



HAL
open science

A high-throughput selection system to quantify the specificity of antibodies

Mégane Boulas

► **To cite this version:**

Mégane Boulas. A high-throughput selection system to quantify the specificity of antibodies. Biotechnology. Université Paris Cité, 2021. English. NNT : 2021UNIP5155 . tel-04556532

HAL Id: tel-04556532

<https://theses.hal.science/tel-04556532>

Submitted on 23 Apr 2024

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.

Université de Paris

**Frontières de l'Innovation en Recherche et Education
FIRE (ED 474)**

CIRB, UMR 7241 CNRS, U 1050 INSERM, Collège de France

A high-throughput selection system to quantify the specificity of antibodies

Par Mégane Boulas (Matysiak)

Thèse de doctorat de Sciences du vivant appliquées, biotechnologie et
ingénierie des biosystèmes moléculaires

Dirigée par Clément Nizak
Et par Olivier Rivoire

Présentée et soutenue publiquement le 14/12/2021

Devant un jury composé de :

Agathe Urvoas, Professeur, Université Paris Saclay, rapporteur

Philipp Bucher, Group leader, EPFL, rapporteur

Hervé Nozac'h, Chargé de recherche, Université Paris Saclay, examinateur

Sébastien Lacroix-Desmazes, Directeur de recherche, Université de Paris, examinateur

Clément Nizak, Directeur de recherche, Sorbonne université, co-directeur

Olivier Rivoire, Chargé de recherche, Collège de France, co-directeur

LISTE DES ÉLÉMENTS SOUS DROITS

Liste de **tous les éléments retirés** de la version complète de la thèse
faute d'en détenir les droits

Légende de l'image	N° de l'image	Page(s) dans la thèse
A schematic representation of a binding affinity landscape	1.1	14
Sequence specificity landscapes of a DNA binding protein	1.2	16

2 Abstract

3 **A high-throughput selection system to quantify the specificity of antibodies**

4
5 Proteins are phenotypically characterized by their functional properties, e.g. bind-
6 ing affinity and specificity for a target molecule in the case of antibodies. Understanding
7 whether and how much these properties are linked has remained largely elusive so far. Ad-
8 dressing this question relies on our capacity to quantify binding specificity, which stands
9 as a challenge. We developed an *in vitro* experimental setup combined with a statistical
10 analysis pipeline to systematically quantify the binding specificity of proteins. In par-
11 ticular, we designed a library of random DNA molecules acting as targets and screened
12 them using SELEX against a pool of recombinant antibodies with frameworks taken from
13 natural antibodies. We found that we could assess global specificity (by estimating the
14 average affinity of antibodies for the DNA library) as well as local specificity (by measur-
15 ing the diversity of DNA sequences bound by the antibodies). We showed that selection
16 experiments can be optimized in order to provide a robust quantification of affinity and
17 specificity in a high throughput fashion. These results pave the way for a systematic map-
18 ping of the relationship between affinity and specificity.

19

20 **Keywords**

21 antibody, specificity, SELEX, high-throughput sequencing

22

Résumé

25 Un système de sélection à haut débit pour quantifier la spécificité des anticorps

26

27 Les protéines sont des biomolécules fondamentales de tout être vivant. Leurs fonctions très
28 variées sont contrôlées par leur structure tridimensionnelle ; elle-même déterminée par leur séquence
29 d'acides aminés. Le rôle des anticorps, par exemple, est de reconnaître et neutraliser des molécules
30 cibles (appelées antigènes) qui pourraient représenter une menace pour l'organisme. Leur fonction
31 est directement liée à leur affinité (force d'interaction) et spécificité (précision) de liaison à ces
32 molécules cibles. Bien que de nombreux travaux aient été réalisés dans le but d'étudier l'affinité
33 des anticorps d'une part, et leur spécificité d'autre part, nous avons à ce jour une compréhension
34 limitée de la relation entre ces deux propriétés fonctionnelles. Par ailleurs, tandis que la métrique
35 d'affinité est clairement définie (elle correspond à la constante de dissociation entre la protéine et
36 sa cible), il n'existe pas de métrique consensuelle de spécificité. Nous faisons donc face à deux
37 principaux défis : Comment quantifier la spécificité de liaison de protéines ? Et quelle est la
38 nature de la relation liant l'affinité et la spécificité de liaison des protéines ? De manière à y
39 répondre, il s'agit tout d'abord de mettre au point une méthode visant à mesurer la spécificité. Ici,
40 nous proposons un procédé expérimental *in vitro* ainsi qu'une approche statistique permettant une
41 quantification systématique de la spécificité de liaison des protéines. Nous avons mis au point une
42 banque de molécules d'ADN (utilisées comme ligands) avec une structure en tige-boucle que nous
43 avons sélectionnées par SELEX contre un ensemble d'anticorps recombinants (dont la structure est
44 issue d'anticorps naturels possédant différents niveaux de maturation -ou différentes affinités- pour
45 une cible biologique). Nous avons ensuite utilisé le séquençage à haut débit afin d'identifier les
46 molécules d'ADN ayant les plus hautes affinités pour ces anticorps. Nous proposons une métrique
47 de la spécificité globale (en estimant l'affinité moyenne des anticorps pour la banque d'ADN) ainsi
48 que de la spécificité locale (en mesurant la diversité des séquences d'ADN sélectionnées par les
49 anticorps). Nous avons par ailleurs optimisé le protocole de sélection dans le but d'obtenir une
50 quantification robuste et à haut débit de l'affinité et de la spécificité. L'utilisation d'une banque
51 de molécules d'ADN nous a permis de mesurer différentes spécificités locales et globales pour des
52 anticorps caractérisés par différents niveaux de maturation. En particulier, nous avons constaté
53 que les anticorps moins maturés ont une affinité moyenne pour la banque de molécules d'ADN
54 plus élevée, et qu'ils sélectionnent avec moins de précision les séquences d'ADN par rapport aux
55 anticorps maturés. Ces résultats ouvrent la voie vers une étude systématique de la relation entre
56 affinité et spécificité.

57

58 Mots-clés:

59 anticorps, spécificité, SELEX, séquençage à haut-débit

61 Acknowledgements

62 I would like to thank the following people who have helped me undertake this project:
63 my thesis directors Clément Nizak and Olivier Rivoire, for their supervision, teaching and
64 trust; my husband Ihab, for his support, encouragement and our valuable discussions; as
65 well as all the members of my lab.

66 Contents

67	Abstract	3
68	Résumé	5
69	Acknowledgements	7
70	1 Introduction	13
71	1.1 Proteins are characterized by a set of functional properties	13
72	1.2 Measuring specificity of DNA-binding proteins	14
73	1.3 Antibodies as a model system	17
74	1.4 Our approach	19
75	2 Measuring the average binding affinity of antibodies for DNA ligands	23
76	2.1 Context & Problem	23
77	2.2 How we estimated average binding affinities of antibodies for the DNA library	23
78	2.2.1 Presentation of our minimalist system	23
79	2.2.2 Presentation of the selection experiments	24
80	2.3 Results	25
81	2.3.1 Comparison of binding affinities between 2 groups of antibodies . . .	25
82	2.3.2 Are the differences that we have seen systematic within each group	
83	of antibodies?	26
84	2.4 Summary	28
85	2.5 Discussion	29
86	3 Measuring the sequence diversity of DNA ligands selected by antibodies	31
87	3.1 Context & Problem	31
88	3.2 How we analyzed the sequences of the selected DNA ligands	32
89	3.2.1 Presentation of our minimalist system	32
90	3.2.2 Analysis of the deep-sequencing data in 3 steps	32
91	3.3 Results	33
92	3.3.1 Can antibodies discriminate DNA sequences ?	33
93	3.3.2 Within each group, do different antibodies select the same DNA	
94	sequences ?	34

95	3.3.3	To what extent does the constant region of DNA molecules impact the selection of the loop region ?	36
96			
97	3.3.4	Does the antibody framework impact the sequence diversity of the selected DNA molecules ?	38
98			
99	3.3.4.1	Resolution = DNA loop	38
100	3.3.4.2	Resolution = DNA motifs inside the loop	39
101	3.4	Summary	45
102	3.5	Discussion	46
103	4	High-throughput experimental measurements of binding affinities	49
104	4.1	Introduction	49
105	4.1.1	Context & Problem	49
106	4.1.2	State of the art	50
107	4.1.3	Our approach	51
108	4.2	Presentation of the calibration experiment	52
109	4.2.1	Our ligand-target system	52
110	4.2.2	The calibration experiment is a high-throughput technique to measure the K_d of a protein for 218 ligands in parallel	52
111			
112	4.2.3	ELISA is used as a low-throughput technique to measure the K_d of a protein for one ligand at a time	54
113			
114	4.3	Results	54
115	4.3.1	Our calibration experiment allows us to generate 218 binding curves at the same time	54
116			
117	4.3.1.1	This experimental setup provides a high-throughput access to enrichment values	54
118			
119	4.3.1.2	The comparison of the high-throughput calibration and ELISA data shows a linear relationship	56
120			
121	4.3.1.3	A combined use of SELEX and ELISA can provide access to affinity distributions under linear conditions	57
122			
123	4.3.2	A simple binding model based on the specific binding of the DNA ligands to the antibody does not describe the experimental data	58
124			
125	4.3.2.1	Why do we need to identify the factors that link the experimental data to the KDs we want to infer?	58
126			
127	4.3.2.2	Presentation of the selection model & identification of the factors that link our data to the affinities	59
128			
129	4.3.3	There is a tradeoff between the diversity of the library and the resolution of the experiment	63
130			
131	4.3.3.1	Experimental data suggest that the design of the DNA ligand and library impacts the resolution of the selection experiment	63
132			
133	4.3.3.2	Simulated data suggest a tradeoff between the diversity of the ligand library and the resolution of the experiment	65
134			

	<i>CONTENTS</i>	11
135	4.4 Summary	67
136	4.5 Discussion	67
137	5 Conclusion, discussion and perspectives	71
138	5.0.1 Conclusion	71
139	5.0.2 Discussion and perspectives	72
140	5.0.2.1 Limits of our approach	72
141	5.0.2.2 Next experiments	72
142	5.0.2.3 Comparison of our results with previous works	73
143	6 Supplementary material	77
144	6.1 Design	77
145	6.1.1 Presentation of the synthetic antibodies	77
146	6.1.2 Presentation of the DNA library	77
147	6.1.3 Calibration experiment	78
148	6.2 Protocols	80
149	6.2.1 Production of antibodies	80
150	6.2.2 ELISA assay	81
151	6.2.3 SELEX assay	82
152	6.2.3.1 On plate	82
153	6.2.3.2 On beads	83
154	6.2.4 Deep-sequencing of the DNA library	83
155	6.3 Analysis	84
156	6.3.1 Binding score based on SELEX data	84
157	6.3.2 Binding score based on ELISA data	85
158	6.3.3 The basic analysis of the deep-sequencing data informs us on 3 ex-	
159	perimental biases	85
160	6.3.3.1 The bias in the initial library	85
161	6.3.3.2 The non-specific binding of the DNA library to the plate	85
162	6.3.3.3 The sampling bias	86
163	6.3.4 Sequence logo of the most selected loops	86
164	6.3.5 Analysis of the motifs inside the loop	87
165	6.3.6 Modeling the sampling bias	87
166	6.3.7 Impact of the calibration protocol on the shape of the binding curves	88
167	6.4 Experiments	88
168	6.4.1 Reproducibility of the SELEX experiments	88
169	6.4.2 Comparison of selection with control experiment	89
170	6.4.3 Impact of the constant region on the selection of the DNA loops by	
171	Germline antibodies	90
172	6.5 Supplementary figures	91
173	Bibliography	99

174 Chapter 1

175 Introduction

176 1.1 Proteins are characterized by a set of functional prop- 177 erties

178 Proteins are part of the fundamental biomolecules of living organisms. They perform
179 a vast array of functions including the catalysis of biological reactions, the transport of
180 small molecules and the formation of the cell structure. Their function is related to the
181 3D structure that they adopt and which is determined by their amino-acids sequence.
182 Many studies have been dedicated to the study of the relationship between the structure
183 of proteins and their function, with the ultimate goal of predicting functional properties
184 (e.g. stability, catalytic activity, affinity, etc.) from amino-acid sequences. However,
185 and while proteins are characterized by an ensemble of properties that can potentially be
186 linked to each other, most of the studies only focus on the investigation of one particular
187 functional property .

188 For example, the binding of proteins to ligands which initiates every biological reac-
189 tions depends on 2 major functional properties: (1) the affinity (i.e. the binding strength
190 between a protein and its ligand) and (2) the specificity (i.e. the ability of a protein to
191 discriminate between different ligands). These functional properties are related by defini-
192 tion, since the specificity of a protein is defined as the distribution of its affinity over every
193 possible ligand. However, we don't know if these two properties are intrinsically linked
194 and if they are, what is the nature of their relationship.

195
196 One possible schematic view of specificity is the binding landscape (Fig. 1.1) [[George
197 and Gray, 1999](#)] where every possible ligand is mapped in a 1 or 2 dimensional space (to
198 simplify the representation). In such a space, the distance between ligands corresponds
199 to their relative differences ¹. Each ligand has a particular binding affinity (or binding
200 energy) for the protein and the ensemble of these binding affinity values characterizes the

¹The way to define how ligands differ from one another can be argued; if ligands are DNA molecules, an example for a metric would be the Hamming distance (i.e. the number of positions at which the nucleotides are different between 2 DNA sequences).

201 topography of the landscape.

202

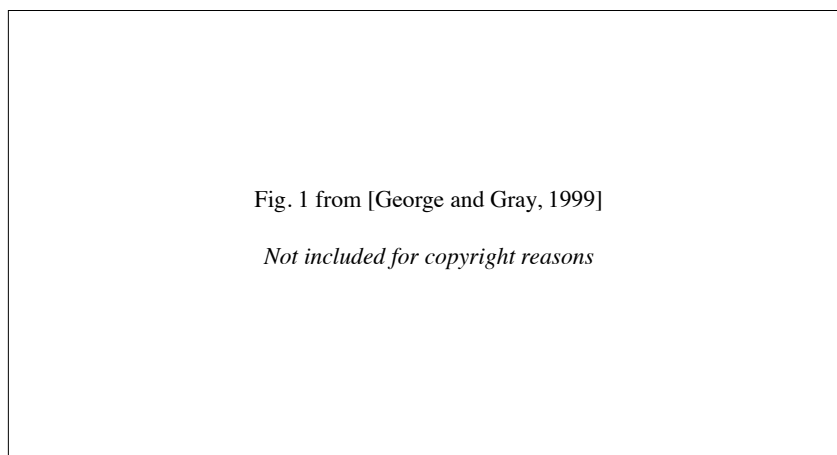


Figure 1.1: **A schematic representation of a binding affinity landscape.** All possible antigen-binding sites are shown on the x axis, with the most similar adjacent to each other. The affinity of the antibody for each antigen is shown on the y axis. Coloured arrows show different trajectories of mutating proteins (antibodies) that increase their binding affinity for antigen-binding sites. Taken from [George and Gray, 1999].

203 We can use the binding landscape picture to describe two levels of specificity:

- 204 • At the larger scale, we can verify if the protein of interest has a high affinity for
205 very different ligands by looking at the global ruggedness of its binding landscape.
206 The ability of a protein to interact with very different ligands is characteristic of its
207 global specificity
- 208 • At the smaller scale, we can verify if the protein has different affinities for ligands
209 that are very similar in structure by looking at the local ruggedness of its binding
210 landscape. The ability of a protein to discriminate between similar ligands is char-
211 acteristic of its local specificity.

212

213 Measuring protein's specificity is the first step towards understanding its function.
214 However, it is difficult (if not effectively impossible) to determine the full binding landscape
215 of a protein, as it would require the experimentalist to measure its affinity for a myriad of
216 possible ligands. Still, it is possible to explore subregions of this landscape: this is what
217 research efforts over the past decades have thrived to measure [Jolma et al., 2013] [Prigent
218 et al., 2018] [McGeary et al., 2019] [Tonikian et al., 2008].

219 1.2 Measuring specificity of DNA-binding proteins

220 As a case study, let us consider the example of a category of proteins for which binding
221 affinity and specificity are crucial in their biological function: transcription factors (TFs).

222 Their main function is the control of gene expression achieved by physically interacting
223 with genomic DNA. Human transcription factors, for instance, recognize DNA motifs (i.e.
224 a precise sequence of 4 to 20 nucleotides) among the billions of more or less accessible
225 nucleotides that constitute the genome. The affinity and specificity of these transcription
226 factors for DNA motifs crucially affect the way in which they bind to their target sequence.

227

228 The first studies of transcription factors were based on the identification of the major
229 DNA motif that they recognize. However, it has been shown that relevant binding sites are
230 not only the ones with the highest affinity [Jiang and Levine, 1993] [Crocker et al., 2015].
231 Consequently, following studies focused on the mapping of the full binding landscape of
232 these transcription factors in the space of all possible DNA motifs.

233

234 A variety of high-throughput techniques have been developed to study the binding
235 landscape of transcription factors in the space of DNA sequences. Among those tech-
236 niques, 2 are particularly popular: (1) the HT-SELEX (high-throughput Systematic evo-
237 lution of ligands by exponential enrichment) approach, during which up to billions of short
238 DNA fragments are mixed together and screened against a transcription factor for their
239 affinity for the latter [Gu et al., 2013] [Chen et al., 2016] [Jolma et al., 2010] [Jolma et al.,
240 2013] and (2) the PBM (Protein Binding Microarray) based approaches during which up
241 to millions of short DNA fragments are immobilized on a chip and the proteins are allowed
242 to interact with these DNA motifs according to their respective affinities [Carlson et al.,
243 2014] [Le et al., 2018] [Isakova et al., 2017].

244 In both approaches, the experimentalists design a library of short DNA ligands containing
245 every possible DNA motifs and they expose the DNA library to the transcription factor
246 of interest. The output of these assays are binding scores that are indirectly linked to
247 the affinities: the output of HT-SELEX assays are enrichment values that indicate how
248 much the DNA ligands increased their prevalence during selection against a TF; while the
249 output of PBM-based approaches are indirect measurements of the number of TFs that
250 interacted with each DNA ligand.

251

252 The resulting binding scores are either directly used to assess the specificity profile of
253 TFs, or are further analyzed with theoretical models in order to be converted into binding
254 energies or dissociation constants (definition of dissociation constant in 4.1.1).

255

256 Using the binding data, the specificity profile of TFs are generally depicted in 2 ways:

- 257 • The binding scores values are represented as a distribution (Fig.1.2, Left) or as a
258 binding landscape (Fig.1.2, Upper Right) where the ensemble of DNA sequences is
259 mapped in a 1 or 2 dimensional space (their distance depending on their Hamming
260 distance²) and their corresponding binding scores characterize the topography of the

²The Hamming distance is the number of positions at which the nucleotides are different between 2 sequences.

261 landscape. The distribution of binding scores informs us about the number of DNA
 262 ligands that have a high affinity for a given TF while the landscape informs us about
 263 both the number of high affinity DNA ligands and the similarity of their sequences.

- 264 • The sequences with the highest binding scores are directly compared as a sequence
 265 logo (Fig.1.2, Left) where the 4 nucleotides are represented as stacked letters for each
 266 position and their relative sizes are proportional to their frequency. More sophisti-
 267 cated analyses have been proposed to generate the sequence logo that represents the
 268 most selected DNA sub-motifs (within the selected sequences) [Jolma et al., 2010]
 269 [Rube et al., 2018][Isakova et al., 2017]. Online tools (<https://meme-suite.org>) are
 270 also available to identify or compare motifs from high-throughput selection data.
 271 The sequence logo is a helpful representation to quickly identify the DNA motifs
 272 with the highest affinities for transcription factors.

273

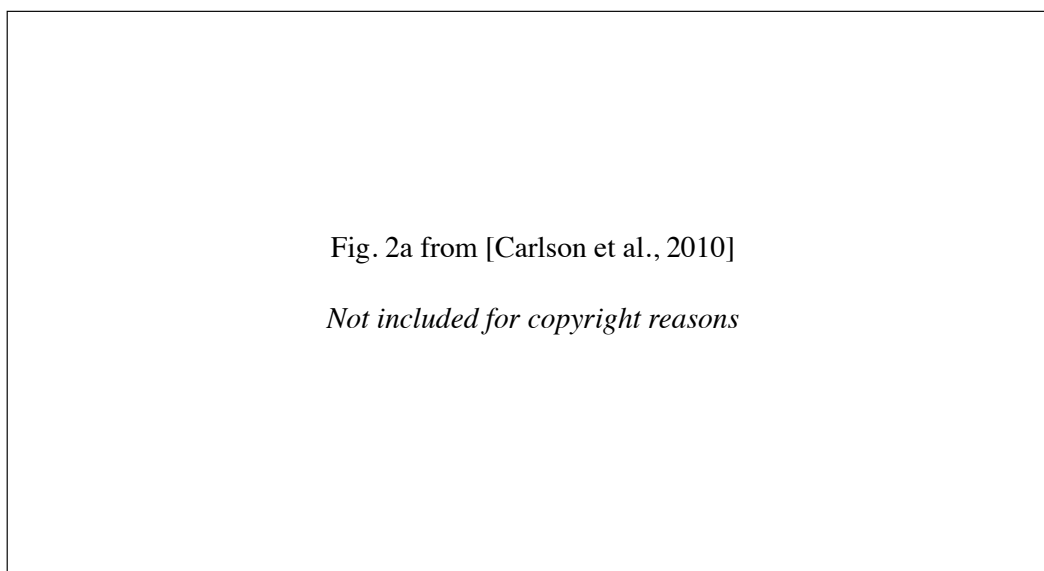


Figure 1.2: **Sequence specificity landscapes of a DNA binding protein.** Left: distribution of the binding score obtained with CSI (Cognate Site Identifier) approach. The sequences with the highest binding scores are compared in a sequence logo. Upper Right: Circular sequence specificity landscape displays the binding intensities of a given DNA ligand across all DNA permutations on the CSI array, with every sequence displayed on the plot. We note the number of mismatches from the most selected motif (seed motif) on each concentric circle (SSL: Sequence Specificity Landscape, PWM: Position Weight Matrix). Taken from [Carlson et al., 2010].

274 However, this representation has 2 major limitations: (1) it only shows the most se-
 275 lected DNA sequences or motifs (and as we mentioned above, DNA motifs with small
 276 affinities can be critical in the function of TFs) and (2) it assumes that each position in
 277 the sequence contributes additively to the overall affinity between the DNA motif and the
 278 TF (it provides no information about the correlation between positions). As a result, more

sophisticated models have been proposed to include partial or total dependency between the nucleotides along the selected DNA sequences [Rastogi et al., 2018] [Robinson et al., 2003].

All in all, these studies have shown how the high-throughput measurement of binding scores can be used to get a sense for the specificity landscape of TFs. However, on one hand, binding scores cannot necessarily be assumed to be proportional to affinity values (and consequently the specificity profiles are merely qualitative in nature). On the other hand, there has been efforts to convert binding scores into dissociation constants to obtain a quantitative specificity landscape, but only at the cost of heavy experimental setups and sophisticated theoretical models [Isakova et al., 2017]. We propose in chapter 4, a new high-throughput approach to simply and quickly measure multiple dissociation constants in a single experiment in order to obtain quantitative specificity profile.

In this section, we presented the investigations on TF specificity as a case study for the approaches used to measure specificity profiles. We next present the advantages associated to using a different model system – another protein whose function is critically linked to affinity and specificity: the antibody.

1.3 Antibodies as a model system

Antibodies are major players in immune response. Their function lies in identifying and neutralizing pathogens that can infect our organism by binding target molecules called antigens. Antibodies are known for their high binding specificity for target molecules. As an example, they must be able to differentiate antigens (that belong to pathogens) from self-antigens (that belong to the host organism). They are also commonly used as tools to outline cellular structures by specifically recognizing a protein of interest. However, antibodies are also known for their ability to interact with very different antigens (i.e. their polyspecificity) (the case of broadly neutralizing antibodies is presented below). As an example, some theoretical studies suggest that antibodies must be able to interact with several target molecules in order to allow our immune system to recognize the infinite number of possible antigens we can encounter. To illustrate this idea, a conceptual formalism called “shape space”³ has been proposed 40 years ago to represent the antibody antigen binding and to explain how antibodies can recognize antigens they never encountered before [Perelson and Oster, 1979][Lapedes and Farber, 2001].

³In a shape space, antibodies and antigens are represented as points. Their coordinates in this space can be physical or chemical properties and the euclidean distance between antigens and antibodies is linked to their affinities. Each antibody can be surrounded by a sphere that represents its specificity: any antigen located within the sphere can be recognized by the antibody. Shape space representations have been used to understand how a finite number of antibodies could recognize an infinite number of antigens. Namely, hypotheses have been formulated regarding the optimal size of the sphere (i.e. specificity) N antibodies should have so their spheres fill the space S and ensure that all the antigens are recognized by the antibodies [Perelson and Oster, 1979].

312

313 Once an antibody encounters an antigen for the first time, it undergoes an acceler-
314 ated evolutionary process called affinity maturation that consists in cycles of mutations,
315 selection and amplification. Over months or years, antibodies accumulate mutations in
316 their sequence to increase their affinity for this antigen. During this maturation process,
317 it has been reported that specificity of antibodies evolves as well. However, it is not clear
318 whether and how the affinity and specificity coevolve over the course of antibody maturation.
319 Different scenarios have been proposed [Eisen and Chakraborty, 2010].

320

321 In the first scenario, antibodies are initially able to recognize multiple antigens but lose
322 their polyspecificity over the course of affinity maturation by decreasing their flexibility.
323 It has been showed with crystallographic data that non-maturated antibodies can have
324 multiple 3D conformations (allowing the binding to multiple ligands) while maturated
325 antibodies can only have one conformation (allowing the binding to one single ligand)
326 [James, 2003].

327

328 In the second scenario, antibodies develop the ability to recognize very different anti-
329 gens over the course of affinity maturation. Typical examples are Broadly Neutralizing
330 antibodies that have maturated against the HIV virus for years. Numerous studies have
331 shown that broadly neutralizing antibodies developed the ability to recognize mutating
332 antigen but also self-antigens [Haynes et al., 2013] [Prigent et al., 2018]. Similar observa-
333 tions have been made for artificially maturated antibodies, in the context of ACT therapy⁴.
334 It has been shown that antibodies with very high affinity for a target molecule developed
335 auto-reactivity (i.e. they recognized self-antigens) [D’Ippolito et al., 2019]. The authors
336 showed that a fine balance between affinity and specificity was necessary to develop arti-
337 ficially maturated antibodies that maximize the elimination of the antigen and minimize
338 the interaction with self-antigens. A theoretical study proposed a simplistic model to un-
339 derstand the relationship between affinity and specificity of proteins [Savir and Tlusty,
340 2007]. Using a minimalist model of ligand and target, the authors showed that a flexible
341 ligand must not perfectly match its target (i.e. must not have a maximal affinity for its
342 target) but should have a slight mismatch to maximize its binding specificity in the pres-
343 ence of similar competing targets. Indeed, if 2 competing targets are very similar, and if
344 the ligand has a high affinity for the first target, it will have a medium affinity for the
345 second target. However, if the ligand has a medium affinity for the first target, it will have
346 a negligible affinity for the second one, and thus its specificity will be increased. These
347 results suggest that a maximal specificity is possible at the cost of a non-optimal affinity.

348

349 A third scenario can also be considered where there is no fundamental link between

⁴During ACT therapy, T cells of a patient are collected, maturated *in vitro* to increase their affinity for an antigen, amplified and re-injected into the patient to increase the efficiency of his immune system against the pathogen

350 the affinity and specificity of antibodies.

351

352 The 2 first scenarios suggest the existence of different selective pressures: (1) In the first
353 one the antibodies evolve to bind a single antigen while in the second one (2) antibodies
354 evolve to bind similar antigens. However, we still don't know if the selective pressure
355 alone can account for the non-trivial relationship that is observed between the affinity and
356 specificity of antibodies. In order to better understand what links these two properties,
357 we first need to quantify the specificity of antibodies. The main question we address in
358 this project is the following: can we quantitatively measure the specificity of antibodies
359 (with potentially different specificity profiles) ?

360 We propose to study the specificity of a pool of recombinant antibodies (called synthetic
361 antibodies in this manuscript) using a DNA ligand library and a high-throughput *in vitro*
362 selection assay (we expect the studied antibodies to have different specificity profiles and
363 thus we expect to observe different binding behaviors with the DNA ligand library we
364 designed).

365 1.4 Our approach

366 We studied a pool of synthetic antibodies (i.e. recombinant antibodies) that have been
367 designed and studied as part of previous work [Boyer et al., 2016] [Schulz et al., 2021].
368 The synthetic antibodies are composed of 2 main parts (Fig. 1.3): (1) A scaffold (also
369 called framework) corresponding to a 100 amino acids sequence directly taken from the
370 VH region of a human antibody and (2) a CDR3 region corresponding to a 4 amino acids
371 sequence acting as a binding site.

372

373 2 different frameworks have been designed using 2 human antibodies. One has never
374 matured against an antigen, it is called "Germline"(-reversed). The other has matu-
375 rated against the HIV virus and has accumulated 15 mutations in its VH region (it is
376 called "Limited"). From these 2 frameworks, 2 libraries have been designed by random-
377 izing the CDR3 region of the antibody sequences. Each library had a diversity of 10^5
378 different antibodies.

379

380 The authors screened each antibody library separately (by Phage Display) against 2
381 different DNA ligands in order to identify the synthetic antibodies with the highest affinity
382 for each DNA targets.

383

384 Measurement of specificity at a small scale:

385

386 Then the authors investigated the specificity of a subset of synthetic antibodies using
387 2 low-throughput selection experiments. They build a minimalist library of antibodies
388 composed of the best binders for each DNA ligand, with Germline and Limited frameworks

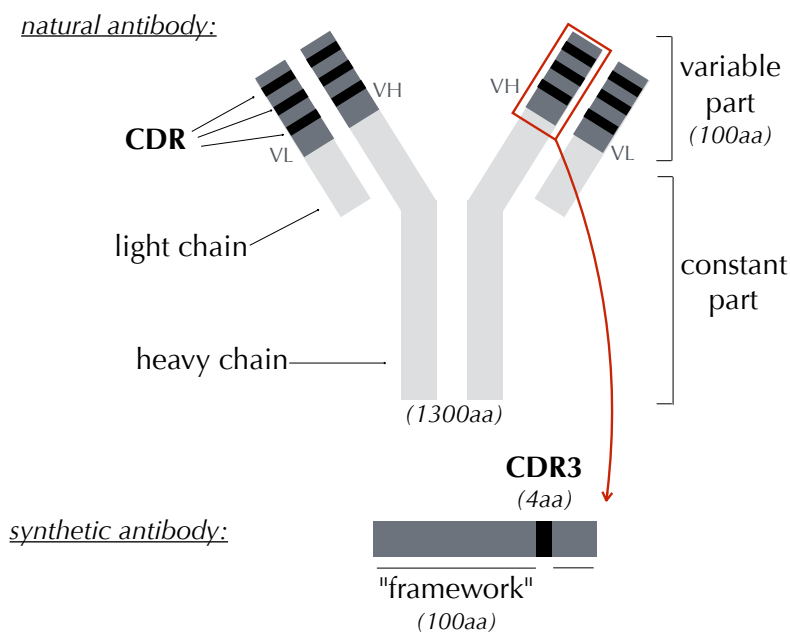


Figure 1.3: **Our protein model.** The proteins are recombinant antibodies with frameworks taken from human antibodies. We call them “synthetic antibodies”. As part of previous work [Boyer et al., 2016], several synthetic antibodies have been designed with different framework and CDR3 regions.

389 then they selected the antibodies against both DNA ligands, separately (Fig. 1.4).

390 They showed that different synthetic antibodies have different specificities for 2 DNA
 391 ligands (some antibodies have a higher affinity for the first DNA ligand than for the other
 392 one and vice versa). These results motivated the use of this system to study specificity on
 393 a larger scale with a high-throughput approach.

394

395 **Our project consists in measuring specificity at a larger scale:**

396

397 The authors previously selected (via Phage Display) 10^5 different antibodies against 2
 398 DNA ligands. Here we select (via SELEX) a library of 10^4 DNA ligands against different
 399 synthetic antibodies (we chose antibodies with blue and green frameworks in Fig. 1.4,
 400 with potentially different specificity profiles) in order to measure their specificity for DNA
 401 ligands. We have 2 main goals:

402 (1) We want to propose a quantitative measurement of specificity (chapter 4)

403 (2) We want to understand the relationship between the sequence of antibodies and their
 404 specificity for DNA ligands. More precisely, we ask if synthetic antibodies with various
 405 sequences have different average affinities for DNA ligands (chapter 2) and if they can
 406 discriminate between DNA ligands with similar sequences (chapter 3).

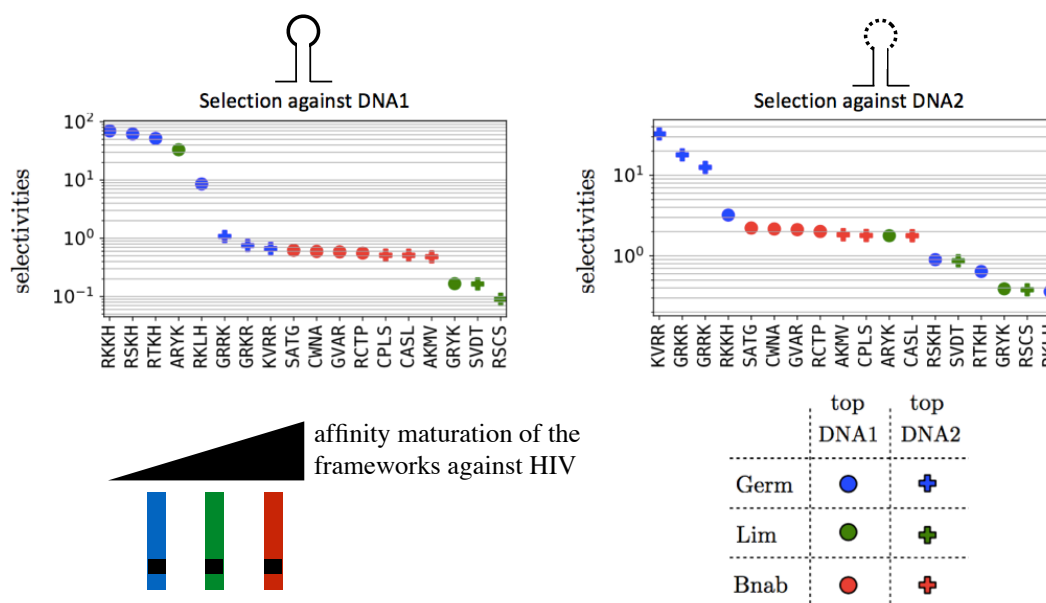


Figure 1.4: **Comparison of selectivities for antibodies with different frameworks and CDR3 regions.** The authors build a minimalist library of antibodies with 3 different frameworks (blue, green and red, more or less matured against a common antigen). The minimalist library has been selected by Phage Display against 2 DNA ligands, separately. For each antibody, the selectivity has been calculated as the ratio between its number of copies after and before the selection (selectivities are positively linked to affinities). The measured selectivities are presented in the upper plots. Taken from [Schulz et al., 2021].

407 Chapter 2

408 Measuring the average binding 409 affinity of antibodies for DNA 410 ligands

411 2.1 Context & Problem

412 Antibodies have the ability to interact with a broad range of ligands such as proteins,
413 peptides, polysaccharides, lipids, DNA and RNA. Specificity profile of antibodies can
414 be schematized as a binding affinity landscape where ligands are mapped in a one or
415 two dimensional space and where the corresponding binding affinities characterize the
416 topography of the landscape.

417 We propose an approach to investigate the global specificity of a set of antibodies by
418 exploring a sub-region of their binding landscape where ligands are nucleic acids. Using
419 selection assays, we estimated the average affinity of antibodies for DNA ligands. We
420 asked whether we could measure different binding affinities to DNA ligands for antibodies
421 with different sequences. We answered the question in 2 steps. First we compared the
422 affinities between 2 groups of antibodies with different frameworks. Then we verified if
423 the differences we observed were systematic within each group.

424 2.2 How we estimated average binding affinities of antibod- 425 ies for the DNA library

426 2.2.1 Presentation of our minimalist system

427 The protein models we worked with are synthetic antibodies. They have been designed
428 and studied as part of previous work [Boyer et al., 2016][Schulz et al., 2021]. They are
429 composed of 2 regions: the framework that forms the scaffold of the antibody and the
430 CDR3 region that is directly involved in the binding of the antibody to its target (Supp.
431 6.1.1). We studied 2 groups of synthetic antibodies: each group has a different framework.

432 The first group is composed of 6 antibodies and is called “Germline”, the framework
433 comes from a human Germline(-reversed) antibody. The second group is composed of 5
434 antibodies and is called “Limited”, the framework comes from a human antibody that
435 has matured against the HIV virus. Within each group, the antibodies share the same
436 framework but have different CDR3 regions. In each group, some CDR3 have been chosen
437 because they have the highest affinity (from their group) for a particular DNA ligand
438 (the corresponding antibodies are called “TOP”) and the other CDR3 have been chosen
439 because they have a lower affinity for the same ligand (the corresponding antibodies are
440 called “Random”).

441

442 We used 2 libraries of 32,768 DNA molecules (each) as ligands to measure the global
443 specificity of the synthetic antibodies (Supp. 6.1.2). The DNA molecules are single-
444 stranded and have a stem-loop 3D structure. They share the same sequence in the stem
445 region but have different sequences in the loop region (their structure is similar to ligands
446 used to select antibodies in [Boyer et al., 2016] and [Schulz et al., 2021] but their loops
447 are randomized).

448 2.2.2 Presentation of the selection experiments

449 We had two approaches to study the binding affinity of synthetic antibodies: We used
450 the SELEX approach to have a qualitative measurement of the average affinity of each
451 antibody for one of the DNA ligand libraries, and we used the ELISA approach to have a
452 quantitative measurement of affinities between each antibody and few DNA ligands.

453

454 1- SELEX approach:

455 We used SELEX [Jolma et al., 2010] to estimate average binding affinity of each synthetic
456 antibody for a population of thousands of DNA molecules. During SELEX experiments,
457 multiple copies of the synthetic antibody are immobilized on a solid surface and the DNA
458 library is exposed to the antibody for 1 hour. After 1 hour, the DNA molecules that didn’t
459 interact with the antibody are washed out and the DNA molecules in complex with the
460 antibody are collected (Supp. 6.2.3, we performed a single round of selection). We then
461 estimated the concentration of recovered DNA molecules using an electrophoresis gel to
462 generate a binding score for each synthetic antibody (Supp. 6.3.1).

463

464 2- ELISA approach:

465 We used ELISA to measure the dissociation constant (i.e. affinity) of synthetic antibodies
466 for few DNA molecules, separately. During ELISA experiment, we fixed the concentration
467 of the DNA molecules, varied the concentration of the antibodies and measured the con-
468 centration of the antibody-DNA complex at equilibrium for each condition (Supp. 6.2.2).
469 As a result we obtained a binding curve from which we could directly measure the disso-
470 ciation constants of the complex and generate a binding score (Supp. 6.3.2).

471

472 For both approaches, high binding scores suggest that antibodies have high affinities
473 for the DNA library and thus can interact with DNA ligands.

474

475 2.3 Results

476 2.3.1 Comparison of binding affinities between 2 groups of antibodies

477 We compared the binding affinities of Germline and Limited antibodies. If the two groups
478 have similar binding affinities, we expect them to have similar binding scores for the DNA
479 library.

480

481 We first used SELEX to calculate the binding scores of 1 TOP antibody in each group.
482 For each antibody, we did 4 selection experiments where we varied the experimental condi-
483 tions (we tested 2 different DNA libraries, and 2 different concentrations for each library).
484 Thus we obtained 4 binding scores for the antibody coming from the Germline group,
485 and 4 binding scores for the antibody coming from the Limited group (Fig. 2.1A). The
486 results show binding scores that are higher for the Germline antibody than for the Limited
487 antibody in average. If we look at each specific condition separately (matching symbols
488 correspond to the same experimental conditions), the Germline antibody has systemati-
489 cally a higher binding score than the Limited antibody.

490

491 Then we used ELISA to calculate the binding scores of 3 TOP Limited and 3 TOP
492 Germline antibodies for 1 to 3 DNA molecules (Fig. 2.1B). We used the same protocol to
493 measure all the affinity values but we couldn't accurately measure the dissociation con-
494 stant of the Limited antibodies (their affinity are too low for these DNA molecules so we
495 didn't manage to obtain their full binding curves). Each unsuccessful measurement of K_d
496 is displayed below the "detection threshold" line. The binding scores are systematically
497 higher for Germline antibodies than for Limited antibodies.

498

499 Conclusion:

- 500 • **SELEX and ELISA experiments show that the TOP Germline antibodies**
501 **have a higher average affinity for the DNA ligand library than the TOP**
502 **Limited antibodies we tested.**

503

504 Discussion:

- 505 • We compared Germline and Limited antibodies that were the best binders from
506 their respective group, for a given DNA ligand. We systematically observed a higher

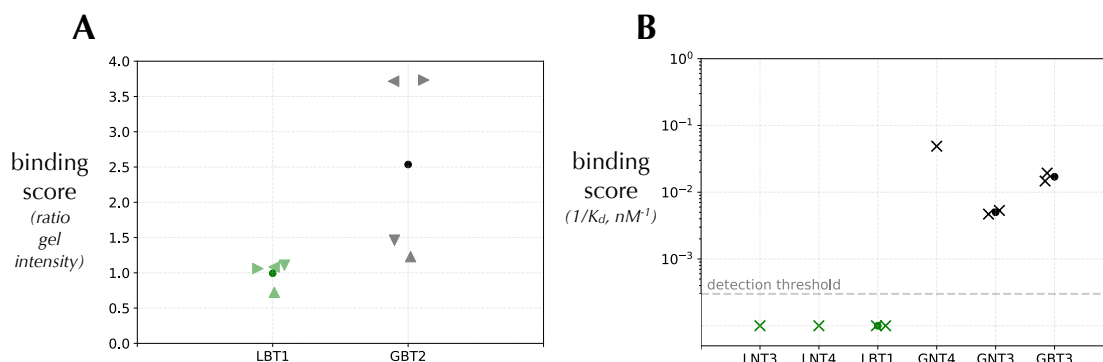


Figure 2.1: **Germline and Limited antibodies have different average affinities for DNA ligands** A. The Germline antibody has a higher average affinity for the DNA ligand library than the Limited antibody, independently of the experimental conditions. The binding scores represent the amount of DNA ligand library collected at the end of the selection experiment (measured with SELEX). The average affinity of 1 Limited (LBT1) and 1 Germline (GBT2) antibody are compared for 4 different selection protocols (DNA libraries with 2 different constant regions and 2 DNA library concentrations). Matching symbols correspond to the same experimental protocols. B. 3 Germline antibodies (GNT4, GNT3, GBT3) have higher affinities for the tested DNA ligands than 3 Limited antibodies (LNT3, LNT4, LBT1). The binding scores have been calculated with the dissociation constants measured by ELISA. The binding scores below the detection threshold correspond to affinities that could not be measured experimentally because they are too low.

507 average affinity for the Germline antibodies. These results suggest that the frame-
 508 work of the synthetic antibodies control the maximal average affinity they can reach
 509 for the DNA library. The Germline framework allows antibodies to have a higher
 510 average affinity for DNA than Limited framework. One explanation could be that
 511 maturation against HIV has reduced the selection potential of Limited framework
 512 based antibodies.

- 513 • The different average affinities between Germline and Limited antibodies for DNA
 514 ligands suggest that these antibodies have different global specificities that are con-
 515 trolled by their framework.

516 2.3.2 Are the differences that we have seen systematic within each group 517 of antibodies?

518 We previously studied the effect of the frameworks of TOP synthetic antibodies on their
 519 affinities for DNA ligands. Now we want to study the effect of the CDR3 regions of both

520 TOP and Random antibodies. If the CDR3 region has no impact on the binding affini-
 521 ties, we expect all of the antibodies (TOP and Random) that share the same framework
 522 (Germline or Limited) to have similar binding scores.

523

524 We first used the SELEX technique to calculate the binding scores of antibodies with
 525 the same framework but different CDR3 regions (TOP and Random). We did the same
 526 analysis for both groups, Germline and Limited and we tested 2 different DNA libraries
 527 in parallel (DNA library 1 & 2). The results are presented in (Fig. 2.2). For each exper-
 528 imental condition we observe the same result: the binding scores of the TOP antibodies
 529 are higher in average than the binding scores of the Random antibodies.

530

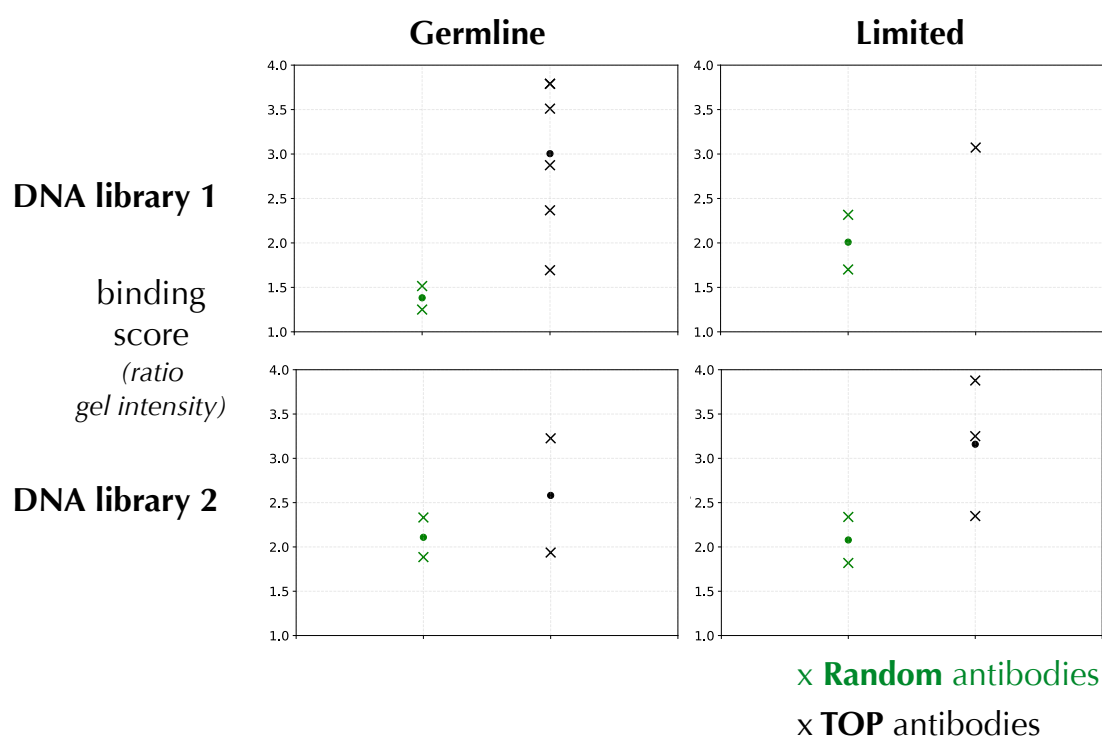


Figure 2.2: **Antibodies with different CDR3 sequences have different average affinities for DNA ligands.** On average, TOP antibodies have a higher affinity for the DNA ligand library than Random antibodies. Crosses represent the individual binding scores and dots the averages. We grouped CDR3 corresponding to TOP antibodies (black crosses) and CDR3 corresponding to Random antibodies (green crosses). The binding scores represent the amount of DNA library collected at the end of the selection experiment (measured with SELEX). We didn't systematically test all of the antibodies against both DNA libraries, which explains the variable number of Germline and Limited antibodies in the different plots.

531

Then we used ELISA to calculate the binding score of 5 different Germline antibodies

532 (that have different CDR3 sequences) for 1 to 6 DNA molecules. The results are presented
 533 in Fig. 2.3. When the affinities were too low to be experimentally measured, we displayed
 534 the binding scores below the detection threshold. We can see that TOP antibodies have
 535 systematically a higher binding score than Random antibodies.

536

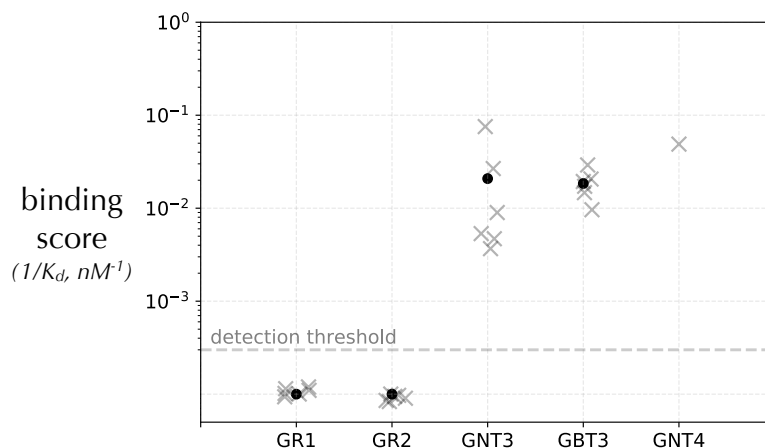


Figure 2.3: **Antibodies with different CDR3 sequences have different affinities for the tested DNA ligands.** TOP antibodies (GNT3, GBT3, GNT4) have higher affinities for the tested DNA ligands than Random antibodies (GR1, GR2). The binding scores have been calculated with the dissociation constants measured by ELISA. We measured the affinities of 5 antibodies for 1 to 6 different DNA ligands. The 6 DNA molecules are composed of 3 different DNA loops (i.e. DNA ligands) with 2 different constant regions for each loop. We used the same experimental conditions to measure the affinities of both TOP and Random Germline antibodies but we couldn't measure the dissociation constant of Random antibodies (their affinity for DNA ligands are too low).

537 **Conclusion:**

- 538 • **SELEX and ELISA experiments show that antibodies with different CDR3**
 539 **sequences have different average affinities for the DNA ligand library.**

540 **2.4 Summary**

- 541 • We measured different average affinities between the 2 groups of antibodies (Germline
 542 & Limited).
 543 – On average, Germline antibodies have a higher affinity for DNA ligands than
 544 Limited antibodies.
 545 • The average affinity is not systematic among antibodies that share the same frame-
 546 work.

547 – TOP antibodies have a higher average affinity for DNA ligands than Random
548 antibodies.

549 **2.5 Discussion**

550 • Our results suggest that binding affinities of antibodies are controlled by both the
551 framework and the CDR3 sequences on 2 different levels:

552 – The framework controls the dynamic range of antibody affinity for DNA ligands

553 * The comparison of TOP antibodies with Germline and Limited frame-
554 works shows a higher average affinity for Germline antibodies than for
555 Limited antibodies.

556 * This observation is consistent with the fact that natural Germline anti-
557 bodies must recognize a broad range of potential targets whereas Limited
558 affinity matured antibodies are specialized in the binding of a particular
559 biological target.

560 – The CDR3 region controls the affinity of the antibody within the dynamic range
561 fixed by the framework

562 • For each antibody, we estimated the average of the affinity distribution for the DNA
563 ligand library. It would be more informative to have access to the full distribution of
564 affinities (because different affinity distributions can have the same average), using
565 the calibration assay presented in chapter 4.

566 Chapter 3

567 Measuring the sequence diversity 568 of DNA ligands selected by 569 antibodies

570 3.1 Context & Problem

571 In chapter 2, we defined the global specificity of antibodies as their ability to bind very
572 different ligands. Using the binding landscape analogy, we presented the global specificity
573 as the long-range ruggedness of their binding landscape. In this chapter, we study the
574 local specificity of antibodies, i.e. their ability to discriminate between ligands that are
575 similar to each other. We can describe local specificity as the short-range ruggedness of
576 their binding landscape.

577 We propose an approach to investigate the local specificity of antibodies by testing their
578 ability to discriminate between sequences of DNA ligands (dubbed “DNA sequences”). We
579 used the same library that was employed in chapter 3 to study global specificity, combined
580 with a high-throughput selection assay. We asked whether different types of antibodies
581 have the same ability to discriminate DNA sequences. We answered the question in 3 steps.
582 First we verified if synthetic antibodies are able to discriminate DNA sequences. Then we
583 compared the selected DNA sequences for antibodies sharing the same framework. And
584 finally we compared the selected DNA sequences for antibodies with different frameworks.
585 Measuring the ability of antibodies to discriminate molecules that are similar to each
586 other in the fitness landscape is a first step in understanding the local specificity of these
587 antibodies.

588 **3.2 How we analyzed the sequences of the selected DNA** 589 **ligands**

590 **3.2.1 Presentation of our minimalist system**

591 In this section, we studied 4 Germline and 4 Limited antibodies. In each group, half are
592 TOP antibodies and half are Random antibodies. We used 2 libraries of 32,768 DNA
593 molecules (DNA library 1&2, Supp. 6.1.2) to measure the local specificity of these anti-
594 bodies.

595 **3.2.2 Analysis of the deep-sequencing data in 3 steps**

596 We selected both DNA ligand libraries against the 4 synthetic antibodies using SELEX
597 (Supp. 6.2.3, we performed a single round of selection). During SELEX, we exposed a
598 DNA ligand library to one antibody and identified the DNA sequences that interacted
599 the most with the latter using deep-sequencing. Deep-sequencing is used to measure the
600 number of occurrences of each DNA ligand before and after its selection. The selected and
601 non-selected DNA libraries are prepared for deep-sequencing with 2 successive PCR reac-
602 tions during which the DNA ligands are amplified and extended with extra nucleotides.
603 The goals of these extra nucleotides are to link each DNA molecule to the selection exper-
604 iment it comes from and to allow its deep-sequencing (for more detail, see protocol 6.2.4).
605 We performed the analysis of deep-sequencing data in 3 successive steps:

606

607 **Step1: Comparison of the copy number of each DNA sequence after and** 608 **before its selection**

609 The principle is to compare affinities of DNA sequences for the antibody by looking at
610 their copy number after and before their selection. DNA sequences that have a high affini-
611 ty for the antibody are significantly more present after than before the selection in the
612 deep-sequencing data, while DNA sequences with a low affinity for the antibody are as
613 much present after than before the selection. We typically observe 2 modes in the results
614 of a selection experiment (Fig. 3.1): The first mode is composed of DNA sequences that
615 have the highest affinities for the antibody. They are located above the diagonal (we call
616 them outliers). The second mode is composed of DNA sequences with lower affinities
617 for the antibody. They are located around the diagonal (we call them bulk sequences).
618 The copy numbers of the outliers after their selection are principally controlled by their
619 affinity for the antibody while the copy numbers of the sequences in the bulk are a mix of
620 selection and experimental noise (for more details, see Supp. 6.3.3) (the study of affinities
621 of sequences from the bulk requires a more sophisticated analysis to extract the selection
622 signal). In order to correct the bias due to the non-uniform representation of each DNA
623 sequence in the initial library, we calculate the enrichment value for each sequence (see
624 below).

625

626 Step2: Measurement of enrichments

627 We quantified the selection of each sequence by measuring its enrichment during selection.
628 The enrichment of a sequence is calculated as the ratio between its number of copies after
629 and before the selection (in this study, we systematically normalized the enrichments so
630 their sum equal 1). We only measured enrichments for copy numbers >10 because smaller
631 copy numbers are not representative of the frequency of the sequences in the total popu-
632 lation (they are dominated by experimental noise, Supp. 6.3.3).

633

634 Step3: Sequence logo

635 We evaluated the diversity of the nucleotides present in the most selected DNA sequences
636 using sequence logos (Supp. 6.11). Sequence logos are commonly used to analyze the
637 diversity of a pool of sequences by representing the information content of each position
638 along the DNA sequence. The information content of a position is maximal when all the
639 sequences of the pool have the same nucleotide at this position.

640 3.3 Results**641 3.3.1 Can antibodies discriminate DNA sequences ?**

642 We first asked if synthetic antibodies discriminate DNA sequences. If they do, we would
643 expect to observe sequences that are significantly more present after than before the se-
644 lection in the deep-sequencing data. We asked this question for the TOP (CDR3 region
645 with a high affinity for a particular DNA sequence) and the Random (CDR3 region with
646 a low affinity for the same DNA sequence) antibodies, separately.

647

648 1- Can TOP antibodies discriminate DNA sequences ?

649

650 We selected a DNA library against one TOP Germline and one TOP Limited anti-
651 body. For each selection experiment, we compared the number of copies of the sequences
652 after and before their selection (Fig. 3.1). We observe outliers in both selection experi-
653 ments. In order to test if the results we observe are significant, we replicated the selection
654 experiments and compared the enrichments between replicates (Fig. 6.9). The highest
655 enrichments are reproducible from one replicate to another (we also compared the enrich-
656 ments between selection and control experiments for each antibody in Supp. 6.4.2 and we
657 showed that the outliers were specifically selected by antibodies). These results confirm
658 that both antibodies discriminate DNA sequences.

659

660 2- Can Random antibodies discriminate DNA sequences ?

661

662 We then tested the ability of one Random Germline and one Random Limited anti-
663 body to discriminate DNA sequences. The results are presented in (Fig. 3.2). We don't

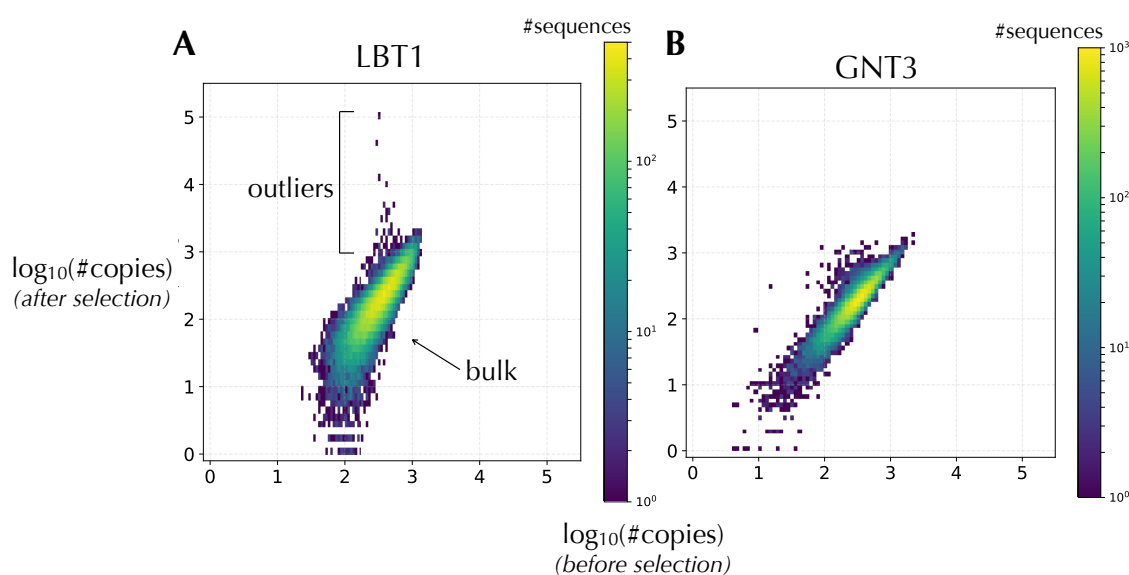


Figure 3.1: **Synthetic antibodies discriminate DNA sequences during SELEX experiment.** Each dot corresponds to a DNA sequence. The number of copies before selection are indicated on the x axis and the number of copies after selection are indicated on the y axis. The selection is specific when the number of copies after the selection is significantly higher than the number of copies before the selection. A. We selected a DNA library against one Limited antibody (LBT1). B. We selected a DNA library against one Germline antibody (GNT3). We observe 2 modes in both graphs: (1) One mode where the copy numbers before and after the selection are similar, this mode is dominated by non-specific binding (2) One mode where the copy numbers are significantly higher after than before the selection. This mode is dominated by specific binding.

664 observe sequences that are significantly enriched in neither of the 2 selection experiments.

665

666 **Conclusion:**

- 667 • **The analysis of 2 Germline and 2 Limited antibodies shows that TOP**
- 668 **antibodies discriminate DNA sequences (that have the same stem but**
- 669 **different DNA sequences in their loop) while Random antibodies do not,**
- 670 **under the same experimental conditions.**

671 **3.3.2 Within each group, do different antibodies select the same DNA**

672 **sequences ?**

673 Before asking whether Limited and Germline antibodies select the same or different DNA

674 sequences, we first need to ask whether Limited antibodies with different CDR3 regions

675 select the same sequences or not, and conversely for Germline. If different antibodies select

676 the same DNA sequences, we expect a correlation between the measured enrichments for

677 these sequences.

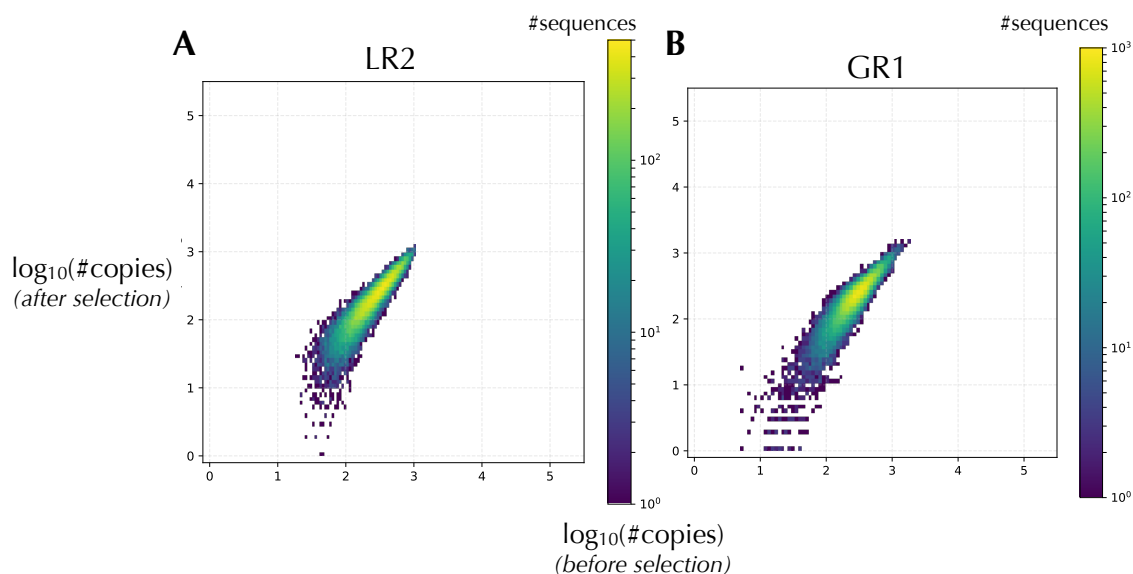


Figure 3.2: **Randomly chosen synthetic antibodies discriminate less DNA ligands than antibodies previously selected against one DNA sequence, under the same experimental conditions.** A. We selected a DNA library against one Random Limited antibody. B. We selected a DNA library against one Random Germline antibody. We don't observe DNA sequences that have been significantly enriched in neither of the 2 selection experiments.

678

679 We compared the enrichments of 2 TOP Germline and 2 TOP Limited antibodies,
 680 the results are presented in (Fig. 3.3). When comparing the 2 Germline antibodies, we
 681 observe a correlation between the most enriched DNA sequences. That means that the
 682 same DNA sequences are selected by both Germline antibodies (We tested 3 other TOP
 683 Germline antibodies, the few most enriched sequences were the same as the one enriched
 684 in Fig. 3.3 (data not shown)). On the contrary, the 2 Limited antibodies select different
 685 sequences (We tested a 3rd Limited antibody that selected DNA sequences different from
 686 the ones in Fig. 3.3B (data not shown)).

687

688 Conclusion:

- 689 • **The study of 2 TOP Germline and 2 TOP Limited antibodies show that**
 690 **antibodies with Germline framework select the same DNA sequences**
 691 **while antibodies with Limited framework selected different DNA sequences.**

692 Discussion:

- 693 • The results suggest that the framework of the synthetic antibodies has an impact
 694 on the role of the CDR3 region during selection
- 695 • We limited our study to few antibodies. For each framework, we would need to

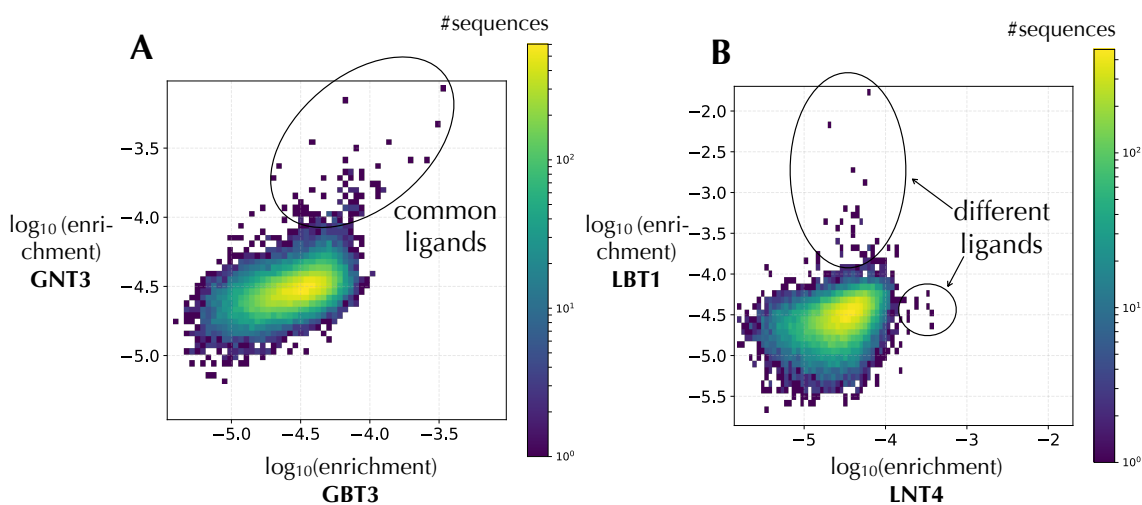


Figure 3.3: **Germline antibodies select the same DNA sequences while Limited antibodies select different DNA sequences.** A. We compare enrichments between 2 Germline antibodies (GNT3 and GBT3). The highest enrichments are correlated for both antibodies, thus these Germline antibodies select the same DNA sequences. B. We compare enrichments between 2 Limited antibodies (LBT1 and LNT4). The highest enrichments are not correlated between both antibodies, thus these Limited antibodies select different DNA sequences.

696 test more CDR3 regions, that are more or less different from each other, in order to
 697 conclude on the role of the framework and CDR3 regions in the binding to the DNA
 698 sequence.

699 3.3.3 To what extent does the constant region of DNA molecules impact 700 the selection of the loop region ?

701 The results encountered so far have been obtained with DNA libraries with different con-
 702 stant regions (DNA library 1&2, Supp. 6.2). We presented the selection of the DNA library
 703 2 by the Limited antibodies and the selection of the DNA library 1 by the Germline anti-
 704 bodies. The next question we can ask is to what extent does the constant region impact
 705 the selection of the loop ? If the constant region of the DNA molecules has no impact
 706 during selection, we expect the antibodies to select the same loops independently of DNA
 707 molecules' constant region.

708
 709 To answer this question, we replicated the selection experiment for TOP Limited an-
 710 tibodies (LBT1 & LNT4) but we changed the constant region of the DNA library. The
 711 results are presented in (Fig. 3.4). In the left plot are compared the enrichments of DNA
 712 loops from library 2 for two Limited antibodies (The most selected DNA sequences are

713 highlighted for each antibody with 2 different colors). In the right plot are compared the
 714 enrichments of DNA loops from library 1 for the same antibodies (the same sequences are
 715 highlighted with the same color-code in plots A and B). The same loops are selected with
 716 both constant regions, but the enrichment values are higher for the DNA library 2.
 717

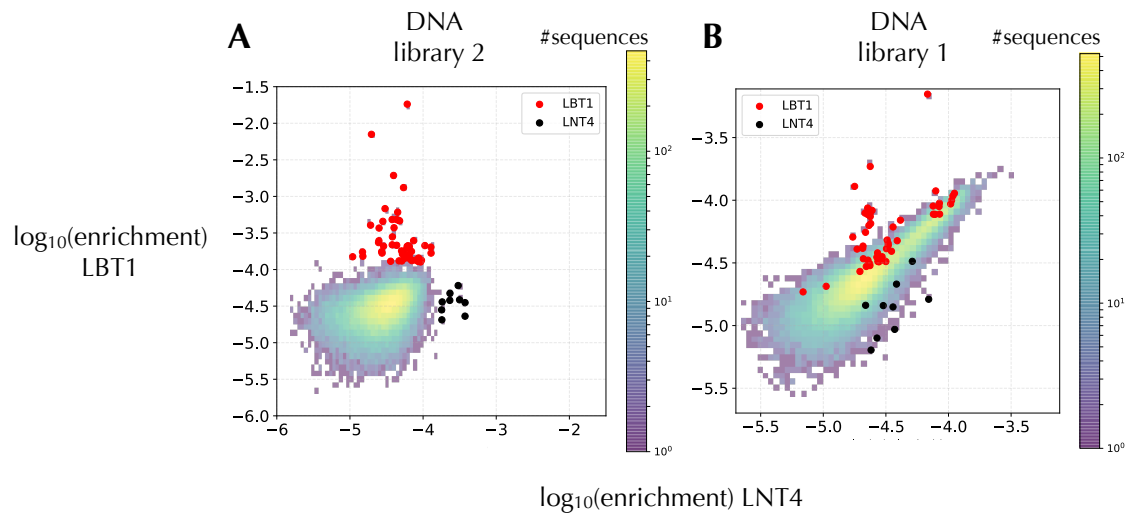


Figure 3.4: **2 Limited antibodies select DNA loops independently of the constant region of the DNA library.** For both Limited antibodies (LBT1 and LNT4), the most enriched DNA loops in the library 2 (marked in red for LBT1 and in black for LNT4) (figure A) are also the most enriched in the library 1 (figure B). The dataset used in figure A is the same as in figure 3.3 B.

718 **Conclusion:**

- 719 • **As far as Limited antibodies are concerned, the constant region of DNA**
 720 **molecules has a small impact on the enrichment of the DNA loops.**
 721

722 **Discussion:**

- 723 • We studied the impact of the constant region of DNA molecules on the selection of
 724 the DNA loops by Germline antibodies as well (Supp. 6.6). Regarding Germline
 725 antibodies, the constant region of DNA molecules has a non-negligible impact on
 726 the selection of the loop (this observation does not change the previous conclusions:
 727 Germline antibodies select the same DNA sequences while Limited antibodies select
 728 different DNA sequences).
- 729 • For Germline antibodies, we observe the strongest selection signal with the DNA
 730 library 1. For Limited antibodies, we observe the strongest selection signal with the

731 DNA library 2. Consequently, for the rest of this chapter, we study the selection of
732 the DNA library 1 by Germline antibodies, and the selection of the DNA library 2
733 by Limited antibodies.

734 **3.3.4 Does the antibody framework impact the sequence diversity of the** 735 **selected DNA molecules ?**

736 We compared the diversity of DNA sequences selected by Germline and Limited antibodies.
737 We studied the selected sequences with 2 levels of resolution: we first studied the selection
738 of the entire sequence of the loop (7 nucleotides) then we studied the selection of shorter
739 sequences, i.e. DNA motifs, inside the loop (less than 7 nucleotides).

740 **3.3.4.1 Resolution = DNA loop**

741 We first analyzed the selection of the entire loop by Germline and Limited antibodies. In
742 order to evaluate the diversity of the selected sequences, we generated the sequence logo of
743 the outliers for 2 TOP Germline and 2 TOP Limited antibodies. The results are presented
744 in (Fig. 3.5). The qualitative comparison of the sequence logos shows that the most se-
745 lected sequences are different between the 2 Limited antibodies and similar between the
746 2 Germline antibodies. The quantitative comparison of the sequence logo shows that the
747 average information content of the selected nucleotides is higher for the Limited antibodies
748 than for the Germline antibodies. Thus, each Limited antibody selects DNA ligands with
749 similar sequences while each Germline antibody selects DNA ligands with more diverse
750 sequences.

751

752 **Conclusion:**

- 753 • **The framework of the antibodies we tested impacts the diversity of the**
754 **selected DNA sequences**
- 755 • **The sequence diversity of the selected DNA loops is smaller for the 2**
756 **Limited antibodies than for the 2 Germline antibodies.**

757

758 **Discussion:**

- 759 • We generated new sequence logo for GNT3 antibody using smaller numbers of DNA
760 sequences in order to verify if the different information contents we observed in
761 Fig. 3.5 are due to antibody binding specificity or sampling size (Supplementary
762 Fig. 6.11). We obtain similar sequence logos independently of the number of DNA
763 sequences we include in the analysis (for GNT3).
- 764 • Because we limited this analysis to the outliers for each selection experiment, we were
765 reduced to the analysis of a sub-sample of DNA sequences (between 9 and 150 out
766 of the 32,768 present in the library). We could redo the same selection experiments

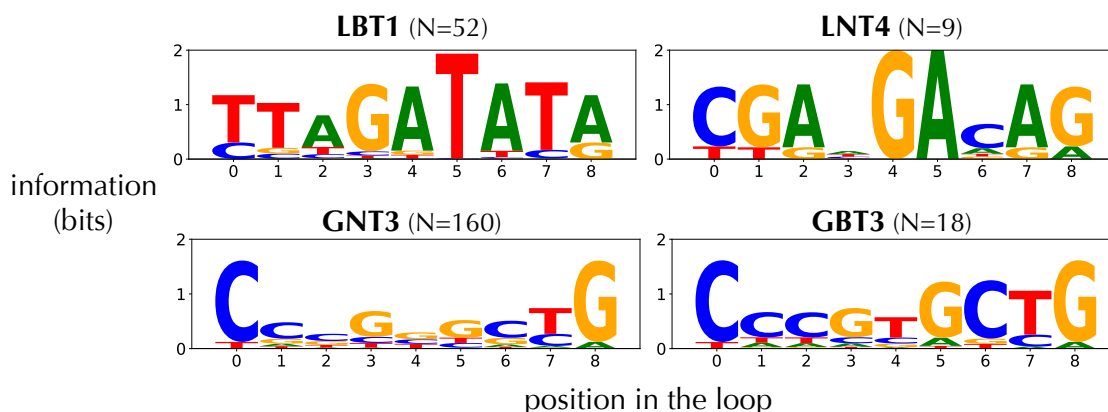


Figure 3.5: **Limited antibodies select DNA ligands with similar sequences (high information content) while Germline antibodies select DNA ligands with different sequences (low information content).** Here we present the 9 nucleotides forming the loop (including the first (position 0) and last (position 8) nucleotides that close the loop). We represented the sequence logo of the most enriched loops for 2 Germline (GNT3, GBT3) and 2 Limited (LBT1, LNT4) antibodies. Each sequence logo has been generated with different numbers of sequences (N) because we only considered the outliers in our analysis (the outliers are dominated by specific binding).

767 with optimal experimental conditions to maximize the number of outliers for each
 768 antibody

- 769 • The optimal experimental conditions that maximize the number of outliers can be
 770 identified with a calibration experiment (Chapter 4)
- 771 • It is likely that Germline antibodies select loop sequences that disorganize the stem-
 772 loop structure of the DNA library 1. Most of the enriched DNA loops have a sequence
 773 that matches a sub-region of the stem. These loop sequences can perturb the stem
 774 loop folding of the DNA molecules and generate alternative 3D structures. The
 775 selection of these particular DNA sequences has only been observed for Germline
 776 antibodies.

777 3.3.4.2 Resolution = DNA motifs inside the loop

778 We previously analyzed the selection of the entire loop by antibodies. The sequence logos
 779 in Fig. 3.5 showed conserved nucleotides at consecutive positions which suggests the se-
 780 lection of DNA motifs inside the loop.

781

782 1) Do antibodies select DNA motifs ?

783

784 If an antibody selects specific DNA motifs, we expect to observe the same motifs mul-
785 tiple times in the most enriched sequences.

786

787 We did a simple test where we identified a potential DNA motif (“ATAT”) in the
788 sequence logo of a TOP Limited antibody (Fig. 3.5, LBT1) and we highlighted all the
789 DNA sequences that contain this motif (at the 4th position in the loop) in the Fig. 3.6.
790 Half of the most selected DNA sequences share this motif.

791

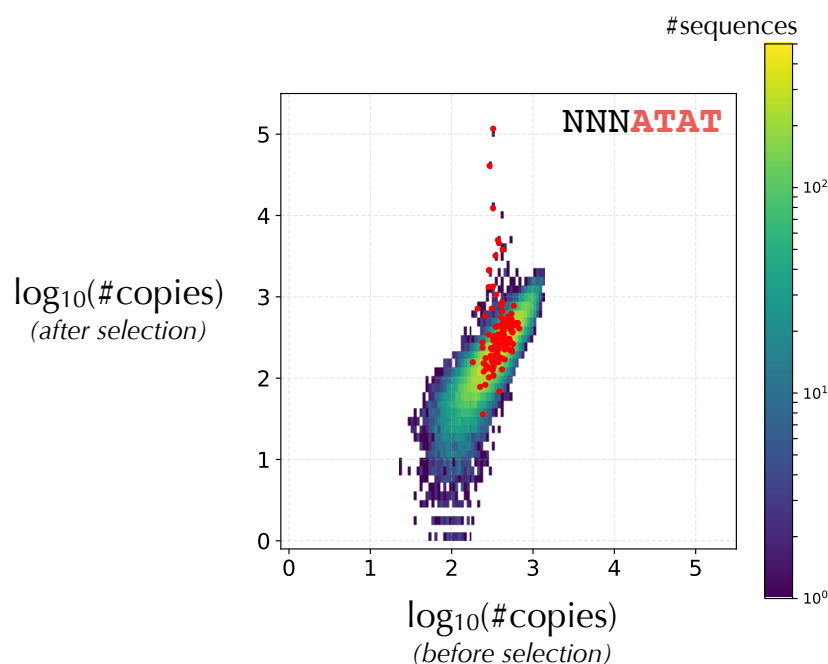


Figure 3.6: **Synthetic antibodies can select 4 nucleotides motifs inside the loop of the DNA molecules.** We highlighted in red all the sequences that share the same motif “ATAT” at the fourth position in their loop. The DNA motif is present in half of the outliers. The dataset used in this figure is the same as in figure 3.1 A.

792

Conclusion:

793

- The TOP Limited antibody (LBT1) selects DNA motifs inside the 7 nucleotides loop of the DNA molecules.

794

795

796

Discussion:

797

- We observe DNA sequences containing this motif in the TOP region of the bulk.

798

- This observation suggests that despite the fact that the experimental noise is dominant in the bulk (Supp. 6.3.3), the distribution of the sequences in this region also

799

800 depends on their affinity for the antibody.

801

802 2) Can we include all the sequences (outliers and bulk) in the study of the selection of
803 motifs ?

804

805 The analysis of the selected DNA loops showed enrichments that were not reproducible
806 (between replicate experiments) for the smallest values (Fig. 6.9). We propose to analyse
807 the selection of DNA motifs inside the loops and see if we can obtain more reproducible
808 enrichments.

809 To do so, we compared the enrichment values between replicate experiments and we
810 used the same datasets for both analyses (first and second levels of resolution). The com-
811 parison of the enrichments with the first analysis (selection of loops) is presented in Fig.
812 3.7A and the comparison of the enrichments with the second analysis (selection of motifs)
813 is presented in Fig. 3.7B (For more details about the analysis of motifs, see Supp. 6.3.5,
814 here we analyzed the 4 nucleotides motifs located at the beginning of the loops). The
815 analysis of selected motifs gives more reproducible enrichment values (The coefficient of
816 determination that was used to assess the reproducibility of the measurements was in-
817 creased by almost one order of magnitude).

818

819 **Conclusion:**

820 • The “whole-loop” level of resolution limits our study to the outliers (they
821 are the only sequences with reproducible enrichments).

822 • The “motif” level of resolution allows us to include all the sequences
823 (outliers and bulk sequences) in our analysis.

824

825 **Discussion:**

826 • The reason why enrichments of motifs are more reproducible than enrichments of
827 loops can be explained by the fact that the enrichments of motifs correspond to the
828 average enrichments of more than 100 sequences containing these motifs in their
829 loop. Consequently, the sampling bias that was dominant in these sequences is av-
830 eraged out and we can partially extract the selection signal from these sequences.

831

832 3) Do all motif parameters carry information ?

833

834 In this section we call parameters the length of the motifs we study and their position
835 along the loop. We previously chose arbitrary parameters to study the selection of DNA
836 motifs. Here we ask if all parameters can be used to study the selection of DNA motifs
837 by synthetic antibodies: do the length of the motifs and their position on the loop matter

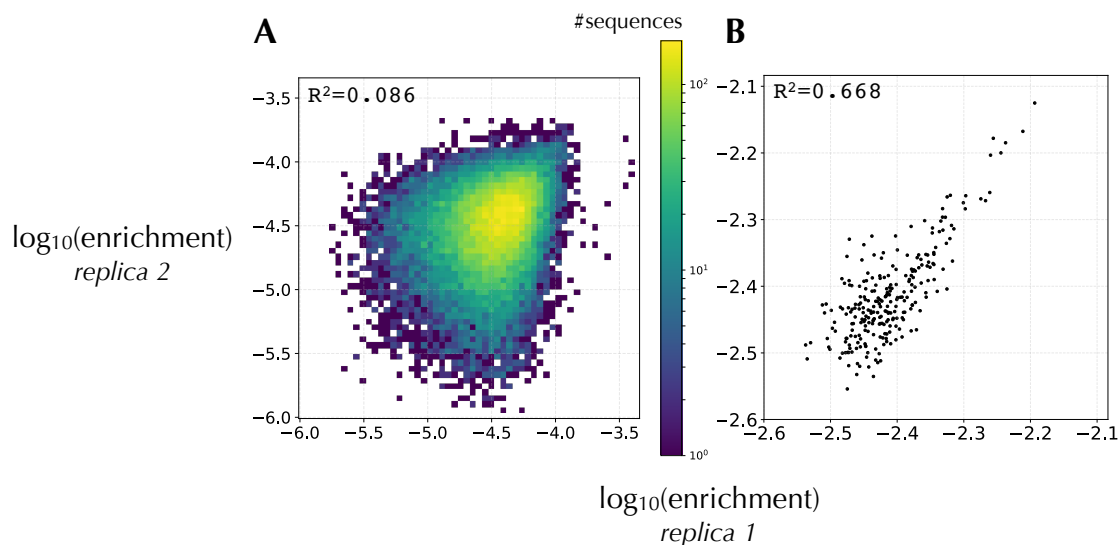


Figure 3.7: **The average enrichments of the motifs are more reproducible than the individual enrichments of the DNA loops.** We used the same datasets (DNA library selected by LNT4, in 2 replicates) to compare 2 types of analysis A. We compared the enrichments of individual DNA loops (1st level of resolution) between 2 replicate experiments. The R^2 score indicates a low reproducibility between the 2 replicate experiments. B. We compared the average enrichment of the 4 nucleotides DNA motifs (2nd level of resolution) (located at the beginning of the loop) between replicate experiments. This analysis allows a 7.8 folds increase of the R^2 score and thus increases the reproducibility of the enrichments.

838 ? We had two different criteria to estimate the quality of the parameters we chose:

839

840 **1- The parameters must maximize the cross-validation score.**

841 The principle of cross-validation is the following: we split the dataset (containing the en-
 842 riched sequences) in 2, test the same analysis on both half of the dataset separately and
 843 then compare the enrichments of each motif calculated in each sub-dataset. If we mea-
 844 sure the same enrichments in both sub-datasets (equivalently, if the R^2 score associated
 845 to the comparison of both dataset tends to 1), it means that the analysis can be used to
 846 accurately describe the data (it will give the same results for replicate experiments). For
 847 each set of parameters we tested, we systematically measured the corresponding cross-
 848 validation score.

849

850 **2- The parameters must minimize the correlation between the selection and
 851 its control experiment.**

852 The principle is to test the same analysis on both the selection experiment (where we
 853 select a DNA library against one antibody) and its control experiment (where we select
 854 the same DNA library against an empty plate). Then we compare the enrichments of each

855 motif calculated for each experiment. If the enrichments are the same between the exper-
856 iments (equivalently, if the R^2 score associated to the comparison of both dataset tends
857 to 1), it means that the parameters we chose didn't allow us to identify the motifs that
858 were specifically selected by the antibody. In this case we must change the parameters
859 and repeat the procedure until we observe DNA motifs that are specifically selected by the
860 antibody (and are not selected in the control experiment). For each set of parameters we
861 tested, we systematically compared the enrichments measured in the selection and control
862 experiments, in order to identify DNA motifs that are specifically selected by the antibody.

863
864 We do not present the systematic investigation of these parameters. We present in
865 (Fig. 3.8) 2 sets of parameters (length=3, position=1 & length=4, position=4): In the
866 first one, the parameters maximize the cross validation score but generate also a high
867 correlation between the selection experiment and its control. In the second one, we obtain
868 a smaller cross-validation score but a more specific signal of the selection by the antibody.

869
870 **Conclusion:**

- 871 • **The choice of the length and position of the motifs in the loop is critical**
872 **to capture the specific selection of motifs by antibodies.**

873
874 4) Can the same parameters be used to study the selection of motifs for different antibod-
875 ies (Germline and Limited)?

876
877 We previously showed that different antibodies can select different DNA loops (Fig.
878 3.5). Here we ask if these antibodies can also select different DNA motifs. We addressed
879 this question using a single set of parameters for all the antibodies (length=4 and posi-
880 tion=4). We analyzed the selected motifs of different Limited and Germline antibodies in
881 2 steps: first we compared the enrichments of the selected motifs, then we compared their
882 DNA sequences (for each antibody, we verified the cross validation score and the correla-
883 tion between selection and control experiments, as explained in the previous section, see
884 Supplementary Fig. 6.14 & Supplementary Fig 6.15).

885
886 The comparisons of the enrichments between Limited and between Germline antibod-
887 ies are presented in (Fig. 3.9). The results show that a unique set of parameters allows us
888 to identify specific motifs for each Limited antibody and common motifs for both Germline
889 antibodies (We observed the same trends when we compared the enrichment of the entire
890 loop (Fig. 3.3)). Note that despite the fact that both Germline antibodies tend to select
891 the same DNA motifs, we can see that the most selected motifs are not all the same for
892 both antibodies.

893
894 We then studied the diversity of the selected motifs using the sequence logo representa-

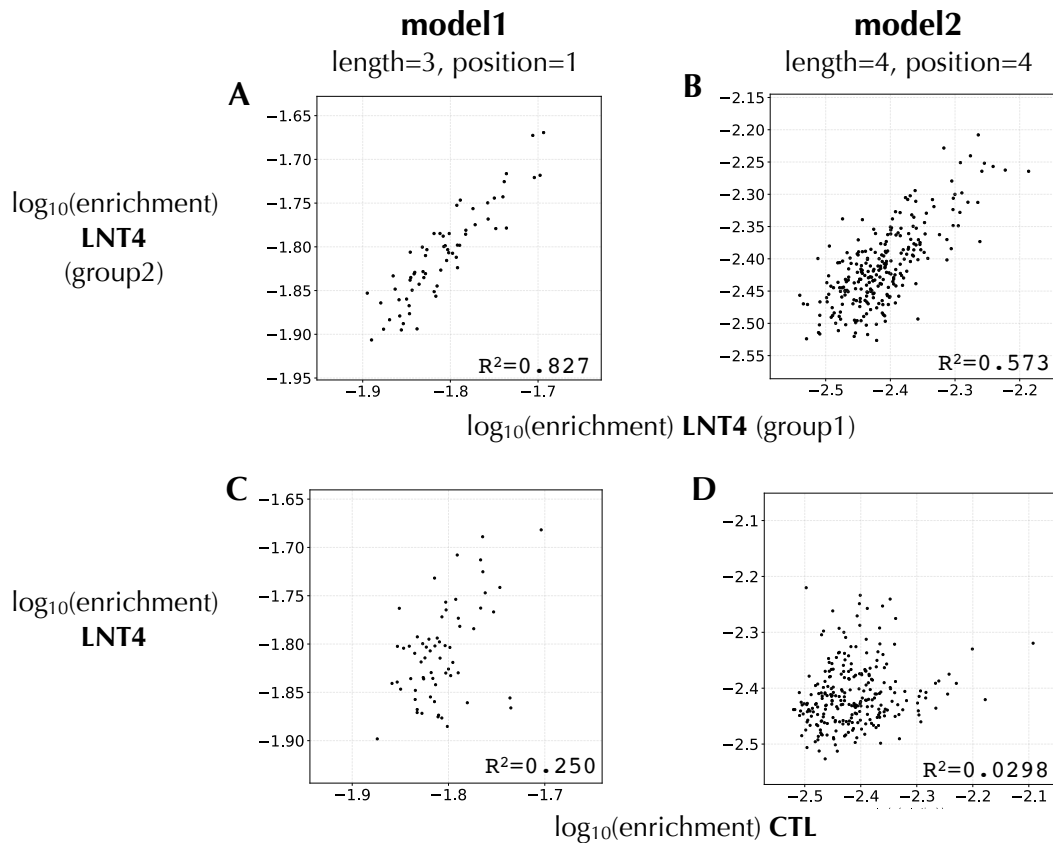


Figure 3.8: **The choice of the length and position of motifs inside the loops in the second level of analysis is critical to measure specific binding.** The TOP plots (figA and B) show cross-validation results of average enrichments of DNA motifs (a good cross-validation is characterized by a high R^2 score). The bottom plots (fig C and D) show the comparison of enrichments between selection experiment (performed with an antibody) and control experiment (performed without antibody). Specific selection signal is characterized by orthogonal enrichments and low R^2 score. In the first column, we consider motifs of 3 nucleotides located at the beginning of the loop. Despite the high cross validation score (figA), we cannot clearly identify DNA motifs that are specifically selected by the antibody (figC). In the second column, we consider motifs of 4 nucleotides located at the end of the loop. The cross validation score is smaller (figB) but we can identify DNA motifs specifically selected by the antibody (figD).

895 tion. We build a sequence logo of the 5 most selected DNA motifs (corresponding to more
 896 than 500 selected DNA sequences) for each Limited and Germline antibodies. The data
 897 are shown in (Fig. 3.10). For both Limited antibodies, we can notice the presence of the
 898 motifs “ATAT” for LBT1 and “GACA” for LNT4 (they are also visible in the sequence
 899 logo of the most enriched DNA loops, in Fig. 3.5 LBT1, LNT4). We observe similar
 900 sequence logos for LNT4 and GBT3 antibodies. The average information content in each
 901 sequence logo is similar for both the Limited and Germline antibodies.

902

903 **Conclusion:**

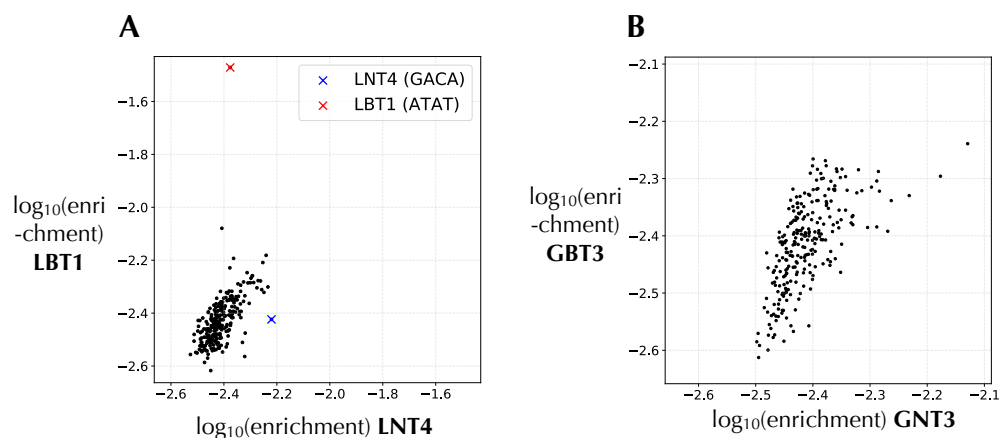


Figure 3.9: **Limited antibodies select different DNA motifs while Germline antibodies select the same motifs, in average.** We calculated the average enrichments of 4 nucleotides motifs located at the end of the loop for 2 Limited (LNT4 and LBT1) and 2 Germline (GNT3 and GBT3) antibodies. A. Comparison of enrichments between the 2 Limited antibodies. Both antibodies selected different DNA motifs. The red cross corresponds to the most selected DNA motif 'ATAT' by LBT1, that we could already observe in the sequence logo of the most selected loops (Fig. 3.5, LBT1). The blue cross corresponds to the most selected DNA motif 'GACA' by LNT4, that we could already observe in the sequence logo of the most selected loops in (Fig. 3.5, LNT4). Comparison of enrichments between the 2 Germline antibodies. Both antibodies select the same DNA motifs.

- 904 • The use of a unique set of parameters allows us to identify motifs that
 905 are specifically selected by Germline and Limited antibodies.

906

907 **Discussion:**

- 908 • For each antibody, we can measure specific binding of DNA motifs with different set
 909 of parameters.

- 910 • The next challenge would be to systematically and automatically look for the optimal
 911 parameters for each antibody.

- 912 • We can test models where the position of the motifs is not fixed within the loop.

913 **3.4 Summary**

- 914 • The first level of analysis informs us that synthetic antibodies can discriminate DNA
 915 loops of 7 nucleotides in a library of 32,768 DNA molecules.

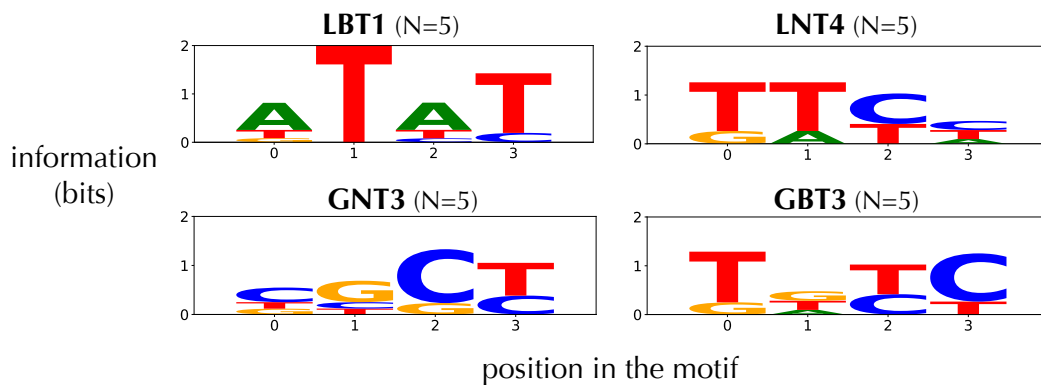


Figure 3.10: **Sequence logo of the 5 most selected DNA motifs for 2 Germline and 2 Limited antibodies**

- 916 • The second level of analysis informs us that synthetic antibodies recognize 4 nu-
917 cleotides DNA motifs inside the loops .
- 918 • The constant region of DNA molecules has minimal impact on the selection of the
919 loop for Limited antibodies.
- 920 • For the 4 tested antibodies, we observed that different Germline antibodies select
921 the same DNA loops while different Limited antibodies select different DNA loops.
- 922 • The framework impacts the diversity of the selected DNA sequences.
- 923 • The diversity of the selected DNA sequences is higher for Germline than for Limited
924 antibodies.
- 925 • The second level of analysis based on the selection of DNA motifs allowed us to
926 extract selection information from sequences dominated by the sampling noise.

927 3.5 Discussion

- 928 • The 2 groups of antibodies we tested select DNA sequences with different levels of
929 diversity
- 930 • In order to verify if the differences we observed are systematic for all the antibodies
931 that share the same framework, we need to study more Germline and more Limited
932 antibodies.

- 933 • When using sequence logos to study the diversity of selected DNA sequences (at
934 the scale of the loop, or at the scale of the motif), we made the assumption that
935 each nucleotide contributes additively to the binding of the DNA molecule to the
936 antibody (we didn't investigate the correlation between the different nucleotides in
937 the sequence).
- 938 • With more time, we could test more sophisticated models that take into account the
939 inter-dependency of the positions in the loops (or in the motifs) by adding parameters
940 that describe the correlation between nucleotides that are far away in the sequence.
- 941 • The measurement of the diversity of selected DNA sequences for different antibodies
942 informed us on their ability to discriminate between similar DNA ligands.
- 943 • Based on these results, we can formulate hypotheses regarding the local specificity
944 of these antibodies for nucleic acids.
- 945 • Despite the small number of antibodies we tested, we can make the hypothesis that
946 Limited antibodies have a higher local specificity than Germline antibodies.
- 947 • Limited antibodies have small affinities for DNA ligands but discriminate with high
948 precision the DNA sequences.
- 949 • Germline antibodies have high affinity for DNA ligands but discriminate with low
950 precision the DNA sequences.
- 951 • Beyond these conclusions, we have set-up tools that can be in general applied to
952 compare binding specificity profiles.

953 Chapter 4

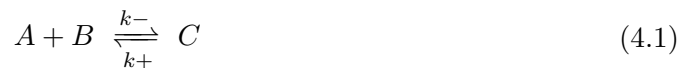
954 High-throughput experimental 955 measurements of binding affinities

956 4.1 Introduction

957 4.1.1 Context & Problem

958 In previous chapters, we schematized the specificity profile of proteins as a binding affinity
959 landscape. We proposed a high-throughput selection assay to investigate the long-range
960 ruggedness of the landscape (or global specificity) by estimating the average affinity of
961 antibodies for DNA ligands (Chapter 2). We then investigated the short-range ruggedness
962 of the landscape (or local specificity) by measuring the sequence diversity of DNA ligands
963 recognized by the antibodies (Chapter 3). In these approaches, we used enrichments,
964 i.e. ratio of the number of copies of DNA ligands after and before their selection by the
965 antibody, as proxy of affinity (the highest enrichments correspond to the highest affinities).
966 However, the relationship between both measurements is not trivial. In order to rigorously
967 investigate the binding affinity landscape of antibodies, we need to systematically measure
968 affinities of antibody for DNA ligands.

969 Let us consider a simple reversible binding reaction where one ligand A meets one
970 target B to form a complex C:



k_+ is the association rate constant and k_- the dissociation rate constant. One commonly used definition of the affinity is the dissociation constant (K_d) with units of moles per liter (the lower the values of K_d , the stronger the binding reaction).

$$K_d = \frac{k_-}{k_+} = \frac{[A]_{eq}[B]_{eq}}{[C]_{eq}} \quad (4.2)$$

971 The dissociation constant (K_d) is generally measured in the lab using equilibrium experi-

972 ments. The principle is to measure the equilibrium concentration of the ligandA-targetB
973 complex over a range of concentrations of one of the reactants (ligand A or target B)
974 in order to build a binding curve from which we can infer the dissociation equilibrium
975 constant. The concentration of complexes are generally measured through chemical and
976 optical assays which allow the measurement of one dissociation constant at a time.

977

978 The general question of this chapter is the following: How to combine a high-throughput
979 selection assay with the quantitative measurement of affinities to obtain the binding affin-
980 ity profile of protein targets for a library of ligands, in a single experiment?

981 4.1.2 State of the art

982 Several high-throughput approaches have been designed to measure multiple dissocia-
983 tion constants in parallel (generally between a library of ligands and a target, except for
984 [Aditham et al., 2021] where they measured the dissociation constants of multiple lig-
985 ands for multiple targets in parallel). In [Aditham et al., 2021] the authors measured
986 the affinities of more than 1500 different transcription factor-DNA complexes by perform-
987 ing each binding experiments in separate environments using microfluidic protein arrays
988 (STAMMP). They used a microfluidic platform and designed more than 1500 reaction
989 chambers in which the binding curves were separately monitored with fluorescence mea-
990 surements.

991 In [Adams et al., 2016], the authors used yeast display to study the interaction between a
992 library of recombinant antibodies (library of ligands) and an antigen (target) by building
993 thousands of binding curves in parallel. During yeast display, the library is expressed
994 on the surface of cells (each cell displays a single type of antibody, and each type of an-
995 tibody is present on multiple cells). The target is labeled with a fluorescent tag so the
996 stronger the affinity between the target and the antibody, the more fluorescent the cell will
997 be. After the incubation, the yeast with the highest fluorescent signal are collected and
998 deep-sequenced. This selection experiment has been repeated several times with different
999 concentrations of targets to build the binding curves. The yeast display approach has also
1000 been used in [Kowalsky and Whitehead, 2016] to measure in a high-throughput fashion
1001 the dissociation constant of dockerin protein domain (target) for a library of thousands of
1002 cohesin domain mutants (library of ligands), in parallel.

1003 In [McGeary et al., 2019], [Lambert et al., 2014] and [Dominguez et al., 2018], the
1004 authors studied the binding of different proteins (targets) to RNA sequences (library of
1005 ligands) using the RBNS approach. The RBNS is a combination of 2 techniques (SELEX
1006 and Bind-n-Seq) that have been adapted to the study of RNA-proteins interaction and to
1007 the generation of binding curves. Multiple binding experiment are performed in parallel
1008 with various concentrations of the protein target. The reactants are exposed for one hour
1009 and collected using a solid surface as nitrocellulose membrane, resin or magnetic beads.

1010

1011 In these studies, the authors managed to significantly enhance the number of dissociation constants we can measure in the lab by performing more than 1500 affinity measurements in a single experiment. In the studies where all the binding reactions happen in the same mix (yeast display and RBNS), the procedure is always the same: the authors fixed the concentration of the library of ligands and varied the concentration of the target. The reactants are free during the binding reaction, and the complexes are collected by various means (sorting according to fluorescent signal or capture with solid surfaces). The studies based on microfluidics require sophisticated experimental setup that are not accessible to all the labs. Using microfluidics or FACS require extra steps of optimization before they can capture the signal of interest. RBNS has been developed for the study of RNA and thus contained sensitive steps like transcription to generate RNA ligands and reverse transcription to deep-sequence RNA sequences. Each new approach has its own sources of noise: In FACS based approaches, the binding model must take into account the variable number of proteins of interest that are expressed on the surface of the yeasts as well as the autofluorescence of the cells that can interfere with the fluorescence of the selection. In the RBNS experiments, the non-specific recovery of the free molecules must be included in the binding model to obtain a more accurate description of experimental data.

1029 4.1.3 Our approach

1030 In this project, we designed a calibration experiment that is a cheap, easy to set-up and high throughput approach to measure multiple K_d in parallel between a library of DNA ligands and a target protein. We re-purposed SELEX (Systematic evolution of ligands by exponential enrichment) which is a high-throughput binding assay used to identify the DNA ligands with the highest affinities for a protein of interest. The output of this assay are deep-sequencing data (the sequence of several millions of DNA molecules are read among those present in the output of the assay) that contain the sequences of the DNA ligands that interacted the most with the protein. While SELEX is generally used to qualitatively identify the best DNA binders for a protein of interest with little knowledge about the fundamental drivers and the crucial parameters of this experiment, we propose to adapt the protocol of SELEX to the problem of measuring dissociation equilibrium constant by performing multiple SELEX experiments over a range of concentrations of one of the reactants. From deep-sequencing data, we measure the number of copies of DNA ligands after and before their selection in order to calculate their enrichments (the enrichments are normalized so their sum equal 1) for each tested concentration, then we build their individual binding curves from which the dissociation constant can be inferred.

1046
1047 In this chapter we propose to: (1) perform the calibration experiment and build the binding curves of 1 target protein for 218 DNA ligands in parallel, (2) identify the factors that link the experimental measurements of enrichments to affinities and (3) explore the

1050 experimental parameters that maximize the resolution of the calibration experiment.

1051

1052 Our goal is to propose an experimental setup that measures quickly and with little
1053 material the affinity profile of a synthetic antibody for a library of single-stranded DNA
1054 ligands. The long term goal is to apply this technique to several synthetic antibodies in
1055 order to quantitatively compare their binding profiles for this DNA library.

1056 4.2 Presentation of the calibration experiment

1057 4.2.1 Our ligand-target system

1058 The target model used in this section is a synthetic antibody (“GNT3”) designed and
1059 studied as part of previous works [Boyer et al., 2016] [Schulz et al., 2021]. The syn-
1060 thetic antibody is a recombinant human heavy chain that consists in a 100 amino-acids
1061 sequence acting as a scaffold (the sequence comes from a Germline(-reversed) antibody),
1062 and a CDR3 sequence of 4 amino-acids (R, T, K and H) acting as a binding site (Fig. 1.3).

1063

1064 The ligand model is a single-stranded DNA molecule with a stem-loop structure (Fig.
1065 6.2). We designed a library of 218 different DNA ligands that have the same stem but
1066 different loop sequences. This minimalist DNA library was used for the proof of concept
1067 of our calibration experiment.

1068

1069 4.2.2 The calibration experiment is a high-throughput technique to mea- 1070 sure the K_d of a protein for 218 ligands in parallel

1071 The goal of this high-throughput calibration experiment is to generate binding curves of
1072 218 DNA ligands in parallel. The procedure is the following:

1073

1074 **step1 : Selection experiments**

1075 The calibration assay consists in 7 SELEX experiments that are performed in parallel, in
1076 separate wells. Different concentrations of the DNA library (0.1pM, 1pM, 10pM, 100pM,
1077 1nM, 10nM, 100nM and 1μM) are added to the same quantity of immobilized antibody
1078 (12.5 pmol). The DNA and antibodies are incubated together for one hour, and the free
1079 DNA molecules are removed with multiple washing steps. The DNA molecules in complex
1080 with the antibody are detached and collected for the next step (Supp. 6.2.3).

1081

1082 **step2 : Deep-sequencing**

1083 Deep-sequencing is used to measure the number of occurrences of each DNA ligand before
1084 and after its selection. The selected and non-selected DNA libraries are prepared for deep-
1085 sequencing with 2 successive PCR reactions during which the DNA ligands are amplified
1086 and extended with extra nucleotides. The goals of these extra nucleotides are to link each

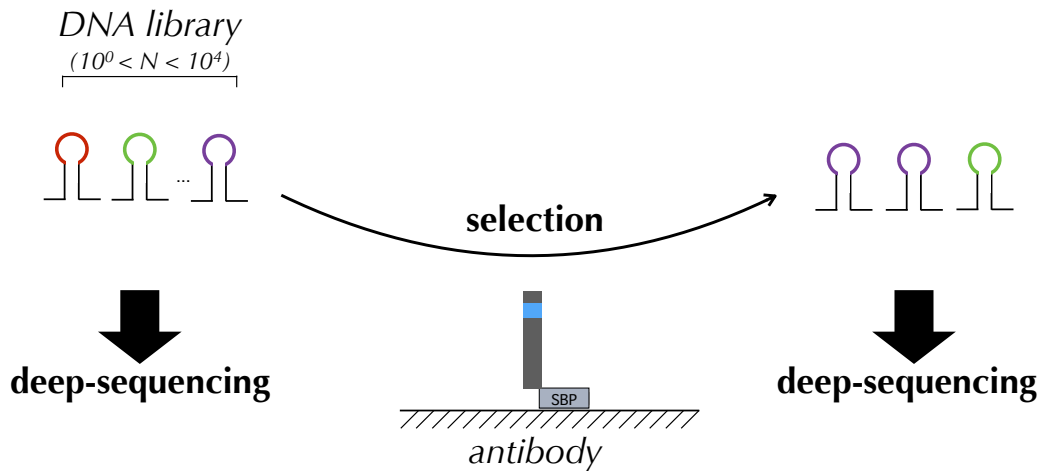


Figure 4.1: **The high-throughput calibration assay is based on SELEX approach.** (The calibration assay is composed of 7 SELEX experiments performed in parallel with the same concentration of antibody but different concentrations of DNA library). For each SELEX experiment, a DNA library (whose diversity can reach 10^4 DNA ligands) is selected against a synthetic antibody immobilized on a solid surface. After one hour of incubation, DNA ligands with the lowest affinities are washed away and DNA ligands with the highest affinities are collected and deep-sequenced to be identified and quantified.

1087 DNA molecule to the selection experiment it comes from and to allow its deep-sequencing
 1088 (for more detail, see protocol 6.2.4).

1089

1090 **step3 : Enrichments and binding curves**

1091 We calculated the enrichment of each DNA ligand as the ratio between its number of
 1092 copies after and before its selection against the antibody (we systematically normalized
 1093 the enrichment values so their sum equal 1). Here we assess that all the calculated en-
 1094 richments are representative of the affinities between the DNA ligands and the antibody
 1095 (i.e. the enrichment values are positively correlated with affinities). This assumption has
 1096 limitations: DNA ligands with small copy numbers are particularly exposed to experi-
 1097 mental biases such as sampling¹ (Supp. 6.3.3). As a result, enrichment calculated with
 1098 small copy numbers does not reflect the affinity of the DNA ligand for the antibody. In
 1099 this section, we assumed that this experimental bias was negligible and we plotted the
 1100 enrichment of each DNA ligand according to the DNA library concentrations that were
 1101 used during selection to build the 218 binding curves.

¹The sampling is the sudden reduction of the number of DNA ligands during the calibration experiment. The most critical sampling steps are the following: (1) during selection, a limited number of DNA ligands can interact with the antibodies and (2) for deep-sequencing, we only collect a small fraction of the total DNA ligand population. During these sampling steps, the small copy numbers are more sensitive to variability and consequently do not reflect the frequency of the DNA ligands in the total population.

1102

1103 To sum-up, we used 7 wells to obtain 218 binding curves in a single calibration exper-
1104 iment. From these binding curves, we proposed 2 approaches to measure the dissociation
1105 constants in this chapter.

1106 **4.2.3 ELISA is used as a low-throughput technique to measure the K_d** 1107 **of a protein for one ligand at a time**

1108 From the library presented above, we took 3 DNA ligands and measured their affinity for
1109 the antibody (GNT3) with another approach (ELISA) in order to compare these results
1110 with the data obtained with the high-throughput calibration assay. ELISA assay described
1111 below allows the measurement of one binding curve at a time:

1112

1113 **step1 : Selection experiments**

1114 The calibration procedure consists in 8 binding experiments that are performed in paral-
1115 lel, in separate wells. Different concentrations of antibodies (one single antibody is used,
1116 in multiple copies) are added to the same quantity of immobilized DNA molecules (one
1117 single DNA sequence is used, in multiple copies). The DNA molecules and antibodies
1118 are incubated together for one hour, and the free antibodies are removed with multiple
1119 washing steps. The antibodies in complex with the DNA molecules are quantified in the
1120 next sep (Supp. 6.2.2).

1121

1122 **step2 : Quantification of the formed complex and binding curve**

1123 The concentration of formed complexes at equilibrium is indirectly measured via the
1124 ELISA colorimetric assay (The absorbance measured in each well is positively correlated
1125 with the concentration of formed complexes). We plotted the measured absorbance ac-
1126 cording to the antibody concentration that was used during the selection to build the
1127 binding curve. From the binding curve, we inferred the dissociation constant (The proce-
1128 dure is explained in Supp. 6.2.2).

1129

1130 To sum-up, we used a total of 24 wells to obtain 3 binding curves. We directly inferred
1131 the dissociation constant from each binding curve using a Hill function.

1132 **4.3 Results**

1133 **4.3.1 Our calibration experiment allows us to generate 218 binding** 1134 **curves at the same time**

1135 **4.3.1.1 This experimental setup provides a high-throughput access to enrich-** 1136 **ment values**

1137 We selected the 218 DNA ligands of the minimalist library against the synthetic antibody
1138 and measured their binding curves using the calibration experiment presented above. The

1139 results are shown in Fig. 4.2. We can observe 3 regimes of DNA concentration. The first
 1140 and third regimes, corresponding respectively to DNA concentrations smaller or larger
 1141 than 10^2 nM, show enrichment values of antibody-DNA complexes that are closer to each
 1142 other. The second regime, corresponding to the DNA concentrations around 10^2 nM, cor-
 1143 responds to the regime where the enrichments are the most different from each other (the
 1144 ranking between the enrichment values is the most conserved among DNA concentrations
 1145 ranging from 10^1 to 10^3 nM ,supplementary Fig. 6.16).
 1146

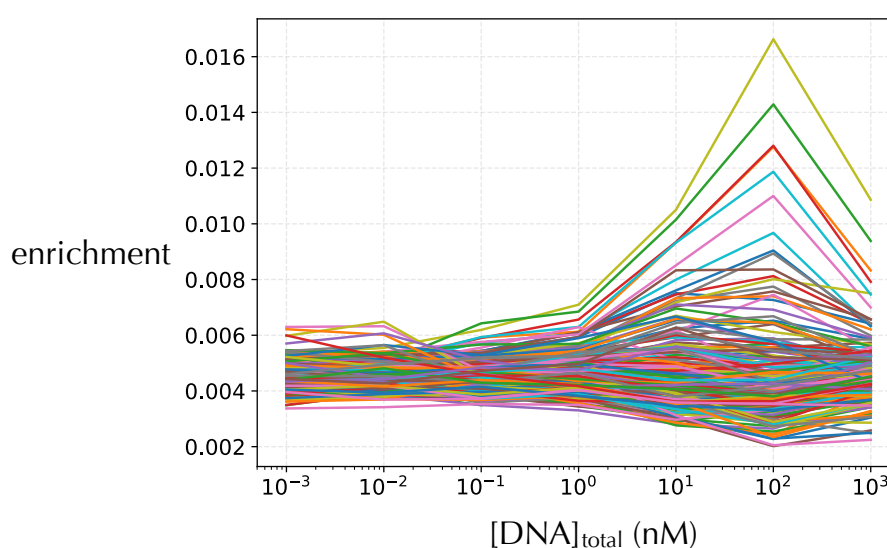


Figure 4.2: **218 binding curves are obtained in a single high-throughput calibration assay.** The enrichments of 218 DNA ligands have been measured for 7 selection experiments with DNA concentrations ranging from 10^{-3} to 10^3 nM. We can identify 3 regimes of DNA concentration: the first regime with $[DNA] < 10^2$ nM, the second regime with $[DNA] \simeq 10^2$ nM and the third regime with $[DNA] > 10^2$ nM. The dynamic range of the enrichment values it the highest in the second regime.

1147 **Conclusion:**

- 1148 ● **The high-throughput calibration assay allowed us to measure 218 binding**
 1149 **curves at the same time.**
- 1150 ● **There is an optimal DNA concentration that maximizes the discrimina-**
 1151 **tion of the DNA ligands during selection.**

1152 **4.3.1.2 The comparison of the high-throughput calibration and ELISA data**
 1153 **shows a linear relationship**

1154 To verify that the enrichments we measured are linked to the affinity of the DNA ligands
 1155 for the antibody, we compared the high-throughput calibration data (for $[DNA]=10^2nM$)
 1156 with ELISA data for 3 DNA ligands. We compared the binding energies with the logarithm
 1157 of the enrichments in Fig. 4.3. We observe a linear relationship between the calibration
 1158 experiment and ELISA data (the lowest binding energy correspond to the most enriched
 1159 DNA ligand).

1160

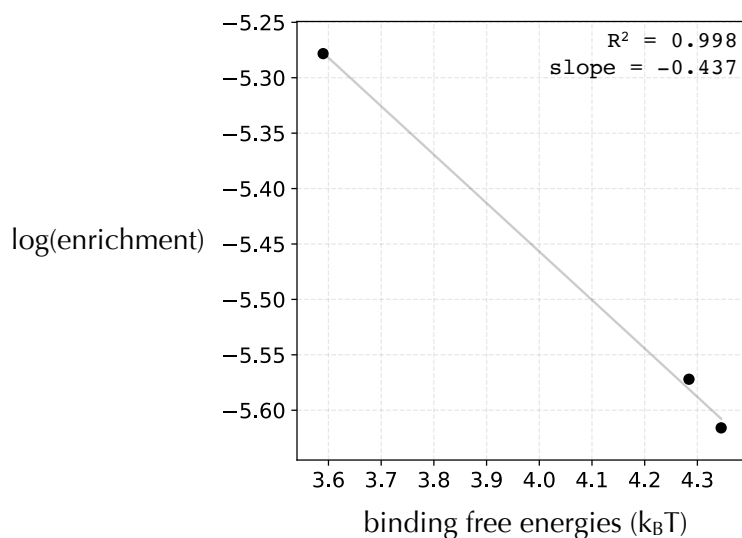


Figure 4.3: **Comparison of calibration and ELISA data for 3 DNA ligands.** We observe a linear relationship between the logarithm of the enrichments measured with calibration and the binding energies indirectly measured with ELISA. The binding free energies are estimated from the experimental dissociation constants by the equation: $\Delta G_i = \ln K_d^i$ with ΔG_i the binding free energy of the complex DNA_i -antibody expressed in units of $k_B T$ and K_d^i the dissociation constant of the complex DNA_i -antibody measured by ELISA. The sequences of the 3 DNA ligands have been picked randomly.

1161 **Conclusion:**

- 1162 • **The ranking of the selected DNA ligands are related to their affinity for**
 1163 **the antibody.**
- 1164 • **There is a linear relationship between the log of the enrichments and the**
 1165 **binding energies for the 3 datapoints we tested (with $[DNA]=10^2nM$).**

1166

1167 **Discussion:**

- 1168 • **The correlation we observe for the 3 datapoints is reinforced by the fact that the 2**
 1169 **experimental setups we compared are very different from each other: the antibody**

1170 is immobilized and its concentration is fixed during the high-throughput calibration
1171 while the antibody is free and its concentration varies during ELISA.

- 1172 • We can predict a linear relationship between the log of the enrichments and the
1173 binding energies for the ligands that are between the two extreme data-points that
1174 we picked for the ELISA.
- 1175 • In order to verify the robustness of the linearity, we could compare the binding
1176 energies and enrichments of more DNA ligands.
- 1177 • High-throughput selection experiments are commonly used to study binding proper-
1178 ties of proteins based on enrichment data without systematically measuring affinities
1179 [Fowler et al., 2010] [Jolma et al., 2013].
- 1180 • Linearity assumption is validated under the condition that the dynamic range of the
1181 studied K_d does not exceed 10 [Kowalsky and Whitehead, 2016]. This condition is
1182 rarely verified in studies working with selection experiments, and thus the building
1183 of binding curves is necessary to accurately measure affinities.

1184 4.3.1.3 A combined use of SELEX and ELISA can provide access to affinity 1185 distributions under linear conditions

1186 We used the enrichment values (for $[DNA]=10^2nM$) to deduce the binding energies of
1187 the 218 DNA ligands based on the linear relationship we observed for the 3 tested DNA
1188 ligands in the previous section. The distribution of the binding energies is presented in
1189 Fig. 4.4. The 2 dashed lines correspond to the binding energies of the tested DNA ligands
1190 (we represented the 2 most different binding energies, the ligands located between these
1191 values represent 52% of the DNA library). The binding energies between these limits have
1192 been interpolated and thus should be more trusted than the binding energies outside these
1193 limits that have been extrapolated. We need more measurements of binding energies to
1194 verify if the extrapolated values are correct.

1195

1196 **Conclusion:**

- 1197 • **Under linear condition, we can deduce binding energies by combining**
1198 **high-throughput calibration and ELISA data.**

1199

1200 **Discussion:**

- 1201 • The assumption of a linear relationship between the calibration experiment and the
1202 ELISA data has limitations.
- 1203 • The long-term goal of our approach is to compare antibodies with different binding
1204 profiles, and thus K_d that can be very different for the same ligands. Consequently,

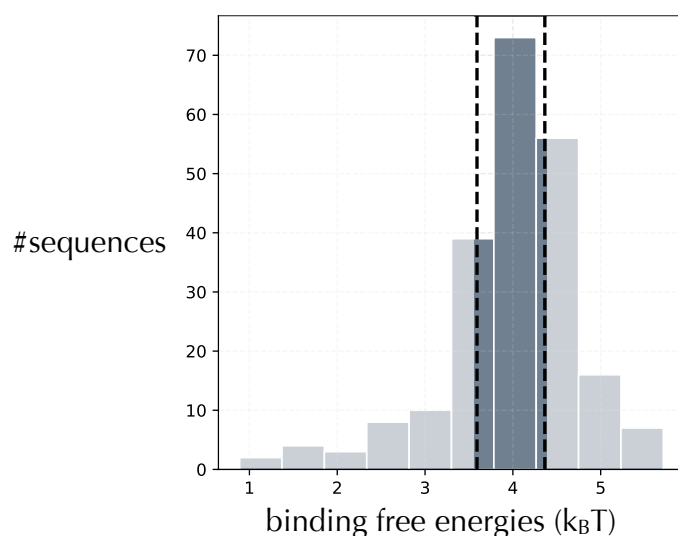


Figure 4.4: **Distribution of the binding energies of the DNA library deduced from the linear relationship between calibration and ELISA data.** The vertical dashed lines correspond to binding energies of the 2 (most different) datapoints we tested. The linear relationship between calibration and ELISA data allowed us to interpolate (deduce energies between the tested datapoints) and extrapolate (deduce energies outside the tested datapoints) the 218 binding energies using their enrichment values. We have a higher confidence in the binding energy values between than outside the tested datapoints because we don't know what is the dynamic range of the binding energies that are linearly related to the enrichments.

1205 we need to find a new approach to measure dissociation constants, that does not
 1206 depend on a linear relationship between the enrichments and binding energies.

- 1207 • We need to understand how the enrichments we measure experimentally are linked
 1208 to the K_d . This is the purpose of the next section.

1209 4.3.2 A simple binding model based on the specific binding of the DNA 1210 ligands to the antibody does not describe the experimental data

1211 4.3.2.1 Why do we need to identify the factors that link the experimental 1212 data to the KDs we want to infer?

1213 1- We identified 3 regimes of selection in the experimental data (Fig 4.1). This observation
 1214 suggests a non-trivial relationship between the enrichments and the K_d that depends on
 1215 the concentrations of reactants we use during calibration.

1216

1217 2- Without further experimental verification, we do not know if there is a linear relation-
 1218 ship between all the binding energies and their corresponding enrichments. Consequently,
 1219 we need to identify the model that links both measurements.

1220

1221 3- The studies presented in the beginning of this chapter reported the presence of exper-
 1222 imental noise in their data. Consequently, they had to adapt their selection model to
 1223 accurately measure the dissociation constants: (1) during yeast display, the authors took
 1224 into account the varying number of antibodies expressed on the surface of the cells as well
 1225 as the auto-fluorescence of the cells during sorting in their binding model [Adams et al.,
 1226 2016] (2) The collection of formed complexes during RBNS experiment was accompanied
 1227 with the unwanted capture of free reagents, so the authors included this non-specific re-
 1228 covery in their binding model as well [McGeary et al., 2019].

1229

1230 Multiple factors can be considered in the link between experimental data and K_d we
 1231 want to measure. To identify these factors, we can simulate the calibration experiment
 1232 with several selection models and identify the one that could possibly explain our data.

1233 4.3.2.2 Presentation of the selection model & identification of the factors that 1234 link our data to the affinities

1235 The output of the calibration experiment are enrichments² of DNA ligand A^i ($i \in \{1, \dots, N\}$),
 1236 obtained for different DNA concentrations. The enrichment of a ligand A_i can be written
 1237 as following:

1238

$$1239 \text{ enr}_{A^i} = \frac{P_{sel,A^i}}{\sum_{j=1}^N P_{sel,A^j}} \quad (4.3)$$

1239 where P_{sel,A^i} is the probability of A_i to be selected (enr_{A^i} normalize P_{sel,A^i} values so
 1240 their sum equal 1, as we did for the enrichment values measured experimentally). We can
 1241 formulate different selection models to link the probability of a A_i to be selected to its
 1242 affinity for the protein target T (the concentrations are measured at equilibrium):

1243

1244 1- Selection model 1 (specific binding):

1245 We make the assumption that the only interaction that happens during selection is the
 1246 binding of the DNA library to the target. In this scenario, A_i can be in 2 states: either
 1247 it is bound to the target ($A^i T$) or it is free in solution (A^i). Its probability of selection
 1248 P_{sel,A^i} is written as following:

1249

$$1249 P_{sel,A^i} = \frac{[A^i T]_{eq}}{[A^i T]_{eq} + [A^i]_{eq}} = \frac{1}{1 + \frac{K_{A^i T}}{[T]_{eq}}} = \frac{1}{1 + \frac{K_{A^i T}}{[T]_{tot} - \sum_{j=1}^N [A^j T]_{eq}}} \quad (4.4)$$

²We experimentally measure the enrichment of each DNA ligand as the ratio between its number of copies after and before the selection. We systematically normalized the enrichment values so their sum equal 1.

1250 with $K_{A^i T} = \frac{[A^i T]_{eq}[T]_{eq}}{[A^i]_{eq}[T]_{eq}}$, the affinity of A_i for T . Note that $[T]_{eq}$ depends on the con-
1251 centrations and the affinities of the other DNA ligands present in the mix during selection.
1252 (Using this selection model and simulations, we studied the impact of the calibration pro-
1253 tocol on the enrichment values by either varying the concentration of the DNA library, or
1254 by varying the concentration of the target in supp. 6.3.7). Note that if the total concen-
1255 tration of target $[T]_{tot}$ is significantly larger than the dissociation constant of the complex
1256 $A^i T$, the probability of selection of the ligand A_i tends to 1 independently of its affinity
1257 for the target (we are in a titration regime). Consequently, every calibration assay must
1258 be performed with $[T]_{tot}$ small enough to discriminate the ligands [Jarmoskaite et al., 2020].

1259

1260 We simulated a high-throughput calibration experiment using this selection model (we
1261 fixed the concentration of the target and varied the concentration of the DNA ligands).
1262 The binding curves are presented in Fig. 4.5 A.

1263

1264 The qualitative comparison of the simulated data with the experimental data (Fig.
1265 4.2) shows significative differences in the regimes of selection. In the simulated data, the
1266 enrichments are independent of the DNA concentration and different from each other. We
1267 observe a single regime of selection. In the experimental data, the enrichments depend on
1268 the DNA concentration and we can identify 3 different regimes of DNA concentrations.

1269

1270 **Conclusion:**

- 1271 • **A selection model based solely on the specific interaction of the DNA**
1272 **library to the antibody does not describe the experimental data.**

1273

1274 **Discussion:**

- 1275 • This selection model potentially describes the experimental enrichments correspond-
1276 ing to the intermediate DNA concentration (10^2 nM) where the enrichments are the
1277 most different from each other (in Fig. 4.2)
- 1278 • The presence of 2 regimes of DNA concentrations in the experimental data that are
1279 not described with this model suggests that there are additional factors that must
1280 be taken into account in the probability of selection of the DNA ligands.
- 1281 • We can formulate hypotheses regarding the nature of these additional factors by
1282 looking at the results of control experiments:
 - 1283 – The non-specific binding of DNA molecules to the plate: we systematically
1284 performed control experiments where we replicated all the steps of the selection,
1285 but no antibody was present in the plate. In every control experiments, we
1286 managed to collect DNA molecules at the end of the selection, suggesting that
1287 the DNA molecules interacted with the plastic of the plate (Supp. 6.3.3).

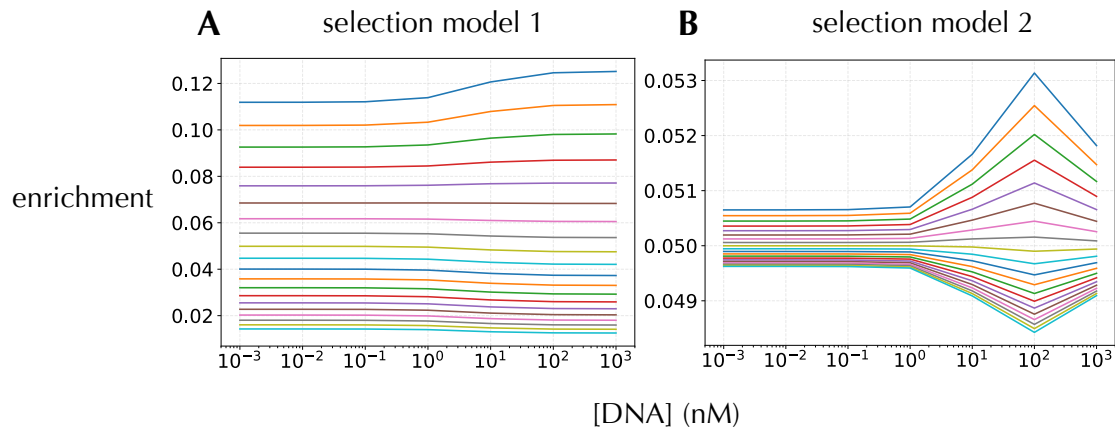


Figure 4.5: **The experimental data are not exclusively described by a simple selection model based on the specific binding of the DNA library to the antibody.** We chose 20 binding energy values ranging from 3.5 to $5.8 k_B T$. We set the antibody concentration to 10nM and tested DNA concentrations ranging from 10^{-3} to 10^3nM . A. The simulated data have been generated with a selection model based on the the specific binding of the DNA ligands to the antibody. The simulated enrichment values are nearly independent of the DNA concentrations contrary to what we observe in experimental data (Fig 4.2) . B. The simulated data have been generated with a selection model based on the the specific binding of the DNA ligands to the antibody, the non-specific binding of the DNA ligands to the plastic of the plate and the unspecific recovery of free DNA molecules. The affinity of the DNA ligand A for the plastic P K_{AP} is set to 20nM and the concentration of the plastic binding sites P $[P]_{tot}$ is set to 200nM . The recovered free DNA molecule represents 1% of the total DNA concentration. The relationships between the enrichment values and the DNA concentrations are qualitatively similar for the simulated and experimental data (Fig. 4.2).

1288 – The recovery of free DNA molecules due to insufficient washing: at the end of
 1289 the binding reaction, we keep the DNA in complex with the target and remove
 1290 the free DNA molecules with multiple washing steps (that dilute the free DNA
 1291 molecules). However, for selections with high concentrations of DNA, the wash-
 1292 ings may become insufficient, leading to a concentration of recovered free DNA
 1293 molecule that is not negligible anymore and that must be taken into account in
 1294 the selection model. This hypothesis is supported by the control experiments
 1295 where we observed proportional increase of recovered free DNA molecule as we
 1296 were increasing the concentration of DNA during the calibration experiment.
 1297

1298 **2- Selection model 2 (specific binding + non-specific binding + free DNA**
 1299 **molecules recovery):**

1300

1301 In this selection model, we considered 2 additional ways for the ligand A^i to pass the
 1302 selection: (1) it binds to the plastic of the plate (we can make the hypothesis that DNA
 1303 ligands interact with the plastic independently of the sequence in their loop, and thus they
 1304 all share the same affinity K_{AP} for the plastic of the plate) and (2) it gets through the
 1305 washing steps.

1306

1307 First, we only consider the first scenario where each DNA ligand can be in 3 different
 1308 states: free in solution (A^i), bound to the antibody (A^iT) and bound to the plastic (A^iP).
 1309 The probability of A^i to be selected is written as:

1310

$$P_{sel,A^i}^1 = \frac{[A^iT]_{eq} + [A^iP]_{eq}}{[A^iT]_{eq} + [A^iP]_{eq} + [A^i]_{eq}} \quad (4.5)$$

1311 Then we consider the next scenario where each DNA ligand can be in 3 states (A^i , A^iP ,
 1312 A^iT) and can get through the washings. The probability P_{sel,A^i}^2 of A^i to be selected is
 1313 written as:

$$P_{sel,A^i}^2 = (1 - \alpha)P_{sel,A^i}^1 + \alpha \quad (4.6)$$

1314 with α the fraction the free DNA molecules that gets through the washings (if the wash-
 1315 ings remove 100% of the free DNA molecules, $\alpha = 0$ and $P_{sel,A^i}^2 = P_{sel,A^i}^1$).

1316

1317 We simulated a calibration experiment using this second selection model. The binding
 1318 curves are presented in Fig. 4.5 B. The simulated data are qualitatively more similar to
 1319 the experimental data with this new selection model. The simulated data show 3 regimes
 1320 of selection that are also present in the experimental data. The first and third regimes cor-
 1321 respond to DNA concentrations where the enrichments are the most similar. The second
 1322 regime correspond to DNA concentrations where the enrichments are the most different.

1323

1324 **Conclusion:**

- 1325 • **The addition of external factors in the selection model (as the non-specific**
 1326 **binding of the DNA ligand to the plastic and the unwanted recovery of**
 1327 **free DNA molecules) generate artificial data that are qualitatively more**
 1328 **similar to the experimental ones.**
- 1329 • **A realistic selection model must contain terms describing experimental**
 1330 **bias in addition to the specific binding of the ligands to the target (this**
 1331 **observation has also been made in the calibration studies presented pre-**
 1332 **viously).**

1333

1334 **Discussion:**

- 1335 • Here we make the assumption that we measure concentrations of complexes at equi-
1336 librium. However, in order to remove the free DNA molecules, we must perform
1337 multiple washing steps that are susceptible to disrupt the equilibrium state of the
1338 reaction. Therefore, during calibration experiment, we perform the washings as
1339 quickly as possible to minimize the disruption of equilibrium.

- 1340 • We identified at least 2 additional factors that could explain the presence of 3 regimes
1341 of selection in the calibration data, based on the analysis of control experiments
1342 (however, we must be careful with the use of additional parameters as they will
1343 always make easier the fit of the model to experimental data, even if the model is
1344 not appropriate).

- 1345 • We can perform extra control experiments to validate the presence of these factors
1346 in our selection model (We can replicate the calibration experiment with different
1347 washing protocols, stronger and weaker, and see if we can predict the change in the
1348 experimental data).

- 1349 • For both selection models, we verified if the enrichments and binding energies were
1350 linearly related under an optimal DNA concentration (Fig. 6.7). The results show
1351 that the most sophisticated selection model (with non-specific binding and non-
1352 specific recovery terms) generates enrichments that can be linearly linked to binding
1353 energies if the dynamic range of these binding energies is small enough. Conse-
1354 quently, the simulated data confirms that a linear relationship can be approximated
1355 between the enrichments and binding energies measured experimentally under par-
1356 ticular conditions.

- 1357 • An additional factor that we did not include in the selection model but that is
1358 not negligible in the experimental data is the sampling of the DNA library (Supp.
1359 6.3.3.3). The sampling noise is systematically observed in our deep-sequencing data
1360 (in particular when we compare the number of copies after and before selection). In
1361 the next section, we study the impact of the sampling noise on the resolution of the
1362 selection experiment.

1363 **4.3.3 There is a tradeoff between the diversity of the library and the**
1364 **resolution of the experiment**1365 **4.3.3.1 Experimental data suggest that the design of the DNA ligand library**
1366 **impacts the resolution of the selection experiment**

1367 In this section we test the limits of our calibration experiment by measuring the resolu-
1368 tion ³ of the data for different library designs. We did 2 separate calibration experiments.

³We define the resolution as the minimal value we can measure with little error. A high resolution allows the accurate measurement of small values. In our case, the values are enrichments. Enrichments

1369 In the first experiment, the library is composed of 218 DNA ligands while in the second
 1370 experiment, the library is composed of 32,768 DNA ligands. We adapted the number of
 1371 sequencing reads to each assay in order to expose them to the same level of sampling dur-
 1372 ing deep-sequencing. However, we used the same number of antibodies for both libraries
 1373 during selection.

1374

1375 For each calibration assay, we qualitatively measured its resolution by comparing the
 1376 enrichment values between near-replicate experiments (we compared enrichments obtained
 1377 with $[\text{DNA}]=10^2$ and $[\text{DNA}]=10^3$ nM). The results are presented in Fig 4.6. The DNA
 1378 library with a high diversity has the lowest resolution (only the highest enrichments are
 1379 reproducible from one experiment to another). The DNA library with a low diversity
 1380 shows a higher resolution with enrichments that are more reproducible independently of
 1381 their value.

1382

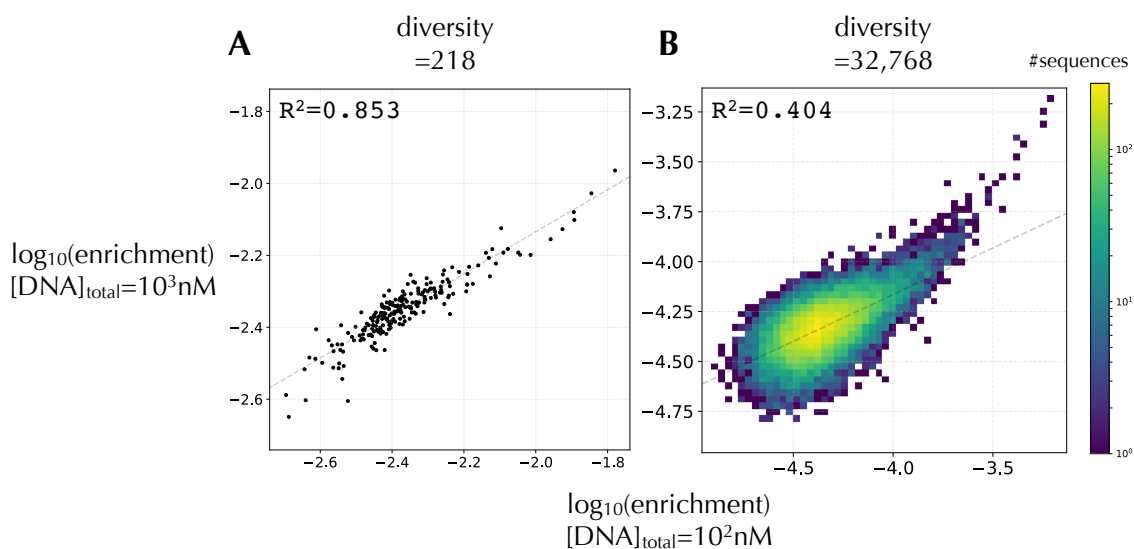


Figure 4.6: **Experimental data show that the resolution of measured enrichments is impacted by the diversity of the DNA library.** We compared the reproducibility of enrichments measured for 2 different DNA libraries. We assessed the reproducibility of the measurements by comparing selection experiments that have a hundred fold difference in DNA concentrations ($[\text{DNA}]=10^2$ and $[\text{DNA}]=10^4$ nM. A. The first library is composed of 218 DNA ligands. B. The second library is composed 32,768 DNA ligands. The enrichment values are more reproducible for the small DNA library than for the big DNA library (The R^2 score is higher for the small than for the big library).

1383 **Conclusion:**1384 **• The diversity of the DNA library impacts the resolution of the experi-**

measured with little error allow us to infer K_d with a high accuracy).

1385 **ment.**

1386 • **The DNA library with the highest diversity has the lowest resolution.**

1387

1388 **Discussion:**

1389 • The reproducibility of the enrichments is limited by the sampling noise that can be
1390 generated at different steps during the selection experiment (principally during the
1391 binding of the DNA ligands to the antibody or during the sequencing step).

1392 • We adapted the number of reads to the diversity of the DNA library so a similar
1393 number of reads was attributed to each DNA ligand (and thus the sampling due to
1394 the deep-sequencing is the same for both assays).

1395 • Consequently, the difference in resolution we observe is due to to the sampling during
1396 the binding step.

1397 • For the same number of selected DNA molecules, the average copy number of DNA
1398 ligands in the high diversity library is smaller and thus more sensitive to variability
1399 during sampling. This is why the enrichments below a threshold value are not
1400 reproducible for this library.

1401 • So far, we studied the impact of the diversity of the library on the resolution of
1402 the experiment. It would be more accurate to consider the distribution of affinities
1403 of the DNA library because it would also impact the resolution of the experiment.
1404 However, during the design of the DNA library, we only have access to its diversity,
1405 this is why we focused our study on this parameter.

1406 **4.3.3.2 Simulated data suggest a tradeoff between the diversity of the ligand** 1407 **library and the resolution of the experiment**

1408 We simulated multiple selection experiments where we fixed the total concentration of the
1409 DNA library and varied its diversity (we used the selection model described in eq. 6.1).
1410 For each diversity value, we simulated different numbers of antibodies because we don't
1411 have access to the exact experimental number (we describe how we simulated different
1412 numbers of antibodies in Supp. 6.3.6). For each experimental condition (diversity and
1413 number of antibodies), we identified the smallest enrichment value we could measure with
1414 less than 10 percents of error and calculated a resolution score. We set the resolution
1415 score as the the ratio between the dissociation constant corresponding to the smallest en-
1416 richment with less than 10% of error and the highest dissociation constant of the library
1417 (as a reminder, the highest dissociation score corresponds to the smallest affinity in the
1418 library) (Fig 4.7).

1419

1420 The simulated data show that the resolution of the experiment is maximal when the
 1421 number of DNA molecules collected at the end of the selection is significantly higher than
 1422 the diversity of the DNA library.

1423

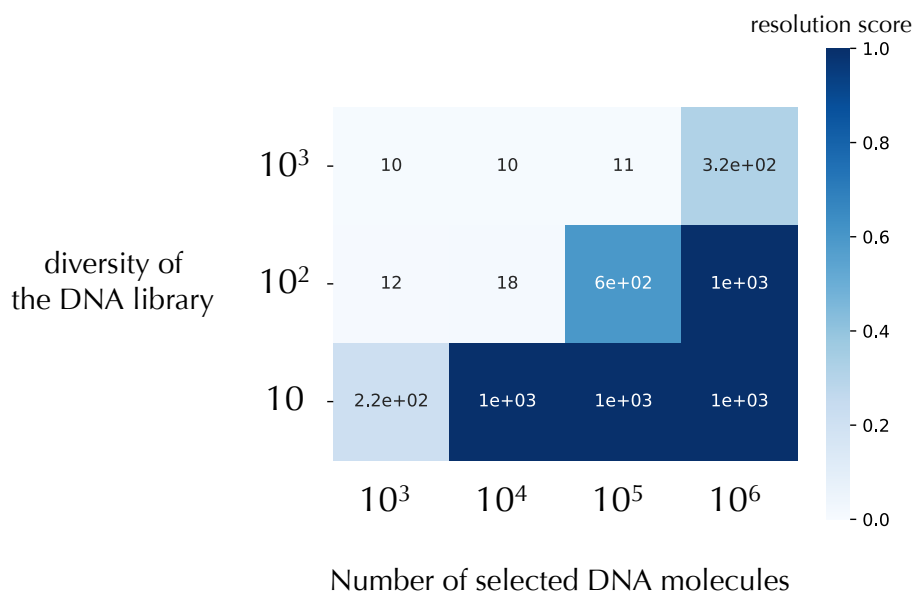


Figure 4.7: **Simulated data show that the resolution of selection experiment is negatively correlated with the diversity of the library, for a fixed number of selected DNA molecules.** We identified the smallest enrichment value we can measure with less than 10% of error for different DNA libraries (with diversity ranging from 10 to 10^3 , and K_d values ranging from 10 to 1000nM) and for different numbers of selected DNA molecules (ranging from 10^3 to 10^6). We set the resolution score as the ratio between the K_d of the smallest enrichment (with less than 10% of error) and the highest K_d in the library. We represented in the matrix the K_d values (in nM) corresponding to these enrichments. A small number of selected DNA molecules results in a high sampling noise that negatively impact the resolution of the experiment.

1424

Conclusion:

1425

- The experimental and theoretical data show a trade-off between the diversity of the DNA library and the resolution of the calibration experiment due to sampling noise.

1426

1427

1428

1429

Discussion:

1430

- In order to increase the number of selected DNA molecules, we would need to increase the number of antibodies per well and take the risk of being in a titration regime where the probabilities of selection do not depend on the affinities anymore (Section 4.3.2.2). There is a maximal diversity we cannot exceed without impacting the resolution of the experiment.

1431

1432

1433

1434

- 1435 • Thus, the only parameter we can optimize during the calibration experiment is the
1436 diversity of the DNA library.
1437

1438 4.4 Summary

- 1439 • We designed a protein-ligand system (composed of a synthetic antibody and a li-
1440 brary of 218 single-stranded DNA ligands) to set-up the high-throughput calibration
1441 experiment.
- 1442 • The high-throughput calibration assay allowed us to build 218 binding curves using
1443 7 wells and few days of experimental work while the regular ELISA approach would
1444 have required 1744 wells corresponding to 27 plates and as many days of experimental
1445 work.
- 1446 • For 3 DNA ligands, we compared their enrichments measured with the high-throughput
1447 calibration experiment, with their binding energies measured with an independent
1448 assay (ELISA). We observed a linear relationship between calibration and the ELISA
1449 data.
- 1450 • The measurement of affinities with no selection model is possible under the condition
1451 that the calibration data are linearly linked to independent measurement of binding
1452 energies (this is generally the case when the dynamic range of K_d is small enough).
1453 If this condition is not verified, we must define a selection model to infer affinity.
- 1454 • The selection model must take into accounts additional factors in order to link the
1455 data to the affinities. Our calibration experiment probably contains at least 2 sources
1456 of noise based on the results of control experiments and the qualitative comparison
1457 of simulated data (generated with the second selection model) with experimental
1458 ones.
- 1459 • Experimental and simulated data show a tradeoff between the diversity of the DNA
1460 library and the resolution of the experiment.
- 1461 • With more time, we could investigate the limits of the linearity observed between
1462 calibration and ELISA data by comparing the data of more than 3 DNA ligands.
- 1463 • In parallel, we could infer the 218 dissociation constants using different selection
1464 models and validate the results with independent measurements of affinities (ob-
1465 tained with ELISA).

1466 4.5 Discussion

1467 The main goal of the calibration experiment is to combine high-throughput selection assay
1468 with quantitative measurement of affinity. The output of the assay is a distribution of

1469 affinities between a protein of interest and a library of DNA ligands. This distribution
1470 that we call “affinity profile” can be used as a proxy of the specificity of the protein. The
1471 long-term goal of the calibration experiment is to measure the binding profile of several
1472 Germline and Limited antibodies and use the parameters (average and standard deviation)
1473 of the affinity distributions to quantitatively compare their specificity.

1475 Chapter 5

1476 Conclusion, discussion and 1477 perspectives

1478 5.0.1 Conclusion

1479 Our work was motivated by the following question: Can we quantitatively measure the
1480 specificity of antibodies ? We decomposed this general question into 3 sub-questions:

- 1481 (1) Do different synthetic antibodies have different average affinities for DNA ligands ?
- 1482 (2) Do different synthetic antibodies select the same DNA sequences, with the same di-
1483 versity ?
- 1484 (3) Can we propose a high-throughput selection assay to measure multiple dissociation
1485 constants in a single experiment ?

1486

1487 To answer the first and second questions, we estimated the average affinity of antibod-
1488 ies with different framework and CDR3 regions for 2 DNA libraries, and we compared the
1489 diversity of the DNA sequences that were selected by these antibodies. The following con-
1490 clusions have been made based on the study of few synthetic antibodies. We observed that
1491 antibodies with different frameworks and CDR3 regions have different average affinities
1492 for DNA and they select DNA sequences with different levels of diversity. More precisely,
1493 we observed that the frameworks and CDR3 regions have different effects on the selection
1494 of the DNA ligands:

1495

1496 **Impact of the frameworks on selection:**

1497 First, we observed that the frameworks control the dynamic range of the affinities that
1498 the CDR3 can reach. More precisely, the Germline framework allows a higher maximal
1499 affinity than the Limited framework. Second, we observed an impact of the framework
1500 on the role of the CDR3 region in the discrimination of the DNA sequences: different
1501 antibodies with the Germline framework selected the same DNA sequences while different
1502 antibodies with the Limited framework selected different DNA sequences.

1503

1504 Impact of the CDR3 region on selection:

1505 First, we observed that antibodies with different CDR3 regions have different affinities for
1506 DNA ligands (within the dynamic range fixed by the framework). In particular, the CDR3
1507 that have been selected for their high affinity for 1 particular DNA ligand have an average
1508 affinity for DNA that is higher in comparison with CDR3 randomly chosen. Second, we
1509 observed that under the same experimental conditions, the CDR3 selected against 1 DNA
1510 ligand can discriminate the DNA library while random CDR3 cannot.

1511

1512 To answer the third question, we set-up a high throughput selection assay to build the
1513 binding curve of 218 DNA ligands in a single experiment. With more time, these binding
1514 curves could be used to infer the dissociation constants of the 218 DNA ligands. The
1515 high-throughput calibration assay is a promising approach to obtain quickly and simply a
1516 more quantitative specificity profile.

1517 5.0.2 Discussion and perspectives**1518 5.0.2.1 Limits of our approach**

1519 Our experimental and theoretical results suggest that there is a tradeoff between the di-
1520 versity of the DNA library and the resolution of the selection experiment. Indeed, we
1521 couldn't accurately measure the enrichment values of the 32,768 DNA ligands from the
1522 high diversity library (the smallest enrichment values were dominated by sampling noise).
1523 Consequently, we cannot rigorously measure the dissociation constant of the ligands with
1524 the smallest enrichment values. These observations set the limit of our system: with the
1525 actual experimental setup, we cannot measure the dissociation constants of 10^5 DNA lig-
1526 ands at the same time.

1527

1528 To address this tradeoff, we can use the same experimental setup and limit our mea-
1529 surement to few hundred K_d (We can generate a new DNA library where we randomize
1530 4 consecutive nucleotides instead of 7, based on the observation that 4 nucleotides motifs
1531 are selected by antibodies). We can also change the experimental setup and perform the
1532 selection step with free DNA ligands and antibodies in solution (the immobilization of
1533 antibodies on a solid surface may prevent the optimal interaction between DNA ligands
1534 and antibodies).

1535

1536 5.0.2.2 Next experiments

1537 Our preliminary results suggests that a DNA library can be used to measure different
1538 binding behaviors between different synthetic antibodies. It motivates the continuation of
1539 the project using the same system.

1540

1541 **Short term experiments:**

1542 (1) With more time, we could continue our analysis of the 218 binding curves to infer
1543 the corresponding dissociation constants. Once the pipeline of the inference is set-up, we
1544 could generate the same binding curves for the different Germline and Limited antibodies
1545 we studied in chapters 2 and 3. Then we could quantitatively compare their specificity by
1546 measuring the parameters (mean and standard deviation) of their distributions of affinities
1547 for the same library of DNA ligands.

1548

1549 (2) So far we only tested the additive model to understand the link between the sequence
1550 of the selected DNA ligands and their affinity for the synthetic antibodies. With more
1551 time, we could test more sophisticated models that take into account the correlations be-
1552 tween the positions in the selected DNA sequences.

1553

1554 **Long term experiments:**

1555 We could perform *in vitro* affinity maturation of Germline antibodies (with multiple cycles
1556 of mutation, selection and amplification) in order to increase their affinity for a particular
1557 DNA ligand. We could then investigate the relationship between the evolving affinity of
1558 these synthetic antibodies and their specificity for DNA ligands.

1559 **5.0.2.3 Comparison of our results with previous works**

1560 The few Germline and Limited antibodies we tested in this project came from larger anti-
1561 body libraries that have been studied as part of previous work [Boyer et al., 2016] [Schulz
1562 et al., 2021]. Germline and Limited antibody libraries have been selected against a single
1563 DNA ligand, separately (Fig. 5.1). The resulting distribution of selectivities (i.e. enrich-
1564 ments, not normalized) for each library was used as a proxy of their selection potential
1565 (i.e. their ability to yield a high affinity for a new target).

1566

1567 The authors showed that Germline antibodies have a higher selection potential than
1568 Limited antibodies: most of the Germline antibodies have a low affinity for the DNA lig-
1569 and that has been tested, but few antibodies have a very high affinity for this DNA ligand
1570 (their distribution of selectivities are characterized by a low average selectivity and high
1571 standard deviation). On the other side, most Limited antibodies have a higher affinity
1572 for the same DNA ligand, but no Limited antibody has a particularly high affinity for
1573 the DNA ligand (contrary to what we observe for Germline library) (their distribution of
1574 selectivities are characterized by a high average selectivity and low standard deviation).

1575

1576 We can equivalently characterize the specificity profile of these antibodies as the pa-
1577 rameters of the distribution of their affinity for the DNA library (the affinities can be
1578 measured with the calibration experiment we present in chapter 4). Based on our obser-
1579 vations, we can predict that Germline antibodies have a higher average affinity for DNA,

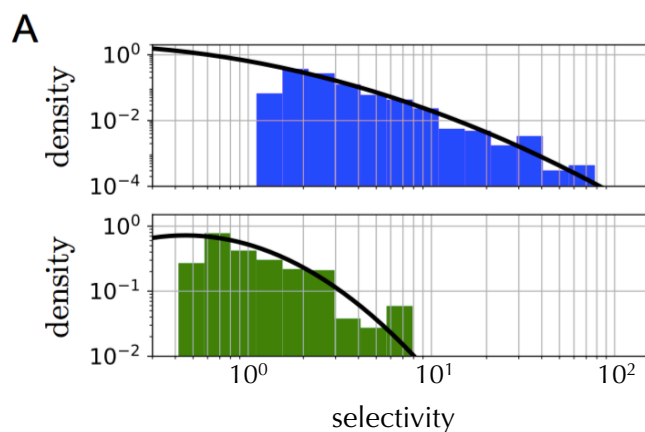


Figure 5.1: **Distributions of selectivities of Germline (in blue) and Limited (in green) antibody libraries.** Both libraries of antibodies have been selected against a DNA ligand by phage Display. The antibodies (before and after their selection) have been deep-sequenced and their selectivity have been calculated as the ratio between their copy numbers after and before the selection (selectivities are equivalent to the non-normalized enrichments measured in our project). Taken from [Schulz et al., 2021].

1580 with a smaller dynamic range of affinities, while Limited antibodies have a smaller average
 1581 affinity for DNA, with a higher dynamic range.

1582

1583 The parameters describing the specificity profiles are potentially opposite to the param-
 1584 eters describing the selection potential, when comparing Germline and Limited antibodies.
 1585 This observation suggests that antibodies with the highest selection potential may have
 1586 the smallest specificity and antibodies with the smallest selection potential may have the
 1587 highest specificity (this hypothesis must yet be validated with the quantitative analysis of
 1588 more Germline and Limited antibodies).

1589

1590 These predictions are surprising as Germline antibodies analyzed in Fig. 1.4 seem to
 1591 be more specific than Limited antibodies for 2 DNA ligands. However, we must keep in
 1592 mind that Phage Display and SELEX are not symmetrical experiments (During Phage
 1593 Display, antibodies are expressed on the surface of phages and DNA ligands have a small
 1594 stem while during SELEX, antibodies are purified and DNA ligands have a long stem).

1595

1596 Taking a step back, we used a minimalist format of antibody in order to simplify the
 1597 study of the relationship between the sequence and the phenotype. The preliminary results
 1598 suggest that mutations in the framework that are potentially far away from the binding
 1599 region may have an impact on the binding specificity of the protein. Our minimalist system
 1600 potentially highlighted the long-range interaction between amino-acids in the sequence (the
 1601 mutations by which Limited differs from Germline framework are 4 to 76 amino-acids away
 1602 from the CDR3 region) and their impact on the specificity of antibodies.

1604 **Chapter 6**

1605 **Supplementary material**

1606 **6.1 Design**

1607 **6.1.1 Presentation of the synthetic antibodies**

A

framework	amino acids sequence
germline	QLQLQESGPGPLVKPSETLSLTC TV SGGSI SSSSY YWGWIRQ PPGK GLEWIG S I YYSG S TYYNPSLKSRVTIS V DTSKN Q FSL KL SSVTAADTAVY Y CARF DYWGQ Q TLVT
limited	QLQLQESGPGPLVKPSETLSLTC I VSGGSI GTTD HYWGWIRQ SPGK GLEWIG T T YYSG K TYYNPSLKSRVTIS I DTSKN H FSL RL ISVTAADTAVY H CARF DYWGQ Q TLVT

B

name	group	framework	CDR3
GNT2	TOP	germline	RKLH
GNT3	TOP	germline	RTKH
GBT3	TOP	germline	GRAT
GNT4	TOP	germline	RSKH
GR1	Random	germline	KVRR
GR2	Random	germline	GWVI

C

name	group	framework	CDR3
LBT1	TOP	limited	RSCS
LNT3	TOP	limited	ARYK
LNT4	TOP	limited	GRYK
LR2	Random	limited	WLLG
LR3	Random	limited	CTSQ

Figure 6.1: **Amino-acid sequences of the frameworks and CDR3 regions.** Blue amino-acids are different between Germline and Limited framework. Red dots correspond to the amino-acids in the CDR3 regions (CDR3 sequences are detailed in tables B and C).

1608 **6.1.2 Presentation of the DNA library**

1609 We used 2 libraries of 32,768 different DNA sequences as potential ligands to study syn-
 1610 thetic antibodies (Fig. 6.2). The libraries are composed of multiple single stranded DNA
 1611 molecules of 54 or 57 nucleotides. Single stranded DNA molecules have the particularity
 1612 to fold on themselves via the complementarity interaction of the nucleotide A with T and

1613 C with G. We designed the DNA molecules to have a stem loop 3D structure with a stem
1614 region where the nucleotides interact with their complementary pair and form double-
1615 stranded DNA (the 2 libraries have different stem regions), and a loop region where the
1616 nucleotide do not interact with others and form a single stranded DNA chain of 7nt. All
1617 the DNA molecules from the same library share a constant region that corresponds to
1618 the stem and that contains the sequences that are necessary for the analysis of the se-
1619 lection experiment. The diversity of the DNA library is located in the DNA loops that
1620 are composed of 7 random nucleotides. We designed the library with 2 possible pairs of
1621 nucleotides (TA and CG) that can close the loop. Thus the library can reach a diversity
1622 of up to $2 * 4^7 = 32,768$ different DNA ligands. We designed these libraries to study
1623 the interaction between the CDR3 region of the synthetic antibodies and the loop of the
1624 DNA ligands. The presence of two libraries allows us to detect a potential impact of the
1625 constant region on the selection of the loops. DNA libraries with different constant regions
1626 are regularly tested when the interaction DNA-protein is studied in the literature, in order
1627 to see if the constant region interferes in the binding [Jolma et al., 2010].

1628

1629 ELISA assay used to measure K_d requires the DNA to be immobilized in each well. To
1630 do so, the DNA is synthesized with a biotin on its 5' end so the molecules get covalently
1631 bound to the streptavidin of the well. In turn, SELEX assay requires the antibody (not
1632 DNA) to be immobilised in each well. Thus, DNA is synthesized with no biotin on its 5'
1633 end.

1634

1635 6.1.3 Calibration experiment

1636 The experimental set-up of our calibration experiment is presented in (Fig. 4.1). While
1637 the authors of the high-throughput studies presented in chapter 4 performed the binding
1638 steps with reactants that are free in solution, we decided to immobilize the antibody on a
1639 solid surface for the 3 following reasons:

1640

1641 1- The long term goal of our approach is to compare the affinity profile of various syn-
1642 thetic antibodies. Thus, we need the experimental conditions to be as constant and robust
1643 as possible from one antibody to another. Saturating a solid surface with the antibody
1644 ensures we always work with the same concentration of antibody independently of the
1645 yield of the purification that can greatly vary from one synthetic antibody to another.

1646 2-Our approach requires the use of one single plate. There is no need for an extra step for
1647 the capture of the formed complexes that are free in the solution. Our approach maximizes
1648 the recovery of the formed complexes.

1649 3-The configuration where the antibody is immobilized by one of its extremities is symmet-
1650 rical to the phage display experiment (used for the selection of the synthetic antibodies)
1651 where one single end of the protein was presented to the DNA ligand.

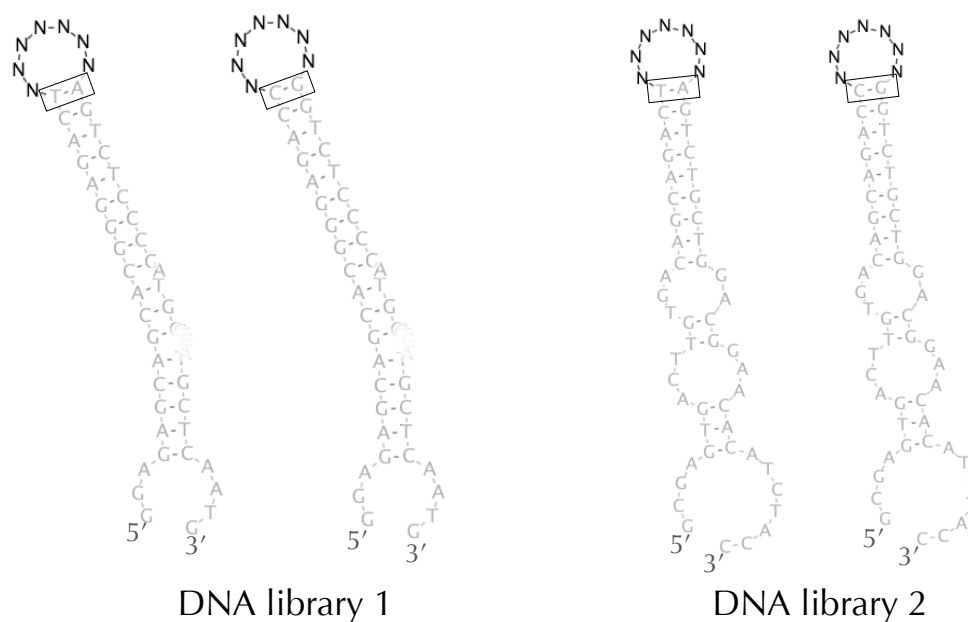


Figure 6.2: **Our ligand system** The ligands are single-stranded DNA molecules with a stem-loop 3D structure. There is a constant region forming the stem and a 7 nucleotides variable region forming the loop. 2 DNA libraries have been designed with different constant regions. The DNA library 1 is presented in figA, and the DNA library 2 is presented in figB.

1652

1653 Working with antibodies immobilized on a solid surface constrains our choice regarding
 1654 which concentration of reactant we fix and which concentration of reactant we vary during
 1655 the binding experiments. In this configuration, we had to vary the concentration of the
 1656 DNA ligand library for the 3 following reasons:

1657

1658 1-If we vary the concentration of the antibodies that are immobilized on the solid surface,
 1659 we vary two parameters at the same time: the concentration of the immobilized antibody,
 1660 and the concentration of the non specific binding sites (If the number of antibodies in the
 1661 plate decreases, the surface of plastic that is not covered increases).

1662 2-Decreasing the concentration of the antibodies on the solid surface is not trivial. The ca-
 1663 pacity of binding of the plates indicated by the manufacturers is generally overestimated.
 1664 It requires systematic tests to evaluate the amount of antibodies that saturates the sur-
 1665 face, every time a new batch of antibody is prepared or a new antibody is studied.

1666 3- Varying the concentration of the antibody requires high amounts of proteins. We are
 1667 often limited in the concentration of proteins we can produce in the lab because of their
 1668 low stability or low yield of production by the cells.

1669

1670 6.2 Protocols

1671 6.2.1 Production of antibodies

1672 Production and purification of the protein:

1673 The gene of the protein is in fusion with a his tag (used for the purification) and a SBP
 1674 tag (used for the immobilization in the SELEX experiment). The gene is located on a
 1675 plasmid with an ampicillin resistance cassette and under the control of the T7 promoter
 (Fig. 6.3).

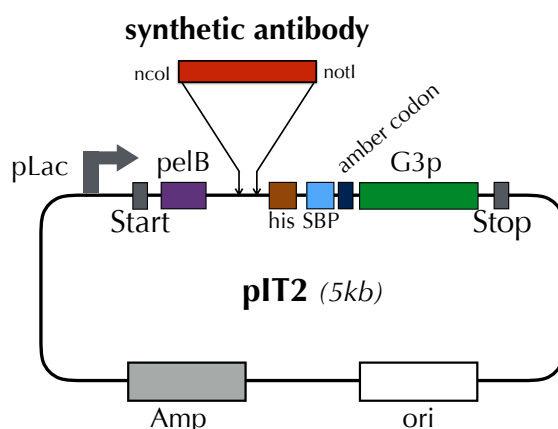


Figure 6.3: Map of the vector that contains the synthetic antibody gene under the control of the T7 promoter.

1676

1677 The plasmid is transformed into an *E. coli* strain (C3019). First, a 5mL liquid culture
 1678 of *E. coli* containing the plasmid is started in 2xYT growth medium with 1% glucose and
 1679 ampicillin at 30°C. The day after, a 100 times dilution of the overnight culture is made into
 1680 a 100mL liquid culture with 0.1% glucose and ampicillin. The OD after dilution should
 1681 average 0.06. The 100mL liquid culture is then grown into a flask at 37°C for 2 hours
 1682 until the OD reaches 0.6. Then IPTG is added (final concentration = 1mM) to turn the
 1683 lac operon on and thus induce the production of the antibody. The induced culture is
 1684 then grown at 30°C for 12 to 16 hours. At the end of the induction process, the 22.8 kDa
 1685 antibodies have been massively produced inside the cell and a non negligible fraction of
 1686 them can be found in the supernatant (their small size allows them to go through the
 1687 cell membrane). The day after, the liquid culture is spinned for 10 minutes at 8000rpm
 1688 in order to collect the supernatant and get rid of the cells. (this procedure is repeated
 1689 twice). From now on the rest of the purification is performed on ice or at 4°C. In order
 1690 to make sure that all the cells and debris are discarded, the supernatant is filtered with
 1691 a 0.2µm membrane. Once the antibodies are released in the supernatant, their hist-tag
 1692 is used to concentrate them in a small volume of PBS buffer. To do so, 1mL of HisPur
 1693 Cobalt Resin is used for the purification of 250mL of supernatant containing the antibod-
 1694 ies. The cobalt resin is transferred to a gravity-flow column and the supernatant is added
 1695 to the column twice. The washing and elution procedures are performed following the

1696 online instructions. In order to get rid of the small unwanted peptides that got collected
1697 with the protein of interest, the resin eluate is then dialysed using the float-a-lyzer kit.
1698 The dialysed solution is aliquoted, snape frozen and stored at -80°C. In order to quantify
1699 and evaluate the purity of the produced antibodies, the sample is run on a protein gel in
1700 denaturing conditions.

1701

1702

1703 **6.2.2 ELISA assay**

1704 The following protocol has been performed using streptavidin-coated plates with different
1705 binding capacities (5-125 pmol/ well). 8 consecutive wells are used to measure the KD
1706 of 1 DNA ligand. Up to 7 different DNA ligands can be studied per 96 well plate (the
1707 last row is reserved to the negative control where no DNA is immobilised in the wells).
1708 The procedure is performed with a multichannel pipett when it is possible. First the wells
1709 are washed 3 times with the washing buffer 1 (PBS Tween 0.1%). Then the biotinylated
1710 DNA is added to each well in a final volume of 100uL of binding buffer/well (BW1X)
1711 (a solution of PBS is added to the control wells). The concentration of added DNA is
1712 at least 5 times higher than the binding capacity of the well. The biotinylated DNA is
1713 incubated in the wells for 1h30 at 20°C or overnight at 4°C with mild shaking. The wells
1714 are then washed once with 200uL of washing buffer 1. Because the antibodies possess a
1715 SBP tag on their C terminal end to be immobilised during the SELEX experiment, all
1716 the streptavidin sites of the wells must be occupied with a biotin molecule during the
1717 ELISA experiment to prevent the unwanted binding of the antibodies to the plate. To
1718 do so, 200uL of PBS-biotin solution (125uM biotin) is added to each well (including the
1719 control wells) and incubated for 30 minutes at 20°C, mild shaking. Then, the wells are
1720 washed 3 times with 200uL of washing buffer 1 and blocked for 30minutes at 20°C at rest
1721 with a blocking solution (0.02%BSA). The blocking solution is then replaced with 100uL
1722 of PBS containing the antibodies at different concentrations. The binding is allowed for
1723 1h at 20C with mild shaking. Once the equilibrium is reached, the wells are washed once
1724 with 200uL of washing buffer 2 (PBST Tween 0.1% & BSA 0.02%) and 100uL of washing
1725 buffer containing the primary antibody anti-polyhistidine and the secondary antibody anti
1726 mouse HRP) are added to the wells and incubated for 1h at 20°C with mild shaking. The
1727 wells are then washed 5 times with 200uL of washing buffer and 100uL of ELISA substrate
1728 is added immediately. After few seconds, 50uL of HCl (1M) is added to each well to block
1729 the reaction and the absorbance of each well (including the control wells) is measured at
1730 450nm using the TECAN platereader.

1731

1732 We constructed the binding curve by plotting the measured absorbance according to
1733 the antibody concentration that was used in each well (We normalized the data so the
1734 absorbance values range from 0 to 1). We fitted the 8 datapoints to the Hill-Langmuir

1735 equation.

$$abs_i = \frac{1}{1 + (K_A/[ab]_i)^n} \quad (6.1)$$

1736 where abs_i (the absorbance measured in the well “i”) and $[ab]_i$ (the concentration of anti-
 1737 body in the well “i”) are experimental measurements and K_A (the antibody concentration
 1738 that saturate half of the DNA molecules) and n (the Hill coefficient) are the parameters
 1739 we infer. Given the absence of competition during binding (we study the binding of anti-
 1740 body with a single DNA ligand) and given that 1 single DNA and 1 single antibody are
 1741 needed to form a complex, K_d equals K_A . We present an example where we fitted the
 1742 experimental datapoints to the Hill equation in Fig. 6.4.

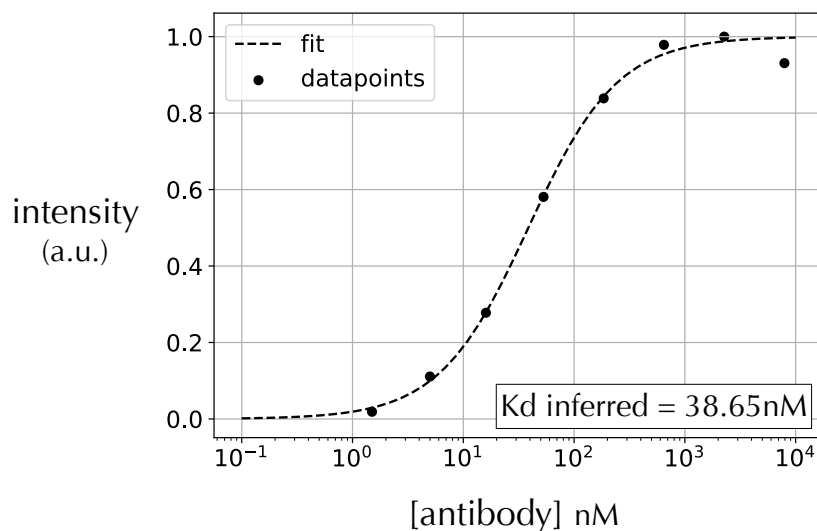


Figure 6.4: **Example of a dissociation constant inference from ELISA data**

1743 6.2.3 SELEX assay

1744 6.2.3.1 On plate

1745 The following protocol has been performed using streptavidin-coated plates with a binding
 1746 capacity of 125pmol/well. First the wells are washed 3 times with 200uL of washing buffer
 1747 2 (PBS Tween 0.1% & BSA 0.02%). Then the antibody is added to each well in a final
 1748 volume of 100uL of PBS (a solution of PBS is added to the control wells). The concentra-
 1749 tion of added antibody are higher than the binding capacity of the well. The antibodies
 1750 are incubated in the wells for 1h30 at 20°C with mild shaking. Then the wells are washed
 1751 3 times with the washing buffer 2 and blocked with a blocking solution (PBS Tween 0.1%
 1752 & BSA 0.02%) for 10 minutes at 20°C at rest. The blocking solution is then replaced with

1753 a mix of PBS Tween 0.1%, BSA 0.02%, poly(dI/dC)-oligonucleotide (150ng/250uL) and
1754 DNA to reach a final concentration of 1nM). The binding reaction is allowed for 1 hour
1755 at 20°C, mild shaking. Once the equilibrium is reached, the wells are washed 25 times
1756 with 200uL of washing buffer 1 (PBST Tween 0.1%). The DNA molecules attached to the
1757 antibodies are then eluted into 50uL of ddH2O by heating at 85°C for 25 minutes. The
1758 solution containing the eluted DNA is then transferred into DNA lobind eppendorf tubes
1759 for long term storage at -20°C.

1760

1761 **6.2.3.2 On beads**

1762 The following protocol has been performed using 10uL of streptavidin-coated beads per
1763 tube. First the beads are washed 3 times with 950uL of washing buffer (PBS Tween
1764 0.1%). Then the antibody is added to the beads (a solution of PBS with no antibody is
1765 added to the control tube). The antibodies are incubated with the beads for 1h30 at 20°C
1766 with mild shaking. Then the beads are washed 3 times using a magnet with the washing
1767 buffer and blocked with a blocking solution (PBS Tween 0.1%, BSA 0.02% & 150 ng of
1768 poly(dI/dC)-oligonucleotide) for 30 minutes at 20°C with mild shaking. The DNA library
1769 is added to reach a final concentration of 250nM in a total volume of 250uL. The binding
1770 reaction is allowed for 1 hour at 20°C, mild shaking. Once the equilibrium is reached,
1771 the beads are washed 5 times using a magnet with 950uL of washing buffer. The DNA
1772 molecules attached to the antibodies are then eluted into 25uL of ddH2O by heating at
1773 95°C for 10 minutes. The solution containing the eluted DNA is then transferred into DNA
1774 lobind eppendorf tubes for long term storage at -20°C.

1775

1776 **6.2.4 Deep-sequencing of the DNA library**

1777 At the end of the selection experiment, the DNA molecules in complex with the anti-
1778 bodies are collected to be prepared for the deep-sequencing procedure. To sum-up, the
1779 preparation consists in 2 successive PCR, a gel purification, a quantification with Qubit
1780 device and a multiplexing of the different selection experiments. 7uL of the DNA library
1781 eluted from the beads, or 14uL eluted from the plate are amplified by PCR (15 cycles).
1782 They are then PCR purified and amplified again (7-15 cycles). The primers used for these
1783 PCR have been designed to add some extra nucleotides on each extremity of the DNA
1784 sequences for 2 reasons: first some bare-codes are added so each sequence can be traced-
1785 back to its selection experiment (this is necessary when several selection experiments are
1786 sequenced at the same time). Second, short DNA sequences called adapters are added
1787 to allow the immobilization of the DNA molecules on the chip, their amplification and
1788 their deep-sequencing. Once the DNA samples have been amplified twice, they are run
1789 on an electrophoresis gel in order to verify if they have the correct length. Then each
1790 DNA sample is extracted from the gel and its concentration is measured using the Qubit

1791 device. The DNA sample of several selection experiments are pooled together and deep-
1792 sequenced. We can choose between different deep-sequencing services. We can choose the
1793 length of the sequence that are read, the direction in which the DNA is sequenced (from
1794 its 5' end to its 3' or in both directions) and the number of DNA molecules that can be
1795 sequenced. We used the Next-seq service to sequence DNA libraries with a small diversity
1796 (218 DNA ligands) with at least 500 000 reads (measuring 75nt) per experiment , in both
1797 directions. We used the Novaseq prestation to sequence the high diversity DNA libraries
1798 (32768 DNA ligands) with at least 8 millions of reads (measuring 100 nt) per experiment,
1799 in both directions.

1800 The sequences containing different barcodes were separated from each other using python.
1801 Only the sequences containing (1) a valid barcode, (2) a perfect match between the reads
1802 forward and reverse and (3) the correct length were analyzed further. around 80% of the
1803 reads obtained by Next Seq and more than 90% of the reads obtained by NovaSeq passed
1804 these selection process.

1805

1806 We used high-throughput sequencing platforms of I2BC (Gif-sur-Yvette) and of ICM
1807 (Salpêtrière, Paris).

1808 **6.3 Analysis**

1809 **6.3.1 Binding score based on SELEX data**

1810 At the end of the SELEX experiment, the DNA molecules in interaction with the antibody
1811 are detached by a heat treatment. The collected DNA are then amplified with 2 successive
1812 PCR reactions (Supp. 6.2.4) and the PCR product that contains amplified DNA is mixed
1813 with a loading dye (that increases the density of the solution), and loaded in an agarose
1814 gel (the gel contains a sensitive DNA gel stain that form DNA-dye complex that emits
1815 green light under UV excitation). Once the DNA sample is loaded, an electrical current
1816 is run through the gel so the negatively charged DNA moves across the gel according to
1817 the length of its DNA molecules. Under UV excitation, the dye attached to the DNA
1818 molecules emits light and the intensity of the signal increases with the number of DNA
1819 molecules present in the sample. We estimated the amount of selected DNA for each anti-
1820 body by measuring the intensity of the DNA signal on the gel using the image J software.
1821 With this approach, we defined a score of binding as the average signal of the selected
1822 DNA normalized with the signal of the molecular-weight size marker. The score of binding
1823 increases with the amount of DNA detected in the gel and thus with the average affinity
1824 of the antibody for the DNA library. This binding score has the advantage of reflecting
1825 the strength of binding of the antibody for the whole library. However, this score is not a
1826 direct measurement of affinity, it only gives an estimation of the amount of DNA molecules
1827 that interacted with the antibody.

1828

1829 **6.3.2 Binding score based on ELISA data**

1830 We used the ELISA technique to measure the affinity between one antibody and one DNA
1831 ligand. The output of the experiment is a dissociation constant. The principle is to per-
1832 form multiple binding experiments in parallel where we fix the concentration of either the
1833 DNA or the antibody, and vary the concentration of the other reactant. Then we plot the
1834 concentration of the formed complexes for each concentration. We thus obtain a binding
1835 curve from which we extract the dissociation constant (for more details, see the ELISA
1836 protocol 6.2.2). The measurement of the dissociation constant allowed us to define a bind-
1837 ing score as the invert of the dissociation constant. Thus a high binding score corresponds
1838 to a low dissociation constant and by definition a high affinity. The advantage of this
1839 binding score is that it is based on a quantitative measurement of the affinity. However,
1840 this score only reflects the binding strength of one antibody for one single DNA ligand
1841 and thus does not give access to the total distribution of the affinities of the DNA library.
1842

1843 **6.3.3 The basic analysis of the deep-sequencing data informs us on 3** 1844 **experimental biases**

1845 **6.3.3.1 The bias in the initial library**

1846 The DNA library is designed to be an equi-molar mix of DNA ligands. However, all the
1847 deep-sequencing data show that the number of copies are not uniformly distributed. We
1848 used the results of a selection experiment presented in Fig. 3.1 as an example. The number
1849 of copies in the initial library (represented on the x axis) ranges approximately from 10
1850 to 10^3 . This figure shows that there is a dynamic range of 2 orders of magnitude in the
1851 initial numbers of copies. Consequently, we must systematically compare the number of
1852 copies after the selection with the number of copies before the selection to take this bias
1853 into account (by calculating enrichments).

1854 **6.3.3.2 The non-specific binding of the DNA library to the plate**

1855 We systematically performed control experiments where the conditions are the same as in
1856 regular selection experiments except that there is no antibody during the binding step. We
1857 systematically collected DNA molecules at the end of the control experiments. The results
1858 of the control experiment are presented in Fig. 6.8. The presence of DNA molecules after
1859 the selection implies that the DNA interacts with the plastic or with the streptavidin of
1860 the plate (the streptavidin is used to immobilize the DNA or antibody molecules during
1861 selection).

1862

1863 6.3.3.3 The sampling bias

1864 We systematically observe more variability in small copy numbers than in high copy num-
 1865 bers (see Fig 3.1 as an example). This suggests that there is a sampling step in the
 1866 selection experiment that is responsible for this variability. We call sampling the action
 1867 of collecting a small amount of the DNA library and use this sample to continue the ex-
 1868 periment. During selection assay, multiple steps involve the sampling of the DNA library,
 1869 we identified 2 critical steps: (1) A small fraction of the DNA library interacts with the
 1870 target during the binding step and (2) we deep-sequence a sub-sample of the total DNA
 1871 ligand population before and after its selection. During these sampling steps, the DNA
 1872 ligands that are present in small number of copies are more sensitive to variability than the
 1873 other (If we take two samples from the same stock of DNA molecules and compare these
 1874 samples, we observe more variation in the number of copies of the less frequent molecules
 1875 than in the number of copies of the most frequent molecules).

1876 6.3.4 Sequence logo of the most selected loops

1877 (We used the online tool LogoMaker to generate our sequence logo). The sequence logo
 1878 is a graphical representation of the nucleotide conservation in a pool of aligned sequences
 1879 (in our case, the pool is composed of the most enriched DNA loops). The x axis of the
 1880 sequence logo represents the different positions of the nucleotides along the sequences and
 1881 the y axis gives information about the representation of the nucleotides at a given posi-
 1882 tion. For each position, the 4 nucleotides are stacked, their relative sizes indicate their
 1883 frequencies in the pool of sequences and the total height indicate the information content
 1884 of the position. The information content has been calculated in 2 steps (as described in
 1885 [Schneider and Stephens, 1990]):

1886

1887 First, the uncertainty measurement has been calculated as

$$H(l) = - \sum_{b=A}^C f(b,l) \log_2 f(b,l) \quad (6.2)$$

1888 where $H(l)$ is the uncertainty at the position l (along the sequence), b is one of the bases
 1889 (A, T, G, and C) and $f(b,l)$ is the frequency of base b , at position l , measured in the pool
 1890 of sequences. The uncertainty of each position is measured in bits and ranges from 0 to 2
 1891 (0 means there is one single base at a given position and 2 means that the 4 bases have
 1892 the same frequency at a given position).

1893 Then, the information content has been calculated as

$$R_{sequence}(l) = 2 - (H(l) + e(n)) \quad (6.3)$$

1894 where $R_{sequence}(l)$ is the amount of information present in the sequence at position l , 2 is
 1895 the maximum uncertainty at any given position and $e(n)$ is a correction factor (used when

1896 the pool of n sequences is small). The information content of each position is measured in
1897 bits and ranges from 0 to 2 (0 means that the 4 bases have the same frequency at a given
1898 position and 2 means there is one single base at a given position).

1899 6.3.5 Analysis of the motifs inside the loop

1900 We can define a model where the parameters are the length “ l ” and the position “ p ” of
1901 the motif along the loop of the DNA molecule. The principle of this analysis is to measure
1902 the average enrichment of all the sequences that share the same DNA motif at the same
1903 position in the loop, for each motif.

1904 6.3.6 Modeling the sampling bias

1905 We simulate a calibration experiment with a sampling noise that describes the binding of
1906 a limited number of DNA molecules to the antibodies. The input parameters of the simu-
1907 lation are m different sequences with m different affinities for the antibody. The antibody
1908 is present in n copies and thus a maximum of n DNA molecules can bind the antibody
1909 and pass the selection. The output data are enrichments of the selected sequences. The
1910 data are simulated in 3 steps:

1911

1912 1- Simulation of the frequencies of each DNA sequence, before and after the selection

1913

1914 **Initial library:** We make the assumption that the DNA sequences are uniformly dis-
1915 tributed in the library before the selection. Consequently, all the sequences have the same
1916 frequency in the initial library.

1917 **Selected library:** If the DNA sequences are uniformly distributed in the initial library,
1918 their frequency after the selection is equivalent to their enrichment (the measurement of
1919 the enrichment is presented in section 4.3.2.2, for different selection models).

1920

1921 2- Introduction of sampling noise to selected DNA sequences

1922

1923 In order to add a sampling noise to the simulated data, we transformed the frequency of
1924 each sequence into number of copies:

1925 **Initial library:** The DNA sequences in the initial library are not impacted by the sam-
1926 pling noise we describe in this section (the sampling happens during selection), conse-
1927 quently their copy numbers are uniformly distributed (the copy numbers are representa-
1928 tive of the frequencies in the total population)

1929 **Selected library:** We introduced sampling noise in the selected DNA sequences¹ by sim-
1930 ulating their copy numbers using a multinomial law: n (=number of DNA molecules that

¹Without sampling noise in the simulation, the relative number of copies are directly linked to the frequencies. With sampling noise in the simulation, the relative number of copies are not directly linked to the frequencies anymore.

1931 bind the antibody) drawings are performed, and m (=diversity of the library) different DNA
1932 ligands can be drawn every time, with a probability p_i corresponding to their frequency
1933 in the selected library. The probability mass function of the multinomial distribution is
1934 defined as following:

$$P(N_1 = n_1, \dots, N_m = n_m) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m} \quad (6.4)$$

1935 with N_i the number of copies of the sequence i that bind the antibody, p_i the enrich-
1936 ment (normalized) of the sequence i (i.e. its frequency in selected library) and n the total
1937 number of sequences that bind the antibodies and pass the selection. We obtain for each
1938 selected DNA sequence, a number of copies that depends on both the selection model and
1939 the sampling noise.

1940

1941 3- Measurement of the enrichment values

1942

1943 We calculate the enrichment of each DNA sequence as the ratio between its number
1944 of copies after and before its selection (we normalize the enrichment values so their sum
1945 equals 1).

1946

1947 **6.3.7 Impact of the calibration protocol on the shape of the binding** 1948 **curves**

1949 Because our calibration protocol is different from the ones commonly used in the literature
1950 (we fixed the concentration of the target and varied the concentration of the library of
1951 ligands while the other studies do the opposite), we studied in this section the impact of
1952 the calibration protocol on the output data of the assay. To do so, we simulated different
1953 calibration protocols and compared the probabilities of selection (using eq. 6.1) as well
1954 as the shape of the binding curves for both protocols (Fig. 6.5). The binding curves are
1955 qualitatively different for both protocols. This results supports the fact that we need to
1956 define a selection model to describe the relationship between the experimental enrichments
1957 and the dissociation constants we want to measure.

1958 **6.4 Experiments**

1959 **6.4.1 Reproducibility of the SELEX experiments**

1960 We evaluated the reproducibility of the SELEX experiments by comparing the measured
1961 enrichments of each sequence between two replicate experiments. In Fig. 6.9 B are pre-
1962 sented the results of the comparison of 2 replicate experiments for one Limited antibody.
1963 The highest enrichments are highly reproducible while the smallest enrichments are not
1964 reproducible between replicates. The limit between the most and the least reproducible en-

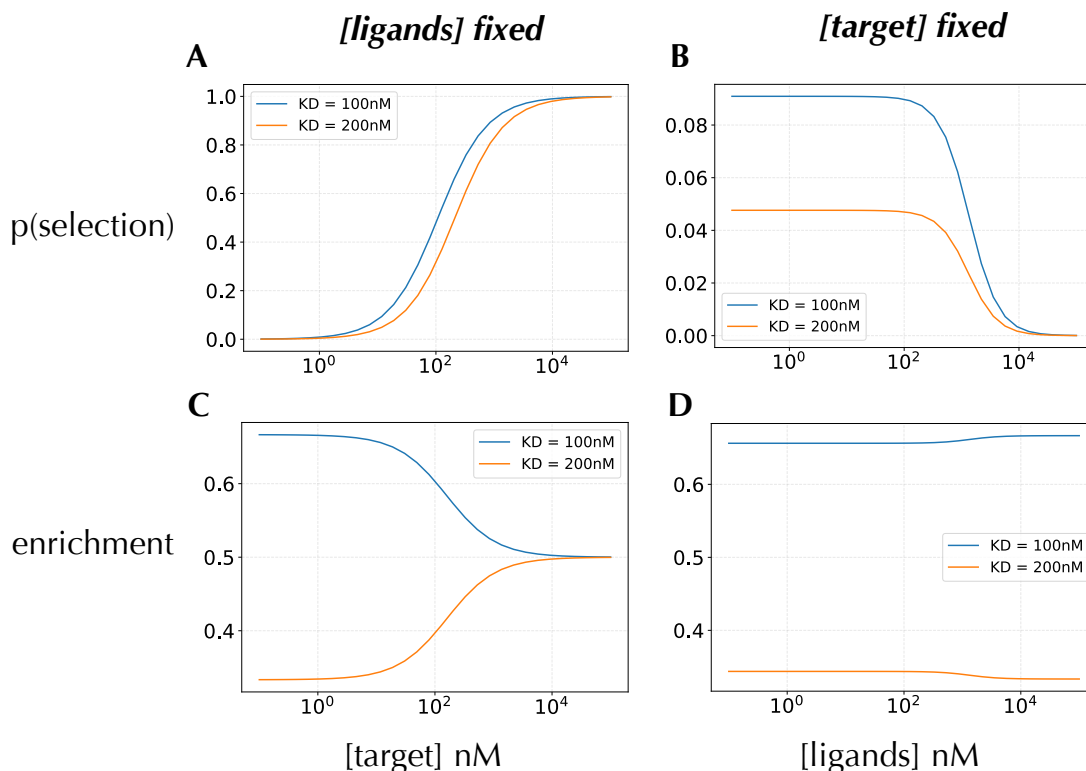


Figure 6.5: **The shape of the binding curves depends on the calibration protocol.** In this figure, the target refers to the antibody, and the ligands to the DNA library. We simulated 2 selection experiments. In the first column, we fixed the concentration of the ligands and varied the concentration of the target. In the second column, we fixed the concentration of target and varied the concentration of the ligands. We plotted the probabilities of selection as a function of the concentrations of targets in figA or as a function of the concentrations of ligands in figB. We presented in figC and D the corresponding binding curves with enrichments measured using eq. 4.3. We obtain different binding curves according to the calibration protocol we use.

1965 richments correspond to the same limit that separates the outliers from the bulk sequences
 1966 in fig 6.9 A. The absence of reproducibility for the lowest enrichments is explained by the
 1967 fact that the copy numbers of the corresponding sequences are dominated by sampling
 1968 noise rather than selection. Consequently, a simple analysis of the selection based on en-
 1969 richments can only be performed for the highest values. We need a more sophisticated
 1970 analysis to extract the selection signal from the sequences in the bulk

1971 6.4.2 Comparison of selection with control experiment

1972 To verify if the most enriched DNA loops have been specifically selected by the antibody
 1973 and not by the plastic of the plate, we systematically compared enrichments of selection
 1974 experiment with enrichments of control experiment (where no antibody is present during
 1975 selection). We compared in (Fig. 6.10) the enrichment of one Germline and one Limited

1976 antibodies with their respective control experiments. For both antibodies, there is no
1977 correlation between the enrichments measured in the selection experiment and the enrich-
1978 ments measured in the control experiment. This results confirms that the most enriched
1979 DNA ligands are selected by the antibody and are not the product of an experimental
1980 bias.

1981 6.4.3 Impact of the constant region on the selection of the DNA loops 1982 by Germline antibodies

1983 We systematically compared the selection of the DNA loops with 2 different constant
1984 regions for every Germline and Limited antibodies. We show in Fig. 6.6 the enrichments
1985 of the 2 DNA libraries for one Germline antibody. We observe no correlation between
1986 the enrichments for this antibody (we observed no correlation for the other Germline
1987 antibodies as well, (data not shown)). The constant region of the DNA molecules has an
1988 impact on the selection of the loops for the Germline antibodies we tested.

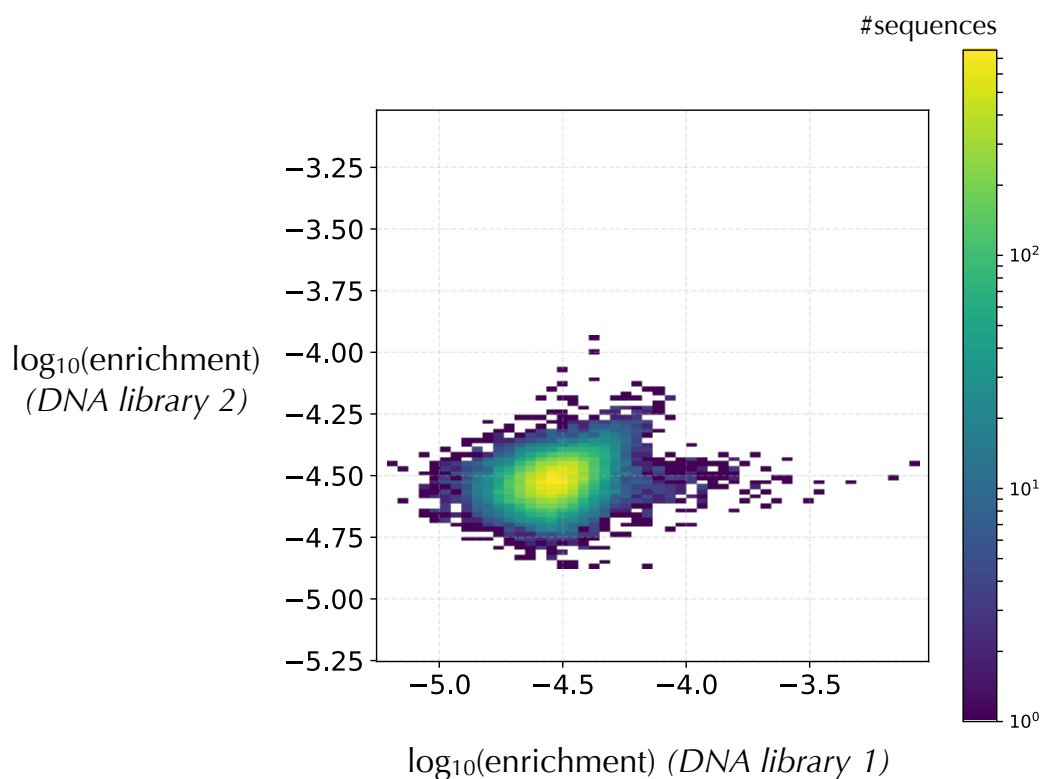


Figure 6.6: **Impact of the constant region on the selection of the loops for 1 Germline antibody.** We compared the enrichments of the DNA loops with 2 different constant regions for the Germline antibody (GNT3). We observe no correlation between the enriched DNA loops. The constant region impacts the selection of the loop for this antibody.

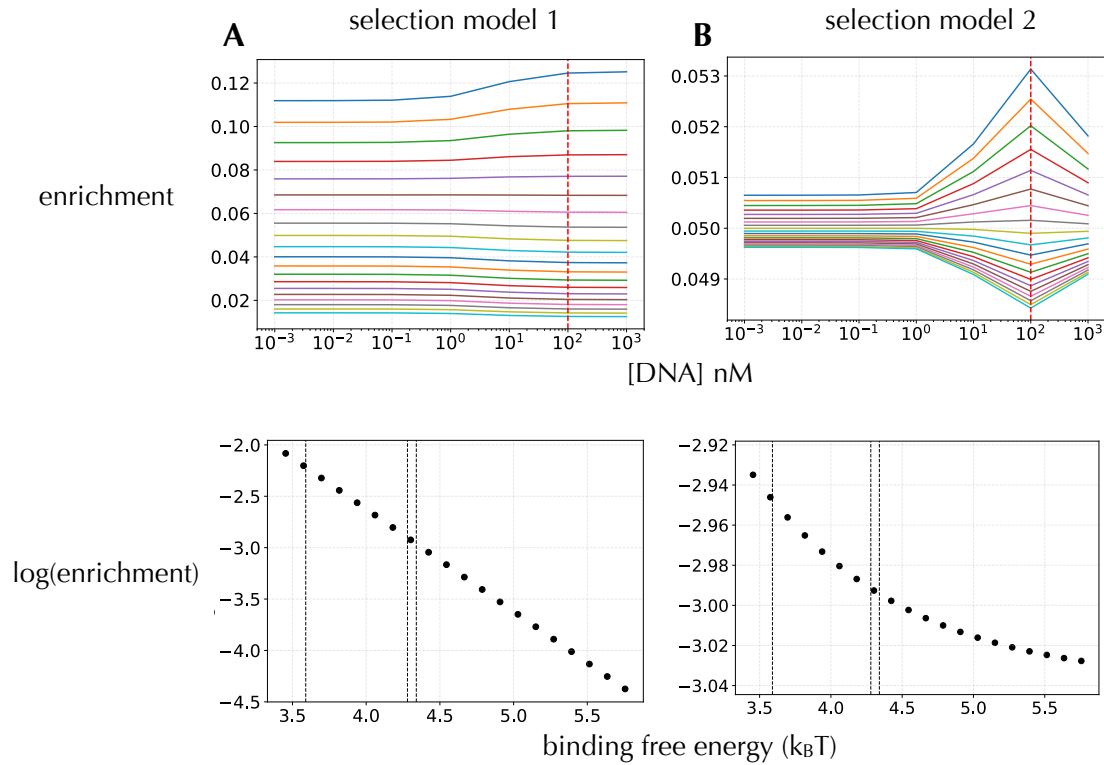
1989 **6.5** Supplementary figures

Figure 6.7: **Simulations show that the nature of the relationship between enrichments and binding energies depends on the selection model.** The parameters of the simulations are the same as in Fig. 4.5. The red dashed lines indicate the DNA concentration regime from which the enrichments have been picked to generate the bottom plots. A. The first selection model predicts a linear relationship between the logarithm of the enrichments and the binding energies. B. The second selection model predicts a non-linear relationship between the logarithm of the enrichments and the binding energies. However, we can make the approximation that the relationship is linear if the dynamic range of the binding energies we consider is small enough. The 3 black dashed lines represent the binding energies of the 3 data-points estimated experimentally by ELISA. These data-points have binding energies that are close enough to be in a linear relationship with their corresponding enrichment values.

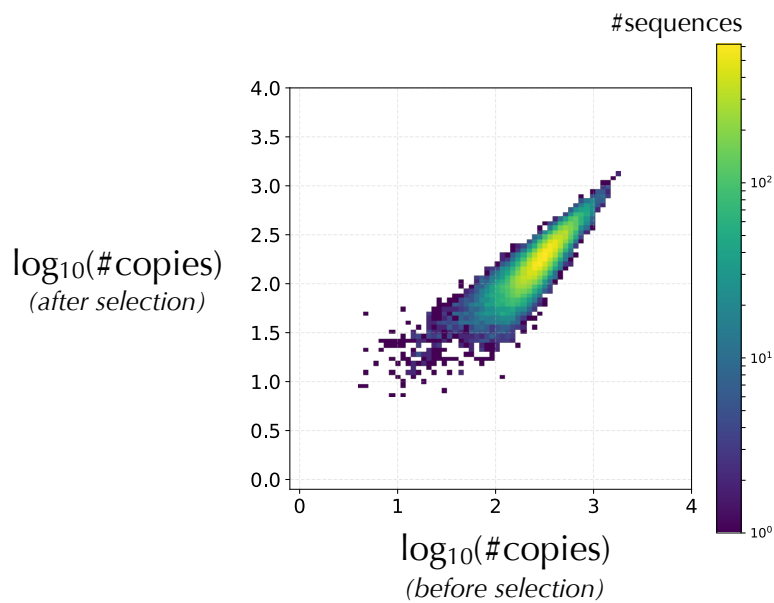


Figure 6.8: **The binding of the DNA library to the plate is revealed by a control experiment.** We performed a control experiment where no antibody was present in the plate during selection. We managed to collect DNA molecules at the end of the selection experiment. Here we present the number of copies of the DNA library before and after the control experiment.

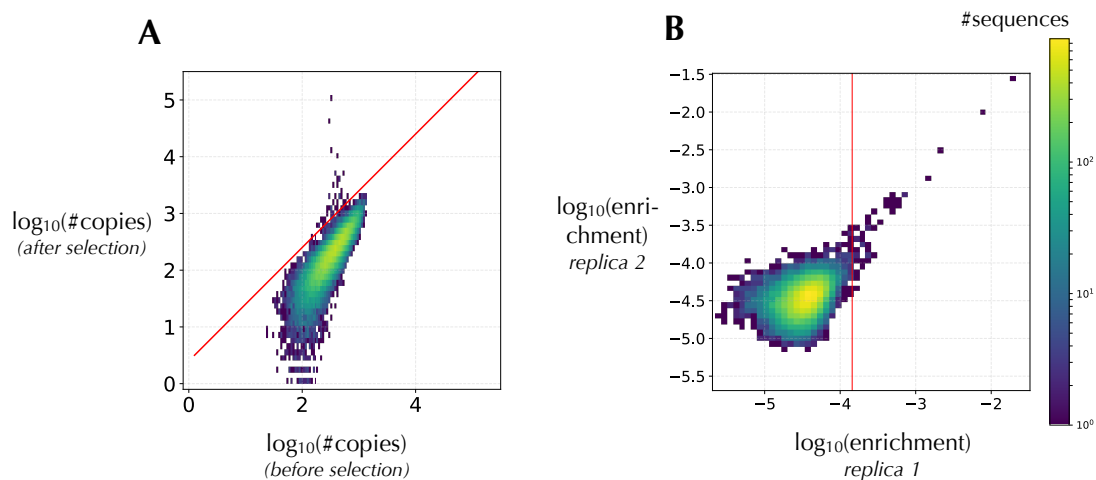


Figure 6.9: **The highest enrichments (dominated by specific binding) are reproducible while the smallest enrichments (dominated by sampling noise) are not reproducible.** We present in figA the results of a selection experiment (DNA library 2 selected by LBT1). The red line corresponds to the enrichment limit above which the sequences are called outliers (they are dominated by specific binding) and below which the sequences are called bulk sequences (they are dominated by sampling noise). In figB are compared the enrichments between 2 replicate experiments. The red line correspond to the same enrichment threshold as presented in figA. The threshold value that separates the outliers from the bulk also separate the reproducible enrichments from the non-reproducible enrichments.

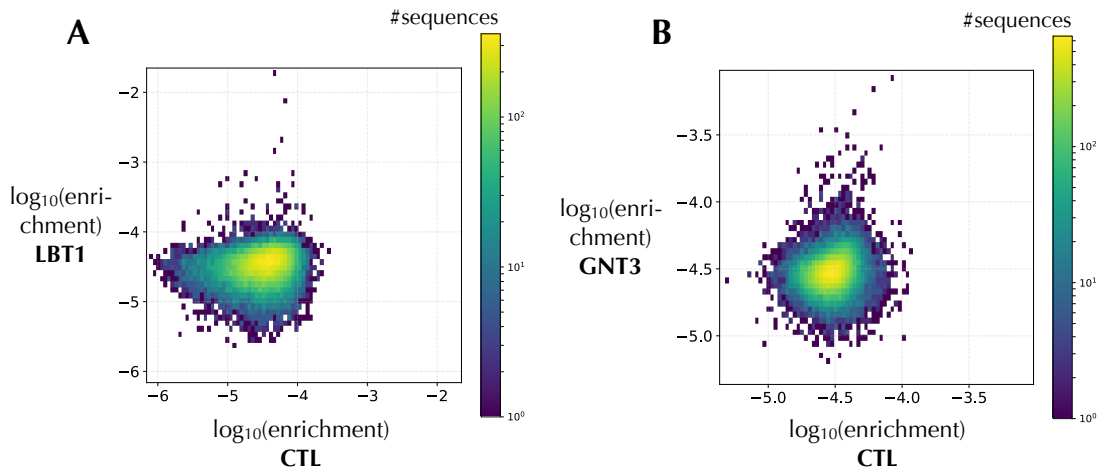


Figure 6.10: The comparison of selection experiment with control experiment shows that the DNA sequences are specifically selected by the antibodies. A. We compare the enrichments between selection and control experiment for one Limited antibody (LBT1). B. We compare the enrichments between selection and control experiment for one Germline antibody (GNT3).

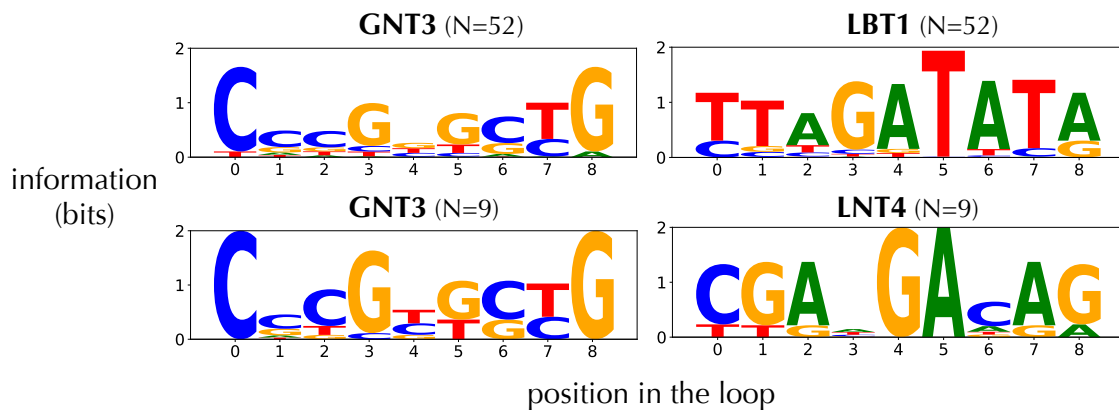


Figure 6.11: Using the same number of selected DNA sequences, the average information content of Limited antibodies is higher than the average information content of Germline antibodies. Here we present the 9 nucleotides forming the loop (including the first (position 0) and last (position 8) nucleotides that close the loop).

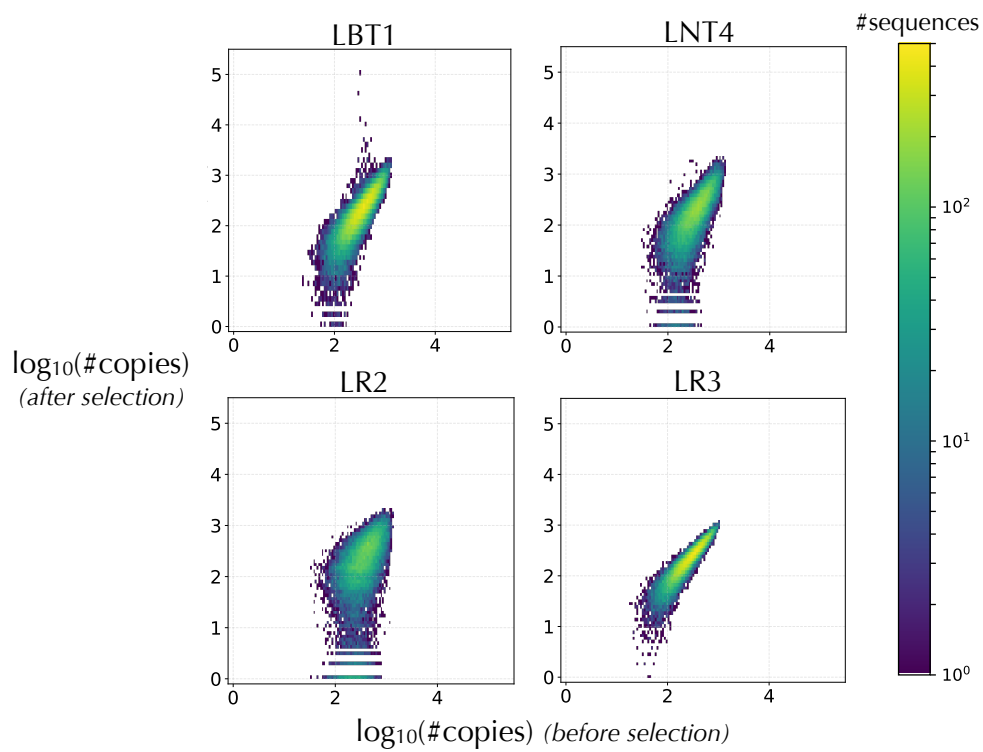


Figure 6.12: Selection results of 2 TOP and 2 Random Limited antibodies.

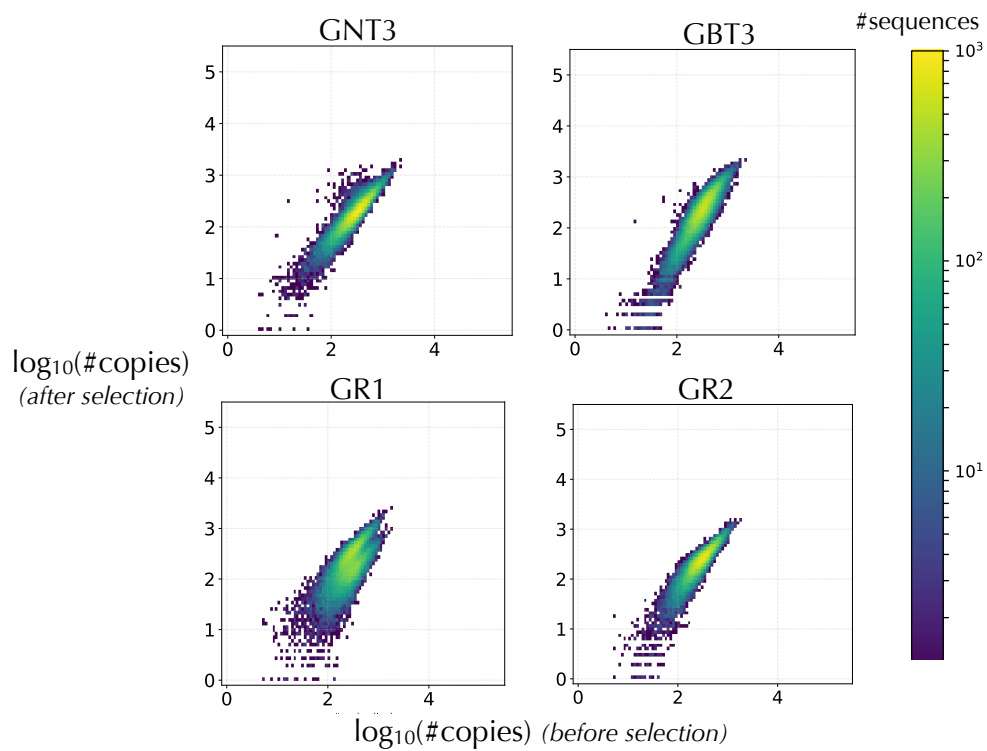


Figure 6.13: Selection results of 2 TOP and 2 Random Germline antibodies

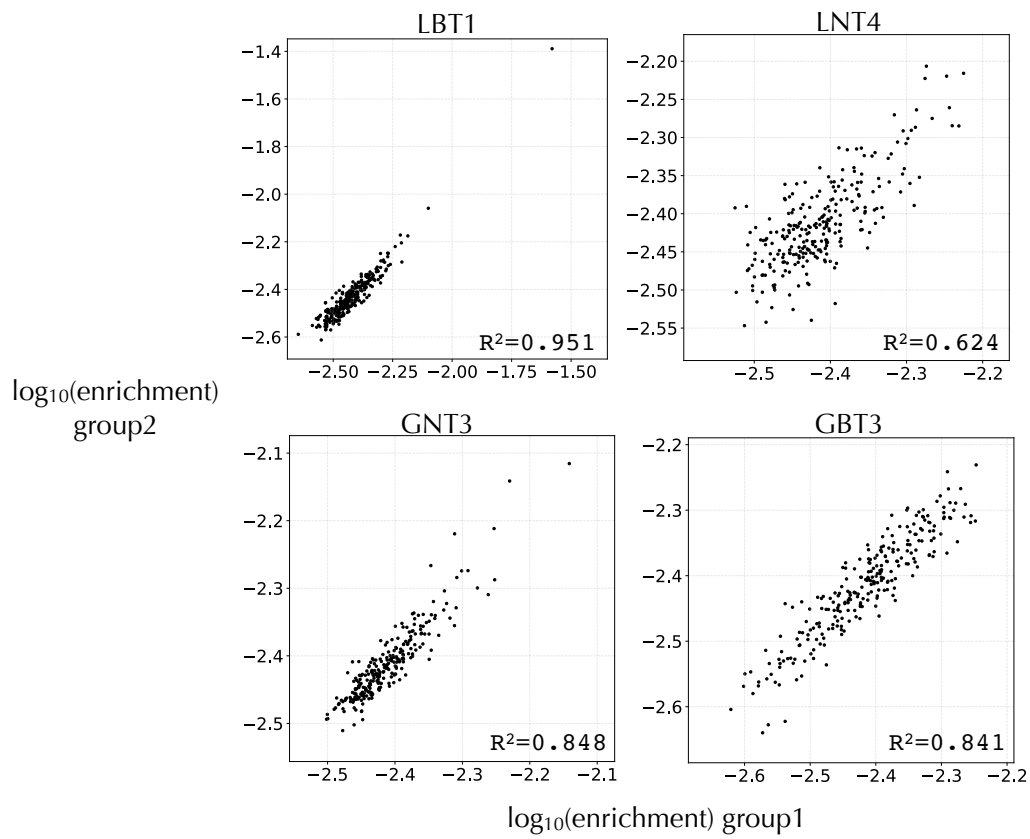


Figure 6.14: Cross-validation results for 2 Germline and 2 Limited antibodies using a unique set of parameters (length motif = 4, position motif = 4)

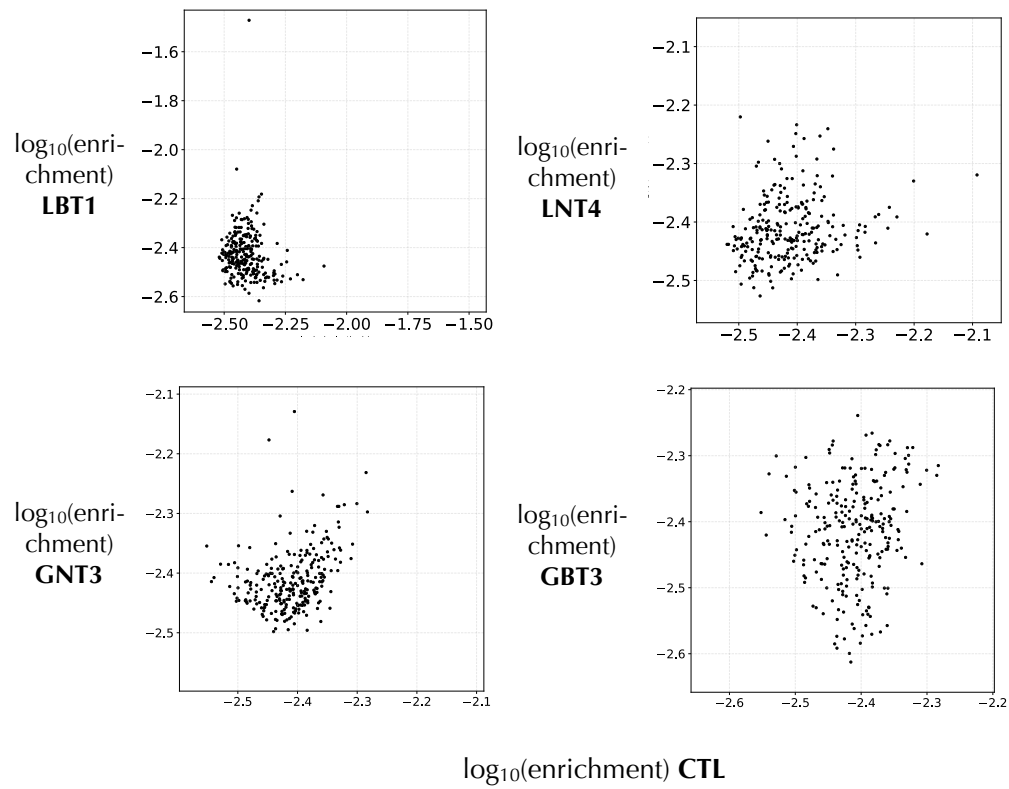


Figure 6.15: Comparison of motif enrichments between selection and control experiments for 2 Germline and 2 Limited antibodies using a unique set of parameters (length motif = 4, position motif = 4)

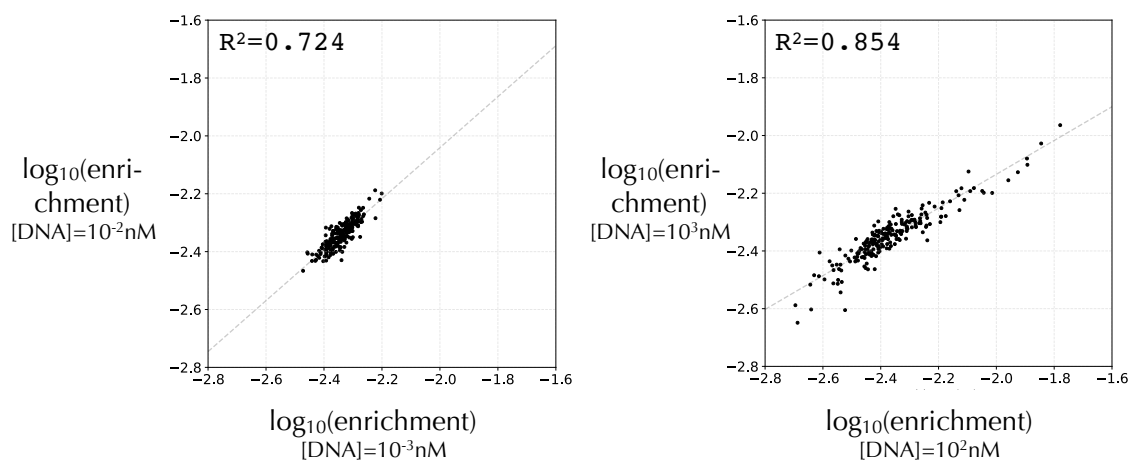


Figure 6.16: **The ranking of the selected DNA sequences is more conserved among intermediate DNA concentration ranging from 10^1 to 10^3 nM.** A. Comparison of enrichments for $[DNA]=10^{-3}$ and $[DNA]=10^{-2}$ nM. B. Comparison of enrichments for $[DNA]=10^2$ and $[DNA]=10^3$ nM. The R^2 score is higher in fig B than in fig A.

Bibliography

1995

1996 Rhys M. Adams, Thierry Mora, Aleksandra M. Walczak, and Justin B. Kinney. Measuring
1997 the sequence-affinity landscape of antibodies with massively parallel titration curves.
1998 *eLife*, 5(DECEMBER2016):1–27, 2016. ISSN 2050084X. doi: 10.7554/eLife.23156.

1999 Arjun K. Aditham, Craig J. Markin, Daniel A. Mokhtari, Nicole DelRosso, and Polly M.
2000 Fordyce. High-Throughput Affinity Measurements of Transcription Factor and DNA
2001 Mutations Reveal Affinity and Specificity Determinants. *Cell Systems*, 12(2):112–
2002 127.e11, 2021. ISSN 24054720. doi: 10.1016/j.cels.2020.11.012. URL [https://doi.
2003 org/10.1016/j.cels.2020.11.012](https://doi.org/10.1016/j.cels.2020.11.012).

2004 Sébastien Boyer, Dipanwita Biswas, Ananda Kumar Soshee, Natale Scaramozzino,
2005 Clément Nizak, and Olivier Rivoire. Hierarchy and extremes in selections from pools
2006 of randomized proteins-supporting informations. *Proceedings of the National Academy
2007 of Sciences*, 113(13):3482–3487, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1517813113.
2008 URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1517813113>.

2009 C. D. Carlson, C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria,
2010 Y. Lee, F. Cerrina, and A. Z. Ansari. Specificity landscapes of DNA binding molecules
2011 elucidate biological function. *Proceedings of the National Academy of Sciences*, 107
2012 (10):4544–4549, 2010. ISSN 0027-8424. doi: 10.1073/pnas.0914023107. URL [http:
2013 //www.pnas.org/cgi/doi/10.1073/pnas.0914023107](http://www.pnas.org/cgi/doi/10.1073/pnas.0914023107).

2014 Jacob C Carlson, Ahmed H Badran, Drago A Guggiana-Nilo, and David R Liu. Negative
2015 selection and stringency modulation in phage-assisted continuous evolution. *Nature
2016 Chemical Biology*, 10(3):216–222, 2014. ISSN 1552-4450. doi: 10.1038/nchembio.1453.
2017 URL <http://www.nature.com/doi/10.1038/nchembio.1453>.

2018 Dana Chen, Yaron Orenstein, Rada Golodnitsky, Michal Pellach, Dorit Avrahami, Chaim
2019 Wachtel, Avital Ovadia-Shochat, Hila Shir-Shapira, Adi Kedmi, Tamar Juven-Gershon,
2020 Ron Shamir, and Doron Gerber. SELMAP - SELEX affinity landscape MAPping of
2021 transcription factor binding sites using integrated microfluidics. *Scientific Reports*, 6
2022 (August):1–13, 2016. ISSN 20452322. doi: 10.1038/srep33351. URL [http://dx.doi.
2023 org/10.1038/srep33351](http://dx.doi.org/10.1038/srep33351).

2024 Justin Crocker, Namiko Abe, Lucrezia Rinaldi, Alistair P. McGregor, Nicolás Frankel,
2025 Shu Wang, Ahmad Alsawadi, Philippe Valenti, Serge Plaza, François Payre, Richard S.

- 2026 Mann, and David L. Stern. Low affinity binding site clusters confer HOX specificity
2027 and regulatory robustness. *Cell*, 160(1-2):191–203, 2015. ISSN 10974172. doi: 10.1016/
2028 j.cell.2014.11.041.
- 2029 Elvira D’Ippolito, Kilian Schober, Magdalena Nauwerth, and Dirk H. Busch. T cell
2030 engineering for adoptive T cell therapy: safety and receptor avidity. *Cancer Im-*
2031 *munology, Immunotherapy*, 68(10):1701–1712, 2019. ISSN 14320851. doi: 10.1007/
2032 s00262-019-02395-9. URL <https://doi.org/10.1007/s00262-019-02395-9>.
- 2033 Daniel Dominguez, Peter Freese, Maria S. Alexis, Amanda Su, Myles Hochman, Tsultrim
2034 Palden, Cassandra Bazile, Nicole J. Lambert, Eric L. Van Nostrand, Gabriel A. Pratt,
2035 Gene W. Yeo, Brenton R. Graveley, and Christopher B. Burge. Sequence, Structure,
2036 and Context Preferences of Human RNA Binding Proteins. *Molecular Cell*, 70(5):854–
2037 867.e9, 2018. ISSN 10974164. doi: 10.1016/j.molcel.2018.05.001. URL <https://doi.org/10.1016/j.molcel.2018.05.001>.
- 2039 H. N. Eisen and A. K. Chakraborty. Evolving concepts of specificity in immune reactions.
2040 *Proceedings of the National Academy of Sciences*, 107(52):22373–22380, 2010. ISSN
2041 0027-8424. doi: 10.1073/pnas.1012051108. URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1012051108>.
- 2043 Douglas M. Fowler, Carlos L. Araya, Sarel J. Fleishman, Elizabeth H. Kellogg, Ja-
2044 son J. Stephany, David Baker, and Stanley Fields. High-resolution mapping of protein
2045 sequence-function relationships. *Nature Methods*, 7(9):741–746, 2010. ISSN 15487091.
2046 doi: 10.1038/nmeth.1492. URL <http://dx.doi.org/10.1038/nmeth.1492>.
- 2047 A. J.T. George and D. Gray. Receptor editing during affinity maturation [1]. *Immunology*
2048 *Today*, 20(4):196, 1999. ISSN 01675699. doi: 10.1016/S0167-5699(98)01408-X.
- 2049 Guangming Gu, Tingting Wang, Yang Yang, Xinhui Xu, and Jinke Wang. An Im-
2050 proved SELEX-Seq Strategy for Characterizing DNA-Binding Specificity of Transcrip-
2051 tion Factor: NF- κ B as an Example. *PLoS ONE*, 8(10), 2013. ISSN 19326203. doi:
2052 10.1371/journal.pone.0076109.
- 2053 Barton F Haynes, Judith Fleming, E William St Clair, Herman Katinger, Gabriela Stiegler,
2054 Renate Kunert, James Robinson, Richard M Scarce, Kelly Plonk, Herman F Staats,
2055 Thomas L Ortel, Hua-xin Liao, and S Munir Alam. Cardiolipin Polyspecific Autore-
2056 activity in Two Broadly Neutralizing HIV-1 Antibodies. 1906(2005):1906–1908, 2013.
2057 doi: 10.1126/science.1111781.
- 2058 Alina Isakova, Romain Groux, Michael Imbeault, Pernille Rainer, Daniel Alpern, Ric-
2059 cardo Dainese, Giovanna Ambrosini, Didier Trono, Philipp Bucher, and Bart Deplancke.
2060 SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Na-*
2061 *ture Methods*, 14(3):316–322, 2017. ISSN 15487105. doi: 10.1038/nmeth.4143. URL
2062 <http://dx.doi.org/10.1038/nmeth.4143>.

- 2063 L. C. James. Antibody Multispecificity Mediated by Conformational Diversity. *Science*,
2064 299(5611):1362–1367, 2003. ISSN 00368075. doi: 10.1126/science.1079731. URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1079731>.
2065
- 2066 Inga Jarmoskaite, Ishraq Alsadhan, Pavanapuresan P. Vaidyanathan, and Daniel Her-
2067 schlag. How to measure and evaluate binding affinities. *eLife*, 9:1–34, 2020. ISSN
2068 2050084X. doi: 10.7554/ELIFE.57264.
- 2069 Jin Jiang and Michael Levine. Binding affinities and cooperative interactions with bHLH
2070 activators delimit threshold responses to the dorsal gradient morphogen. *Cell*, 72(5):
2071 741–752, 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90402-C.
- 2072 Arttu Jolma, Teemu Kivioja, Jarkko Toivonen, Lu Cheng, Gonghong Wei, Martin Enge,
2073 Mikko Taipale, Juan M. Vaquerizas, Jian Yan, Mikko J. Sillanpää, Martin Bonke,
2074 Kimmo Palin, Shaheynoor Talukder, Timothy R. Hughes, Nicholas M. Luscombe, Esko
2075 Ukkonen, and Jussi Taipale. Multiplexed massively parallel SELEX for characterization
2076 of human transcription factor binding specificities. *Genome Research*, 20(6):861–873,
2077 2010. ISSN 10889051. doi: 10.1101/gr.100552.109.
- 2078 Arttu Jolma, Jian Yan, Thomas Whittington, Jarkko Toivonen, Kazuhiro R. Nitta, Pasi
2079 Rastas, Ekaterina Morgunova, Martin Enge, Mikko Taipale, Gonghong Wei, Kimmo
2080 Palin, Juan M. Vaquerizas, Renaud Vincentelli, Nicholas M. Luscombe, Timothy R.
2081 Hughes, Patrick Lemaire, Esko Ukkonen, Teemu Kivioja, and Jussi Taipale. DNA-
2082 binding specificities of human transcription factors. *Cell*, 152(1-2):327–339, 2013. ISSN
2083 00928674. doi: 10.1016/j.cell.2012.12.009. URL [http://dx.doi.org/10.1016/j.cell.](http://dx.doi.org/10.1016/j.cell.2012.12.009)
2084 [2012.12.009](http://dx.doi.org/10.1016/j.cell.2012.12.009).
- 2085 Caitlin A. Kowalsky and Timothy A. Whitehead. Determination of binding affinity upon
2086 mutation for type I dockerin-cohesin complexes from *Clostridium thermocellum* and
2087 *Clostridium cellulolyticum* using deep sequencing. *Proteins: Structure, Function and*
2088 *Bioinformatics*, 84(12):1914–1928, 2016. ISSN 10970134. doi: 10.1002/prot.25175.
- 2089 Nicole J. Lambert, Alex Robertson, Mohini Jangi, Sean McGeary, Philip A. Sharp, and
2090 Christopher B. Burge. structural binding specificity of RNA binding proteins. *Mol Cell*,
2091 54(5):887–900, 2014. doi: 10.1016/j.molcel.2014.04.016.RNA.
- 2092 Alan Lapedes and Robert Farber. The geometry of shape space: Application to influenza.
2093 *Journal of Theoretical Biology*, 212(1):57–69, 2001. ISSN 00225193. doi: 10.1006/jtbi.
2094 2001.2347.
- 2095 Daniel D. Le, Tyler C. Shimko, Arjun K. Aditham, Allison M. Keys, Scott A. Longwell,
2096 Yaron Orenstein, and Polly M. Fordyce. Comprehensive, high-resolution binding energy
2097 landscapes reveal context dependencies of transcription factor binding. *Proceedings*
2098 *of the National Academy of Sciences*, page 201715888, 2018. ISSN 0027-8424. doi:

- 2099 10.1073/pnas.1715888115. URL [http://www.pnas.org/lookup/doi/10.1073/pnas.](http://www.pnas.org/lookup/doi/10.1073/pnas.1715888115)
2100 [1715888115](http://www.pnas.org/lookup/doi/10.1073/pnas.1715888115).
- 2101 Sean E. McGeary, Kathy S. Lin, Charlie Y. Shi, Thy M. Pham, Namita Bisaria, Gina M.
2102 Kelley, and David P. Bartel. The biochemical basis of microRNA targeting efficacy.
2103 *Science*, 366(6472), 2019. ISSN 10959203. doi: 10.1126/science.aav1741.
- 2104 Alan S. Perelson and George F. Oster. Theoretical studies of clonal selection: Minimal anti-
2105 body repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical*
2106 *Biology*, 81(4):645–670, 1979. ISSN 10958541. doi: 10.1016/0022-5193(79)90275-3.
- 2107 Julie Prigent, Annaëlle Jarossay, Cyril Planchais, Caroline Eden, Jérémy Dufloo, Ayrin
2108 Kök, Valérie Lorin, Oxana Vratskikh, Thérèse Couderc, Timothée Bruel, Olivier
2109 Schwartz, Michael S. Seaman, Oliver Ohlenschläger, Jordan D. Dimitrov, and Hugo
2110 Mouquet. Conformational Plasticity in Broadly Neutralizing HIV-1 Antibodies Trig-
2111 gers Polyreactivity. *Cell Reports*, 23(9):2568–2581, 2018. ISSN 22111247. doi:
2112 10.1016/j.celrep.2018.04.101.
- 2113 Chaitanya Rastogi, H. Tomas Rube, Judith F. Kribelbauer, Justin Crocker, Ryan E.
2114 Loker, Gabriella D. Martini, Oleg Laptenko, William A. Freed-Pastor, Carol Prives,
2115 David L. Stern, Richard S. Mann, and Harmen J. Bussemaker. Accurate and sensitive
2116 quantification of protein-DNA binding affinity. *Proceedings of the National Academy of*
2117 *Sciences*, page 201714376, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1714376115. URL
2118 <http://www.pnas.org/lookup/doi/10.1073/pnas.1714376115>.
- 2119 Douglas M. Robinson, David T. Jones, Hirohisa Kishino, Nick Goldman, and Jeffrey L.
2120 Thorne. Protein evolution with dependence among codons due to tertiary structure.
2121 *Molecular Biology and Evolution*, 20(10):1692–1704, 2003. ISSN 07374038. doi: 10.
2122 1093/molbev/msg184.
- 2123 H. Tomas Rube, Justin Crocker, William A. Freed-Pastor, Chaitanya Rastogi, Judith F.
2124 Kribelbauer, David L. Stern, Harmen J. Bussemaker, Ryan E. Loker, Gabriella D.
2125 Martini, Richard S. Mann, Carol Prives, and Oleg Laptenko. Accurate and sensitive
2126 quantification of protein-DNA binding affinity. *Proceedings of the National Academy of*
2127 *Sciences*, 115(16):E3692–E3701, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1714376115.
- 2128 Yonatan Savir and Tsvi Tlusty. Conformational proofreading: The impact of conforma-
2129 tional changes on the specificity of molecular recognition. *PLoS ONE*, 2(5), 2007. ISSN
2130 19326203. doi: 10.1371/journal.pone.0000468.
- 2131 Thomas D Schneider and R Michael Stephens. Sequence logos:. *Nucleic Acids Research*,
2132 18(20):6097–6100, 1990.
- 2133 Steven Schulz, Sébastien Boyer, Matteo Smerlak, Simona Cocco, Rémi Monasson, Clément
2134 Nizak, and Olivier Rivoire. Parameters and determinants of responses to selection in

- 2135 antibody libraries. *PLoS Computational Biology*, 17(3):1–24, 2021. ISSN 15537358. doi:
2136 10.1371/JOURNAL.PCBI.1008751.
- 2137 Raffi Tonikian, Yingnan Zhang, Stephen L. Sazinsky, Bridget Currell, Jung Hua Yeh,
2138 Boris Reva, Heike A. Held, Brent A. Appleton, Marie Evangelista, Yan Wu, Xiaofeng
2139 Xin, Andrew C. Chan, Somasekar Seshagiri, Laurence A. Lasky, Chris Sander, Charles
2140 Boone, Gary D. Bader, and Sachdev S. Sidhu. A specificity map for the PDZ domain
2141 family. *PLoS Biology*, 6(9):2043–2059, 2008. ISSN 15449173. doi: 10.1371/journal.pbio.
2142 0060239.

Résumé

Les protéines sont des biomolécules fondamentales de tout être vivant. Leurs fonctions très variées (comme la catalyse de réactions biologiques ou le transport de biomolécules dans la cellule) sont contrôlées par leur structure tridimensionnelle ; elle-même déterminée par leur séquence d'acides aminés. Le rôle des anticorps, par exemple, est de reconnaître et neutraliser des molécules cibles (appelées antigènes) provenant de pathogènes en se liant à ces dernières. Leur fonction est directement liée à leur affinité (force d'interaction) et spécificité (précision) de liaison à ces molécules cibles. Les anticorps sont connus pour leur haute spécificité de liaison car ils doivent distinguer les antigènes (provenant de pathogènes) des auto-antigènes (provenant de l'organisme hôte). Ils sont, par exemple, couramment utilisés comme outils permettant de révéler des structures cellulaires en se liant spécifiquement à des protéines d'intérêt. Cependant, il a également été démontré que les anticorps ont la capacité d'interagir avec différents antigènes et donc d'être poly-spécifiques (le cas des anticorps neutralisants à large spectre est présenté plus loin). Lorsqu'un anticorps rencontre un antigène pour la première fois, il entre dans un processus de maturation d'affinité durant lequel il accumule des mutations dans sa séquence afin d'augmenter son affinité pour l'antigène. Durant ce processus évolutif, il a été démontré que la spécificité des anticorps est également modifiée. Différentes théories ont été proposées pour décrire la coévolution de l'affinité et de la spécificité durant la maturation des anticorps. Il existe à ce jour deux principaux scénarios : dans le premier, les anticorps ont initialement la capacité de reconnaître plusieurs antigènes, mais perdent cette faculté durant la maturation d'affinité afin de se spécialiser dans la liaison à une molécule cible particulière ; dans le deuxième scénario, les anticorps développent la capacité de reconnaître plusieurs antigènes durant leur maturation d'affinité. Par exemple, certains anticorps ayant mûri durant des années contre le VIH (« anticorps neutralisants à large spectre » ou « bnAb ») ont développé la capacité à reconnaître différentes souches de VIH et parfois des auto-antigènes. Des observations similaires ont été faites concernant les anticorps maturés artificiellement dans le cadre de la thérapie cellulaire « CAR-T » (les lymphocytes de patients atteints de cancers sont collectés, maturés artificiellement pour reconnaître les cellules cancéreuses et réinjectés dans les patients). Ces deux scénarios suggèrent l'existence de différentes pressions de sélection : (1) dans le premier scénario, les anticorps évoluent pour se lier à un unique antigène ; (2) dans le deuxième scénario, ils évoluent pour se lier à différents antigènes. Nous pouvons également envisager un troisième scénario dans lequel l'affinité et la spécificité des anticorps sont indépendantes. Cependant, nous n'avons pas encore élucidé le rôle des pressions de sélection dans l'éventuelle relation liant l'affinité et la spécificité des anticorps.

En résumé, bien que de nombreux travaux aient été réalisés dans le but d'étudier l'affinité des anticorps d'une part, et leur spécificité d'autre part, nous avons à ce jour une compréhension limitée de la relation entre ces deux propriétés fonctionnelles. Par ailleurs,

tandis que la métrique d'affinité est clairement définie (elle correspond à la constante de dissociation entre la protéine et sa cible), il n'existe pas de métrique consensuelle de spécificité. Nous faisons donc face à deux principaux défis : comment quantifier la spécificité de liaison des protéines ? Et quelle est la nature de la relation liant l'affinité et la spécificité de liaison des protéines ?

De manière à y répondre, il s'agit tout d'abord de mettre au point une méthode visant à mesurer la spécificité. Ici, nous proposons un procédé expérimental *in vitro* ainsi qu'une approche statistique permettant une quantification systématique de la spécificité de liaison des protéines. Nous avons mis au point une banque de molécules d'ADN simple brins (utilisées comme ligands) avec une structure en tige-boucle (la diversité de la séquence étant localisée dans la boucle longue de 7 nucléotides). Nous avons sélectionné cette banque par SELEX contre un ensemble d'anticorps recombinants. Ils sont composés de deux parties principales : « l'ossature », constituée de 100 acides aminés (provenant d'anticorps naturels « naïfs » n'ayant jamais maturé contre un antigène ou bien d'anticorps « évolués » ayant maturé contre le VIH) ; et la zone de liaison constituée de 4 acides aminés correspondant au CDR3 (région déterminant la complémentarité) des anticorps naturels (cette région est connue pour son implication dans l'affinité et la spécificité des anticorps pour leurs antigènes). Différentes séquences de CDR3 ont été préalablement choisies pour leurs hautes ou faibles affinités pour une molécule d'ADN particulière. Pour chaque expérience de sélection, nous avons utilisé le séquençage à haut débit afin d'identifier les séquences de molécules d'ADN les plus enrichies durant la sélection (elles correspondent aux molécules d'ADN avec la plus haute affinité pour l'anticorps) ainsi que les molécules d'ADN les moins enrichies.

Nous proposons donc d'étudier la spécificité globale des anticorps en explorant une sous-région de l'espace des ligands où les molécules cibles sont des acides nucléiques. En utilisant les expériences de SELEX, nous pouvons qualitativement comparer les affinités moyennes de différents anticorps pour la même banque d'ADN en estimant la quantité de molécules d'ADN qui ont interagit avec l'anticorps. Nous pouvons également comparer quantitativement des constantes de dissociations mesurées pour un sous-échantillon de molécules d'ADN par ELISA. En particulier, nous avons posé la question suivante : pouvons-nous mesurer différentes affinités moyennes pour des anticorps ayant différentes ossatures (naïve ou évoluée) et différents CDR3 ?

Nous avons répondu à cette question en deux temps : premièrement, nous avons vérifié si nous pouvions mesurer différentes affinités moyennes pour des anticorps naïfs ou évolués, puis nous avons vérifié si des anticorps partageant la même ossature mais ayant différents CDR3 avaient les mêmes affinités ou non pour la banque d'ADN.

Nous montrons que : (1) des anticorps avec différentes ossatures présentent différentes affinités de liaison pour la banque d'ADN, en particulier, les anticorps naïfs ont une affinité moyenne plus élevée que les anticorps évolués ; (2) des anticorps partageant la même ossature mais ayant différents CDR3 ont également des affinités moyennes différentes pour la banque

d'ADN ; (3) L'ossature ainsi que la séquence du CDR3 contrôlent la spécificité globale des anticorps, à différents niveaux.

De-même, nous proposons d'étudier la spécificité locale des anticorps en mesurant leur capacité à discriminer des séquences d'ADN. En utilisant les expériences de SELEX ainsi que le séquençage à haut-débit, nous pouvons comparer la diversité des séquences d'ADN sélectionnées par différents anticorps.

Nous avons posé la question suivante : est-ce que des anticorps avec différentes ossatures (naïve ou évoluée) et différents CDR3 discriminent la banque de molécules d'ADN de la même manière ?

Nous avons répondu à cette question en 3 étapes : premièrement nous avons vérifié si les anticorps recombinants sont capables de discriminer différentes séquences d'ADN ; nous avons ensuite comparé les séquences des molécules d'ADN sélectionnées par des anticorps partageant la même ossature mais ayant différents CDR3, puis nous avons comparé les séquences des molécules d'ADN sélectionnées par des anticorps ayant différentes ossatures.

Nous montrons que : (1) les anticorps recombinants sont capables de différencier des molécules d'ADN ayant différentes boucles de 7 nucléotides parmi une banque de 32768 molécules ; (2) les anticorps reconnaissent des motifs de 4 nucléotides à l'intérieur des boucles d'ADN ; (3) les anticorps naïfs que nous avons testés sélectionnent les mêmes boucles d'ADN tandis que les anticorps évolués sélectionnent différentes boucles d'ADN ; (4) le degré de maturation des anticorps affecte la diversité des boucles d'ADN qu'ils sélectionnent – en particulier, les anticorps naïfs sélectionnent des boucles d'ADN plus variées que les anticorps évolués ; (5) l'étude de la sélection des sous-motifs de 4 nucléotides nous permet d'inclure des données bruitées dans notre analyse.

Nous proposons ensuite une nouvelle approche permettant de mesurer plusieurs centaines de constantes de dissociations en une seule expérience de sélection. Nous avons développé ce procédé expérimental en utilisant un seul anticorps et une banque minimaliste de 218 molécules d'ADN.

Le principe est le suivant : 7 expériences de SELEX sont réalisées en parallèle avec différentes concentrations de banque d'ADN et une concentration fixe d'anticorps. Les mesures d'enrichissement de chaque molécule d'ADN pour chaque concentration de la banque sont utilisées pour construire 218 courbes de liaison. À partir de chacune de ces courbes, une constante de dissociation peut être inférée (correspondant à l'affinité de l'anticorps pour la molécule d'ADN correspondante) en utilisant le modèle de sélection approprié.

Nous montrons que : (1) l'expérience de calibration permet de générer 218 courbes de liaison en utilisant simplement 7 puits (d'une plaque 96 puits) et 2 jours d'expériences contre environ 1500 puits et 1 mois d'expérience avec l'approche standard par ELISA ; (2) le logarithme des enrichissements est linéairement corrélée avec les énergies de liaison pour une concentration optimale d'ADN et à condition que les valeurs extrêmes des énergies de liaison soient suffisamment proches l'une de l'autre ; (3) le modèle théorique de sélection doit inclure

différentes sources de biais afin de décrire les données mesurées expérimentalement ; (4) les données expérimentales et simulées suggèrent un compromis entre la diversité de la banque d'ADN et la résolution de l'expérience de sélection.

En résumé, l'utilisation d'une banque de molécules d'ADN comme un outil de mesure de spécificités globale et locale nous a permis d'identifier des différences de spécificités entre des anticorps recombinants naïfs ou maturés contre une cible biologique et ayant différents CDR3. En particulier, nous avons constaté que les anticorps naïfs ont une affinité moyenne pour la banque de molécules d'ADN plus élevée, et qu'ils sélectionnent avec moins de précision les séquences d'ADN par rapport aux anticorps maturés. Nous avons en parallèle développé une approche plus quantitative de la mesure de spécificité qui nous permettrait, à long terme, de systématiquement relier les spécificités globale et locale à des mesures robustes et à haut-débit d'affinités. Ces résultats ouvrent la voie vers une étude systématique de la relation entre affinité et spécificité.

Les résultats préliminaires de mesures de spécificité peuvent être mis en perspective avec une récente étude de « l'évoluabilité » (décrivant la capacité à évoluer, c'est-à-dire à sélectionner des phénotypes plus favorables sous l'effet de mutations aléatoires) de ces mêmes anticorps. Cette étude a démontré que différentes ossatures étaient responsables de différents niveaux d'évoluabilité tout comme nous avons montré que différentes ossatures étaient responsables de différents niveaux de spécificités.