#### 535 Acknowledgements

536 We thank Prof. Denis Mulleman for providing certolizumab and golimumab. We thank Dr.537 Olivier Lichtarge for reading and advices.

538

#### 540 **Bibliography**

- 541 1. Watier, H., and Reichert, Janice. 2017. Evolution of antibody therapeutics (Chapter 2). In
  542 *Protein Therapeutics*, Jallal B, Vaughan T, Osbur J. Wiley. 25–50.
- 543 2. Van Regenmortel, M. H. V. 2014. Specificity, polyspecificity, and heterospecificity of 544 antibody-antigen recognition. *J. Mol. Recognit.* 27: 627–639.
- 545 3. Cunningham, B. C., and J. A. Wells. 1989. High-resolution epitope mapping of hGH-546 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.
- 548 Nilsson, E. P. Hudson, J. Rockberg, and M. Uhlen. 2014. Proteome-wide Epitope Mapping of
- 549 Antibodies Using Ultra-dense Peptide Arrays. *Mol. Cell. Proteomics* 13: 1585–1597.
- 550 5. Hansen, C. S., T. Østerbye, P. Marcatili, O. Lund, S. Buus, and M. Nielsen. 2017.
- 551 ArrayPitope: Automated Analysis of Amino Acid Substitutions for Peptide Microarray-Based
- 552 Antibody Epitope Mapping. *PLOS ONE* 12: e0168453.
- 553 6. Opuni, K. F., M. Al-Majdoub, Y. Yefremova, R. F. El-Kased, C. Koy, and M. O. Glocker. 2016.
- 554 Mass spectrometric epitope mapping. *Mass Spectrom. Rev.* 9999: 1–13.
- 555 7. Wei, H., J. Mo, L. Tao, R. J. Russell, A. A. Tymiak, G. Chen, R. E. Iacob, and J. R. Engen. 2014.
- 556 Hydrogen/Deuterium Exchange Mass Spectrometry for Probing Higher Order Structure of
- 557 Protein Therapeutics: Methodology and Applications. *Drug Discov. Today* 19: 95–102.
- 558 8. Szymczak, L. C., H.-Y. Kuo, and M. Mrksich. 2017. Peptide Arrays: Development and 559 Application.
- 560 9. Esmaielbeiki, R., K. Krawczyk, B. Knapp, J.-C. Nebel, and C. M. Deane. 2016. Progress and 561 challenges in predicting protein interfaces. *Brief. Bioinform.* 17: 117–131.
- 562 10. Yao, B., D. Zheng, S. Liang, and C. Zhang. 2013. Conformational B-Cell Epitope Prediction
- 563 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
- antibody binding site improves rigid local antibody–antigen docking. *Protein Eng. Des. Sel.*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/ $\beta$ -arrestin/Erk
- 583 module complex. Sci. Rep. 5.
- 18. Igel, C., T. Suttorp, and N. Hansen. 2006. A computational efficient covariance matrix
  update and a (1+ 1)-CMA for evolution strategies. In *Proceedings of the 8th annual 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.

S90 20. Wiedemann, A., L. Mijouin, M. A. Ayoub, E. Barilleau, S. Canepa, A. P. Teixeira-Gomes, Y.
S91 L. Vern, M. Rosselin, E. Reiter, and P. Velge. 2016. Identification of the epidermal growth
S92 factor receptor as the receptor for Salmonella Rck–dependent invasion. *FASEB J.* 30: 4180–
S93 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.
- 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.
- 26. Lee, J. U., W. Shin, J. Y. Son, K.-Y. Yoo, and Y.-S. Heo. 2017. Molecular Basis for the
  Neutralization of Tumor Necrosis Factor α by Certolizumab Pegol in the Treatment of
  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.
- Guo, and Z. Lou. 2013. Comparison of the Inhibition Mechanisms of Adalimumab and
  Infliximab in Treating Tumor Necrosis Factor α-Associated Diseases from a Molecular View. *J. 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.
- 30. Danese, S., G. Fiorino, L. Peyrin-Biroulet, E. Lucenteforte, G. Virgili, L. Moja, and S.
  Bonovas. 2014. Biological agents for moderately to severely active ulcerative colitis: a
  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  $\alpha$ . *mAbs* 2: 428–439.
- 32. Mulleman, D., M. Ohresser, and H. Watier. 2013. *Introduction to Biological and Small Molecule Drug Research and Development: Chapter 14. The case of anti-TNF agents,*. Elsevier
  Inc. Chapters.
- 632
- 633

#### 634 Figure legends

#### 635 Figure 1: Principle and performance of the method

636 A: From the 3D structures of the antibody and the target, HEX generates docking poses and 637 ranks them according to energetic criteria (H-rank). Each of the HEX top-500 docking poses is 638 evaluated using both non-specific and specific scoring functions. A consensus score is 639 computed, and the poses are ranked according to this score. In parallel, for each docking 640 pose, the A-score, C-score and P-score (see hereafter) are computed, and the poses are 641 ranked. The final ranking of poses is a consensus of the 5 different rankings (HEX, consensus, 642 A-ranking, C-ranking, P-ranking). The top 30 solutions are used to compute, for each residue 643 of the target, the frequency at which it appears within the epitope in these top 30 poses, 644 and to design the epitope peptides. B: Ratio of residues of the epitope within the designed 645 peptides identified using MAbTope on the complete test set (blue), cluspro (orange), frodock 646 (green), epipred (violet) and PPiPP (purple). The values obtained when docking the false 647 positives is shown in red. C: Ratio of epitope residues with the designed peptides identified 648 using MAbTope on the complete test set (blue), on unique targets (green), on small targets 649 (orange) and large targets (purple). The values obtained when docking the false positives on 650 unique targets is shown in red.

651

#### **652** 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 657 coloured region represents one designed peptide; a black star indicates that the peptide 658 belongs to the epitope. Red stars outside of these coloured regions indicate residues of the 659 epitope that do not belong to a designed peptide. B: 3D structures of the complexes 660 between the 25 antibodies (cartoon) and gp120 (surface). All the structures have been 661 superimposed on 4ZMJ, the colour code is given in A.

662

Figure 3: Residues present at the interface of docking poses of the Golimumab-TNF andCertolizumab-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- $\alpha$  with golimumab (B) and certolizumab (C). The three TNF- $\alpha$  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).

674

#### 675 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 /

streptavidin-d2. Increasing doses of Certolizumab were added, from 10<sup>-15</sup> to 10<sup>-5</sup> M (empty
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-TNF $\alpha$ . The 691 complex formed by golimumab and TNF- $\alpha$  was detected by the HTRF mix anti-IgG-Tb / streptavidin-d2. As a displacement control, Certolizumab was incubated with Golimumab / 692 693 TNFα complex. HTRF signals were subtracted of the signal obtained with the control 694 antibody (eculizumab).

695

Figure 5: Predicted mutations abolish the binding of Golimumab on complete TNFα andpeptides.

A. HEK293 cells were transfected with either a mock vector, the wild-type TNFα (Val 77 - Leu 233) or with 3 mutated TNFα constructions. The mutations were selected among the amino acids whose side-chain is solvent-exposed within peptides P1-1, P3-1 and P4-1. Cells were fixed, permeabilized and incubated with golimumab (0.33  $\mu$ M). They were then stained by detection of the Flag epitope fused to the different TNF-α constructs (PE-conjugated antiFlag antibody) (upper panels), or of golimumab using an anti-IgG coupled to APC as the secondary antibody (lower panels). The plots show the side-scatter *vs* light intensity for both 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 mutants series and in blue the ones from the P3-3 mutants series. 713

714



- 716 Figure 1



| 719 <sup>Fi</sup> | igure 2 |
|-------------------|---------|
|-------------------|---------|



- 722 Figure 3









#### 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 bestranked 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



#### Figure S3

A: Epitope in cristal structure Five first peptides predicted by: B: MAbTope C: FRODOCK D: Cluspro





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



Figure S5: Golimumab and Certolizumab binding to TNF $\alpha$  are observed using HTRF. Golimumab (A) and Certolizumab (B) were incubated at indicated concentrations with 8 ng of biotinylated AviTag-TNF $\alpha$  and the HTRF sensors. The TNF $\alpha$  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).



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.

747



Figure S7: conformation of the complexes between TNF- $\alpha$  and four different antibodies.

|                   | đ        | 8        | æ        | ¥        | £        | 8        | 64       | 8        | 8        | P10      | P11      | P12      |
|-------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| MAbTope           | 0,33     | 0,59     | 0,72     | 0,80     | 0,83     | 0,85     | 0,85     | 0,87     | 0,88     | 0,88     | 0,89     | 0,89     |
| complete testset  | (± 0.21) | (± 0.24) | (± 0.21) | (± 0.18) | (± 0.17) | (± 0.16) | (± 0.15) | (± 0.14) | (± 0.13) | (± 0.13) | (± 0.12) | (0.12)   |
| MAbTope           | 0,10     | 0,19     | 0,28     | 0,36     | 0,42     | 0,48     | 0,50     | 0,48     | 0,49     | 0,50     | 0,50     | 0,48     |
| negative controls | (± 0.09) | (± 0.14) | (± 0.17) | (± 0.20) | (± 0.22) | (± 0.24) | (± 0.26) | (± 0.24) | (± 0.26) | (± 0.27) | (± 0.27) | (± 0.26) |
| MAbTope           | 0,34     | 0,63     | 0,75     | 0,81     | 0,85     | 0,85     | 0,85     | 0,86     | 0,85     | 0,86     | 0,89     | 0,94     |
| unique targets    | (± 0.20) | (± 0.22) | (± 0.20) | (± 0.19) | (± 0.18) | (± 0.17) | (± 0.15) | (± 0.15) | (± 0.15) | (± 0.07) | (± 0.1)  | (± 0.09) |
| MAbTope           | 0,39     | 0,68     | 0,82     | 0,88     | 0,90     | 0,91     | 0,91     | 0,94     | 0,96     | 0,96     | 0,97     | 1,00     |
| small targets     | (± 0.19) | (± 0.20) | (± 0.14) | (± 0.11) | (± 0.11) | (± 0.08) | (± 0.08) | (± 0.06) | (± 0.06) | (± 0.05) | (± 0.04) | (NA)     |
| MAbTope           | 0,24     | 0,46     | 0,59     | 0,70     | 0,75     | 0,79     | 0,82     | 0,84     | 0,86     | 0,87     | 0,89     | 0,89     |
| large targets     | (± 0.21) | (± 0.24) | (± 0.21) | (± 0.21) | (± 0.19) | (± 0.18) | (± 0.17) | (± 0.15) | (± 0.14) | (± 0.14) | (± 0.12) | (± 0.11) |
| Cluspro           | 0,12     | 0,23     | 0,31     | 0,36     | 0,43     | 0,43     | 0,43     | 0,47     | 0,50     | 0,52     | 0,54     | 0,54     |
| complete test set | (± 0.20) | (± 0.30) | (± 0.33) | (± 0.34) | (± 0.35) | (± 0.35) | (± 0.34) | (± 0.33) | (± 0.33) | (± 0.32) | (± 0.34) | (± 0.33) |
| Frodock           | 0,12     | 0,21     | 0,28     | 0,35     | 0,39     | 0,44     | 0,43     | 0,45     | 0,48     | 0,47     | 0,49     | 0,51     |
| complete test set | (± 0.20) | (± 0.27) | (± 0.32) | (± 0.34) | (± 0.36) | (± 0.37) | (± 0.37) | (± 0.36) | (± 0.36) | (± 0.37) | (± 0.36) | (± 0.36) |
| EpiPred           | 0,09     | 0,17     | 0,24     | 0,29     | 0,35     | 0,39     | 0,41     | 0,44     | 0,49     | 0,49     | 0,49     | 0,47     |
| complete test set | (± 0.17) | (± 0.25) | (± 0.29) | (± 0.32) | (± 0.35) | (± 0.36) | (± 0.36) | (± 0.37) | (± 0.39) | (± 0.39) | (± 0.39) | (± 0.39) |
| PPiPP             | 0,14     | 0,26     | 0,33     | 0,40     | 0,46     | 0,50     | 0,57     | 0,64     | 0,72     | 0,77     | 0,79     | 0,78     |
| small targets     | (± 0.20) | (± 0.25) | (± 0.28) | (± 0.28) | (± 0.29) | (± 0.31) | (± 0.29) | (± 0.29) | (± 0.28) | (± 0.28) | (± 0.29) | (± 0.30) |
|                   |          |          |          |          |          |          |          |          |          |          |          |          |

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

\_