Structural basis for recognition of the FLAG-tag by anti-FLAG M2

J. Wouter Beugelink[1], Els Sweep[2], Bert J.C. Janssen[1], Joost Snijder*[2], Matti F. Pronker*[2]

1: Structural Biochemistry, Bijvoet Centre for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands

2: Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

*corresponding authors: j.snijder@uu.nl, m.f.pronker@uu.nl

Abstract:

The FLAG-tag/anti-FLAG system is a widely used biochemical tool for protein detection and purification. Anti-FLAG M2 is the most popular antibody against the FLAG-tag, due to its ease of use, versatility, and availability in pure form or as bead conjugate. M2 binds N-terminal, C-terminal and internal FLAG-tags and binding is calcium-independent, but the molecular basis for the FLAG-tag specificity and recognition remains unresolved.

Here we present an atomic resolution (1.17 Å) structure of the FLAG peptide in complex with the Fab of anti-FLAG M2, revealing key binding determinants. Five of the eight FLAG peptide residues form direct interactions with paratope residues. The FLAG peptide adopts a 3_{10} helix conformation in complex with the Fab. These structural insights allowed us to rationally introduce point mutations on both the peptide and antibody side. We tested these by surface plasmon resonance, leading us to propose a shorter yet equally binding version of the FLAG-tag for the M2 antibody.

Introduction:

The FLAG-tag/anti-FLAG system is a widely used biochemical tool for protein purification, (co-)immunoprecipitation (IP), Western Blot, immunocytochemistry, chromatin IP, immunohistochemistry and other (*e.g.* chemical biology) applications[1]. The FLAG-tag, consisting of the DYKDDDDK peptide sequence, was originally designed to be a short hydrophilic purification handle with an internal protease cleavage site (Enterokinase, also known as Enteropeptidase/TMPRSS15; -DDDDK-) for protein identification and purification[2]. The short hydrophilic nature of the FLAG peptide and the high specificity and affinity of available antibodies, combined with the possibility of gentle elution with a synthetic FLAG peptide (or, alternatively, low pH or proteolytic release), make it suitable for many applications[1,3,4].

The original anti-FLAG antibody M1 (also known as 4E11) binds calcium-dependent and only to FLAG-tags at the very N-terminus of the (mature) protein[5]. Later iterations of anti-FLAG antibodies such as M2, M5, L5 and 2H8 did not suffer from these limitations, binding calcium-independent to FLAG-tags on both N- and C-termini, or directly following a starter methionine (*e.g.* for cytosolic protein), as well as internal tags (*e.g.* embedded in a flexible linker between two domains)[5–7]. The mouse monoclonal anti-FLAG M2 in particular is widely used due to the availability of a hybridoma cell line[5], and commercial availability of purified anti-FLAG M2 as free IgG and pre-coupled affinity resins. Due to its qualities and widespread popularity, the FLAG-tag/anti-FLAG system has been the subject of intense optimization, both from the peptide and antibody side[6–12].

Previously, structures of other widely used antibody/peptide tag complexes have been described, such as for the anti-Influenza Hemaglutinin (HA)/HA-tag[13,14], anti-cMyc/cMyc-tag[15] and anti-His/His-tag[16]. Despite the wide use of the anti-FLAG-M2 antibody, neither its sequence nor the structural basis for its specific interaction with the FLAG-tag have been publicly available, hampering its application in genetically engineered affinity reagents or structure-based methods to improve binding. We recently used proteomic sequencing in combination with a previously determined incomplete high-resolution crystal structure[17] to determine the protein sequence of the anti-FLAG M2 heavy and light chains, and incorporated these sequences into mammalian expression vectors for recombinant production[18]. We validated recombinantly expressed anti-FLAG M2 using these plasmids by Western blot, yielding indistinguishable results compared to the commercially available antibody[18].

Here, we present a high-resolution structure of the anti-FLAG M2 Fab in complex with the FLAG peptide. This provides us with key residue-specific binding determinants, suggesting possible modifications. Site-directed mutagenesis of the anti-FLAG M2 and the FLAG peptide was combined

with surface plasmon resonance (SPR) to test different variants of both, showing that a shorter version of the FLAG-tag is possible without impeding the affinity.

Results:

Structure determination:

To create a Fab version of our previously described recombinant anti-FLAG M2[18], the heavy chain construct was truncated in between the C_H1 domain and the hinge region, keeping the C-terminal -Ala_3His_8 tag for purification purposes. This construct was co-expressed with the original light chain construct in Expi293 cells, after which the Fab was purified from the cell supernatant using Ni-affinity followed by size exclusion chromatography (see Methods for details). Crystallization screens were set up after mixing the Fab with synthetically produced FLAG peptide, yielding several crystallization hits. Our best crystal diffracted anisotropically to a maximum resolution of 1.17 Å, in the same crystal form as the original *apo* anti-FLAG M2 Fab structure[17] (table 1). Phasing by molecular replacement readily revealed strong F_o-F_c difference density near the paratope region (figure 1), confirming the complex was formed in the crystal, which allowed modelling the first 6 residues of the FLAG peptide (DYKDDD, supp. video 1).

Anti-FLAG M2 binds to the FLAG-tag with both chains:

As anticipated for a highly charged and hydrophilic epitope, most of the interactions formed between the FLAG peptide and anti-FLAG M2 are either through hydrogen bonds, salt bridges or ionic interactions (figure 2, table 2). Five out of the eight FLAG peptide residues (Asp1, Tyr2, Lys3, Asp4 and Asp6) appear to contribute directly to the interaction with M2. FLAG peptide residues Tyr2, Lys3 and Asp4 appear to form crucial interactions, predominantly involving M2 residues heavy chain Glu99 and light chain Arg32 (figure 2, table 2). Tyr2 and Lys3 protrude into the hole between the heavy and light chain variable domains, where they are stabilized by both hydrophilic and hydrophobic interactions. Asp1 and Asp6 form direct salt bridges with light chain paratope residues His31 and Lys55, respectively (figure 2, table 2). The backbone carbonyl of Lys3 forms a hydrogen bond with the side chain carboxamide of light chain Asn33. Similarly, the backbone carbonyl of Asp6 forms a hydrogen bond with the carboxamide side chain of light chain Asn35. FLAG residue Asp5 is oriented away from the M2 binding site and residues Asp7 and Lys8 are not resolved in the electron density, indicating that these residues do not contribute directly to interactions.

Marked conformational changes in the paratope of M2 are apparent comparing the *apo* and FLAG-bound structures (figure 3, supp. video 2). Light chain Arg32 wraps around the FLAG peptide to form a

88 double salt bridge with Asp4. Other paratope residues that substantially change their conformation
89 upon binding are Glu99 and Phe101 of the heavy chain CDR3 loop, both of which form stabilizing
90 interactions with the buried pair of side chains of Tyr2 and Lys3 of the FLAG sequence in the central
91 cavity between heavy and light chains (figure 2). Interestingly, both light chain Arg32 and heavy chain
92 Glu99 and Phe101 are all mutated compared to mouse germline ancestral sequences[19,20], likely the
93 result of somatic hypermutation. A subtle general compaction of the N-terminal (variable) Ig domains
94 of the heavy and light chains is also observed compared to the *apo* structure (figure 3, supp video 2).

95 Another feature observed is that the N-terminal glutamine of the heavy chain forms a pyroglutamate,
96 a phenomenon more commonly observed for antibodies[21,22] (supplementary figure 1). This
97 however does not appear to affect the binding to the FLAG sequence, since the heavy chain N-terminus
98 is distal from the FLAG peptide binding site (25 Å). Several ordered water molecules as well as two
99 chloride ions were observed neighbouring the FLAG peptide binding site (figure 2, supplementary
100 figure 2). Three of these water molecules simultaneously form hydrogen bonds with both paratope
101 and FLAG peptide, mediating indirect interactions between residues FLAG Asp1 and light chain Tyr101,
102 FLAG Tyr2 and heavy chain Tyr50 and between FLAG Asp6 and light chain Asn35.

103 **The FLAG peptide adopts a 3_{10} helix conformation:**

104 Unexpectedly, we observed that the FLAG peptide adopts a 3_{10} helix conformation with two full turns
105 (figure 4c). Whereas in α-helices backbone hydrogen bonds are formed between residues *i* and *i*+4,
106 we see the typical tighter wound 3_{10} helix backbone hydrogen bonding interactions of residue *i* and
107 *i*+3 for the pairs of Asp1-Asp4, Tyr2-Asp5 and Lys3-Asp6 in the FLAG peptide. While a 3_{10} helix is
108 typically energetically less favourable than a regular α-helix, it appears that the FLAG sequence was
109 inadvertently designed with a preference for 3_{10} helix formation[23], with aspartate residues in
110 position 1 and 4. Indeed, both these aspartates form stabilising interactions with their carboxylate side
111 chains to the backbone nitrogen atoms of the N-terminus (Asp4) of the peptide and the backbone
112 amide of Lys3 (Asp1). This suggests that, while not essential for anti-FLAG M2, having the FLAG-tag at
113 the very N-terminus of the protein (without an initiator methionine) might still be beneficial in
114 lowering the energetic penalty for forming a 3_{10} helix, compared to an internal or C-terminal FLAG-tag.
115 Since the N-terminal amino group carries a formal charge at neutral pH and is expected to form a
116 stronger interaction with the carboxylic acid of the side chain of Asp4 compared to an amide nitrogen
117 for an internal or C-terminal tag, it could favour the 3_{10} helix conformation observed in our complex
118 structure.

119

Introducing an extra salt bridge between FLAG-tag Asp5 and heavy chain Lys100:

Surface plasmon resonance to determine the affinity of wildtype and variant FLAG sequences:

Since the side chain of Asp5 is not involved in direct interactions with the antibody paratope, we considered this position to be amenable to mutation with the goal of enhancing the binding affinity. The side chain of heavy chain Lys100 in CDR3 is close to the Asp5 side chain carboxylic acid, but still too far for a direct salt bridge in our structure (6.5 Å, figure 5a). We reasoned that mutating either FLAG-tag Asp5 to glutamate or heavy chain Lys100 to arginine, or combining those two mutations would bring their side chains in close enough proximity in the complex to form an extra stabilizing salt bridge, thereby enhancing the binding affinity.

A series of N-terminally biotinylated synthetic FLAG peptide variants were immobilised to streptavidin-coated SPR chips to determine the binding affinity of anti-FLAG M2. C-terminally truncated variants, as well as two variants with a glutamate in place of Asp5 of the FLAG sequence were tested (DYKDEDDK and DYKDED, table 3). Anti-FLAG M2 Fab and full IgG were used in the mobile phase; the Fab was used to obtain 1:1 binding following a Langmuir isotherm model similar to purification and (co-)IP applications (for monomeric target proteins), the full IgG to quantify avidity-enhanced binding more akin to usage of the antibody in Western blot, immunocytochemistry and immunohistochemistry.

As expected, the apparent affinities were orders of magnitude higher for the full IgG (K_D of 2.95 ± 0.31 nM and 306 ± 8.9 nM for the wildtype peptide, table 3, figure 5, supplementary figures 3 and 4), where a two-site binding model was used to fit the data, compared to the Fab (K_D of 1.91 ± 0.24 µM, one-site binding model). However, the trends between the preference for different FLAG-tag variants were highly similar between the Fab and full IgG data. Truncating the original FLAG sequence DYKDDDDK with one residue from the C-terminal side did not affect the binding affinity, but shorter truncations (in particular the extra short 4- and 5- residue variants DYKD and DYKDD) did dramatically suffer in their binding affinity (52.9 ± 22.1 nM and 154 ± 100 nM, respectively).

The Asp5Glu mutation did appear to subtly benefit binding for the full wt anti-FLAG M2 IgG, with a K_D of 2.95 ± 0.31 nM and 2.23 ± 0.25 nM for the wt and mutated peptide, respectively (p=0.01, table 3 and figure 5). However, no significant difference was observed when comparing a truncated peptide, which measured 9.72 ± 2.6 nM and 10.2 ± 0.48 nM for the wt and mutant peptide, nor did the wt anti-FLAG M2 Fab show a significant difference in binding affinity between wt and Asp5Glu FLAG peptide (1.91 ± 0.24 µM and 1.78 ± 0.13 µM, respectively; p=0.19).

When using the heavy chain Lys100Arg mutated Fab or IgG in the mobile phase, the affinities dropped by about 20-fold for both the original FLAG sequence and the other variants we tested (table 3, figure

152 5, supplementary figure 3 and 4). Possibly, the Lys100Arg mutation locally changes the fold and/or
153 surface properties of the antibody in a way that is disadvantageous for binding the FLAG peptide
154 (either wildtype or Asp5Glu mutant).

155

156 **Discussion:**

157 Our high-resolution structure of the FLAG/anti-FLAG M2 Fab complex reveals the structural
158 determinants of the interaction at the basis of this widely used biochemical tool. It also represents the
159 highest resolution structure of an antibody-antigen complex in the protein databank to date.

160 We only observed well-resolved electron density for the first six out of eight residues of the FLAG
161 peptide (density observed for DYKDDD compared to the full sequence DYKDDDDK), prompting us to
162 determine binding affinities for shorter truncated versions of the FLAG sequence. While truncating the
163 FLAG peptide by deleting the C-terminal lysine residue did not appear to impair binding, shorter
164 truncations did suffer in terms of binding affinity. This is particularly surprising for the six-residue
165 truncated version (DYKDDD), since no density is observed for Asp7, suggesting it does not directly
166 contribute to the binding affinity. Possibly, having the C-terminus at position six (DYKDDDDK) interferes
167 with 3_{10} helix formation or complex formation by having an extra negative charge so close to the
168 epitope/paratope interaction. Of note, while a shorter (*e.g*. 6- or 7-residue) C-terminally truncated
169 version of the FLAG-tag might be beneficial for certain applications, one has to bear in mind that by
170 truncating any C-terminal residue from the original FLAG-tag sequence, the
171 enterokinase/enteropeptidase/TMPRSS15 proteolytic cleavage motif (DDDDK) will be lost.

172 Now that we determined the structural basis of the FLAG/anti-FLAG M2 interaction, it is interesting to
173 revisit previous empirical attempts to determine the binding determinants of this interaction[9,10].
174 Indeed, these studies also found that residues 5, 7 and 8 of the FLAG sequence appear to contribute
175 little to the affinity for anti-FLAG M2. Substituting Asp4 and Asp6 with glutamate still permitted
176 binding[10], which could be explained by these residues still having a carboxylic acid side chain which
177 can still form similar salt bridges to the original aspartate residues, if some flexibility in the peptide or
178 paratope would permit the accommodation of the additional aliphatic $-C_{\gamma}H_2-$. According to a high-
179 throughput phage display screen[9], the main binding determinants in the FLAG sequence are Tyr2,
180 Lys3 and Asp6, similar to what we conclude based on the structure. Asp4 and, in particular, Asp1 were
181 also enriched, whereas position 5, 7 and 8 were completely random, in agreement with their lack of
182 contribution to the binding interface in our structure.

183 In conclusion, our data provide a structural framework for understanding the interaction of the FLAG-
184 tag with its most used antibody, anti-FLAG M2, in agreement with previous empirical studies. With the
185 anti-FLAG M2 sequence now publicly available and its interactions with the FLAG-tag defined with
186 atomic detail, the stage is set for further structure-based optimisation of FLAG-tag based affinity
187 reagents.

188

189 **Materials and Methods:**

190 Construct design:

191 The heavy chain Fab construct for anti-FLAG M2 was generated by deletion PCR and in vivo assembly
192 (IVA) cloning[24], truncating after Gly219 using the following primers:
193 gcggccgcaCATCACCACCATCATCACCATCATTGATAAC (forward) and
194 GTGATGTGCGGCCGCGCCACAGTCGCGCGGCAC (reverse), to ensure an in-frame NotI site (translates to
195 three alanines with an extra adenosine base) followed by the octahistidine tag for purification, both
196 already present in the original full IgG anti-FLAG M2 heavy chain plasmid[18].

197 The K100R mutation was introduced into the heavy chain by site-directed mutagenesis using IVA
198 cloning, using the following primers: CTATTGTGCGCGAGAGagaTTCTATGGTTACGATTATTGGGGCCAAG
199 (forward) and tctCTCTCGCGCACAATAGTAGACAGCACTATCC (reverse). The Fab of this mutant heavy
200 chain construct was generated in the same way as the wildtype version (see above).

201 Protein expression:

202 The IgG and Fab constructs were transiently expressed by mixing two plasmids encoding for the heavy
203 and light chain in a 1:1 (m/m) ratio and transfecting this mixture into Expi293 cells. To this end, plasmid
204 mixture at 30 µg/mL and Polyethylenimine "Max" (PEI MAX®) at 90 µg/mL were combined with the
205 Expi293 cells at $3*10^6$ viable cells/mL in a 1:1:28 (v/v/v) ratio. First the plasmid mixture was dropwise
206 added to the PEI MAX®, and after 30 minutes of incubation this new mixture was dropwise added to
207 the cells. The final concentrations of plasmid and PEI MAX® were thus 1 and 3 µg/mL, respectively.
208 After 6 days the proteins were harvested by spinning down the culture twice for 10 min at 4000 rpm
209 and subsequently filtering the medium using a 0.22 µm filter.

210 Protein purification:

211 Full IgG and Fab versions of anti-FLAG M2 were purified from the Expi293 cell supernatant by a
212 combination of immobilized metal affinity chromatography (IMAC) followed by size exclusion
213 chromatography (SEC). Filtered cell supernatant was loaded onto a 5 mL HisTrap-excel column (Cytiva)
214 equilibrated with IMAC A buffer (500 mM NaCl, 25 mM HEPES pH 7.8) at a flowrate of 5 ml/min. The

215 column was washed with 200-300 mL of either 3% (v/v, Fab) or 5% (v/v, IgG) IMAC B (500 mM NaCl,
216 500 mM imidazole, 25 mM HEPES pH 7.8) in IMAC A, until the UV A280 signal reached a stable baseline.
217 Protein was eluted with 40% IMAC B in IMAC A (wildtype Fab) or 100% IMAC B (K100R Fab and both
218 IgG's). The IMAC eluates were concentrated to 1-1.5 ml using 10 kDa MWCO concentrators (Amicon),
219 before injection onto a HiLoad 16/600 superdex200 SEC column (Cytiva) equilibrated with SEC buffer
220 (150 mM NaCl, 20 mM HEPES pH 7.5). Fab and IgG peak fractions were pooled and concentrated with
221 10 kDa MWCO concentrators (Amicon) to a concentration of 2-6 mg/ml.

222 Crystallization:

223 Fab-FLAG peptide complexes were mixed in a 1:3.5 molar ratio at a protein concentration of 5.6 mg/ml.
224 This was used to set up sitting drop vapour diffusion crystallization screens, mixing 150 nL
225 protein/peptide complex sample with 150 nL reservoir solution, at 293 K. Our best diffracting crystal
226 grew in a condition of 0.1 M Tris-HCl pH 8, 20% (w/v) polyethylene glycol 6000, 0.2 M NH_4Cl. Crystals
227 were cryo-protected with reservoir solution supplemented with 25% (v/v) glycerol before plunge-
228 freezing in liquid nitrogen.

229 Data collection and structure determination:

230 Data was collected at Diamond Light Source beamline I24, equipped with a CdTe Eiger2 9M detector,
231 at a wavelength of 0.6199 Å. Due to the anisotropic nature of the data, three datasets collected on the
232 same crystal were integrated and merged using the multi-autoPROC+STARANISO[25] pipeline, as
233 integrated in Diamond ISPyB[26]. Data was then imported into CCP4i2[27]. Given that the
234 crystallization condition and determined unit cell closely resembled that of the *apo* structure (PDB
235 2G60[17]/7BG1[18]), FreeR flags were copied from this dataset and extended to the higher attained
236 resolution. The *apo* anti-FLAG M2 Fab structure (PDB 7BG1)[18] was used for molecular replacement
237 using PHASER[28]. Density for the FLAG-tag peptide was observed, and the peptide was manually built,
238 after adjusting CDR loops where necessary, in COOT[29]. The structure was then refined using iterative
239 rounds of manual adjustment in COOT and refinement in REFMAC5[30], with the setting "VDWR" set
240 to 2.0. Quality of the geometry was analysed using MolProbity[31]. All programs were used as
241 implemented in CCP4i2 v1.1.0[27].

242 Surface plasmon resonance:

243 N-terminally biotinylated synthetic FLAG peptide variants were ordered from Genscript. These were
244 dissolved in SPR buffer (150 mM NaCl, 20 mM HEPES pH 7.4, 0.005% (v/v) Tween20) by rotating
245 overnight at room temperature, followed by pH adjustment with NaOH to pH 7-8 and sonication.
246 Peptides were printed on a streptavidin-coated SPR chip (P-Strep for full IgG, G-Strep for Fab, Sens B.V.)
247 using a continuous flow microfluidics spotter (Wasatch) and flowing for 1 hour at RT, after which the

chips were washed with SPR buffer for 10 minutes and subsequently quenched with 10 mM biotin in SPR buffer. SPR experiments were performed using an IBIS-MX96 system (IBIS technologies) at 298 K with SPR buffer as the running buffer. Chips were always washed overnight with SPR buffer to have a stable baseline. Analytes were then injected in 2× dilution series in SPR buffer, measuring from low to high concentrations, at a constant temperature of 298 K. Data were analysed using *SPRINTX* (IBIS technologies), extracted in Scrubber2 (Biologic Software), and fit to a one-site or two-site specific binding model in GraphPad Prism (Dotmatics). In short, background signal from adjacent regions without ligand bound were subtracted from signal in regions of interest, and signal was set to zero at the start of the injection series. Injections were then overlayed, and the average signals after reaching equilibrium were plotted against the analyte concentration and fitted with relevant Langmuir models.

Acknowledgements:

Diffraction experiments were performed at beamline I24 at the Diamond Light Source (DLS) synchrotron, Harwell, United Kingdom. We are grateful to local contacts at DLS for assistance in using beamline I24. This research was funded by the Dutch Research Council NWO Gravitation 2013 BOO, Institute for Chemical Immunology (ICI; 024.002.009) to J.S and NWO grant OCENW.KLEIN.026 to B.J.C.J.

Data availability:

The coordinates and structure factors of the crystal structure of the FLAG/anti-FLAG M2 complex have been deposited at the protein databank under accession code 8RMO.

Declaration of competing interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement:

J. Wouter Beugelink: Investigation, Methodology, Formal analysis, Data curation, Validation, Visualization, Project administration, Writing – review & editing. **Els Sweep:** Investigation, Methodology. Writing – review & editing. **Bert J.C. Janssen:** Formal analysis, Validation, Resources, Supervision, Writing – review & editing, Funding acquisition. **Joost Snijder:** Resources, Supervision, Writing – review & editing, Funding acquisition. **Matti F. Pronker:** Conceptualization, Investigation, Methodology, Formal analysis, Data curation, Validation, Visualization, Project administration, Writing – original draft, Writing – review & editing.

Tables:

Table 1: Crystallographic data collection and refinement parameters

Collection statistics	
Space group	$P2_12_12$
Unit cell dimensions (a, b, c; Å)	41.80, 68.43, 134.62
α, β, γ (°)	90.0, 90.0, 90.0
Wavelength (Å)	0.6199
Resolution limits (Å)	73.33 – 1.17 (1.24 – 1.17)
R_{merge}	0.151 (6.867)
R_{meas}	0.153 (7.008)
R_{pim}	0.026 (1.388)
Total no. of reflections	4558427 (163064)
No. of unique reflections	123223 (6161)
Mean I/σ	16.5 (1.5)
Completeness (ellipsoidal) (%)	95.7 (75.0)
Multiplicity	37.0 (26.5)
$CC_{1/2}$	0.999 (0.609)
Anisotropic parameters	
Principal axis	Diffraction limit (Å)
1, 0, 0; along a*	1.17
0, 1, 0; along b*	1.17
0, 0, 1; along c*	1.53
Refinement statistics	
R_{work} / R_{free}	0.157 / 0.185
Non-H atoms (no.)	7396
Protein residues (no. / Bfac; Å^2)	423 / 25.13
Ions (no. / Bfac; Å^2)	3 / 17.74
Water (no. / Bfac; Å^2)	517 / 36.2
Bond length r.m.s.d (Å)	0.011
Bond angle r.m.s.d (°)	1.72
Molprobity score	1.14
Clash score	2.32
Poor rotamers (%)	0.50
Ramachandran favoured (%)	97.32
Ramachandran allowed (%)	2.68
Ramachandran disallowed (%)	0.00

Table 2: Observed interactions (LC: light chain, HC: heavy chain)

Interacting residues:	**Type of interaction:**	**Mutated from germline?**
Asp1-His31 LC	Salt bridge	
Tyr2-Thr33 HC	Hydrophobic/-CH-π	
Tyr2-Glu99 HC	Hydrogen bond	V
Tyr2-His35 HC	Hydrogen bond	

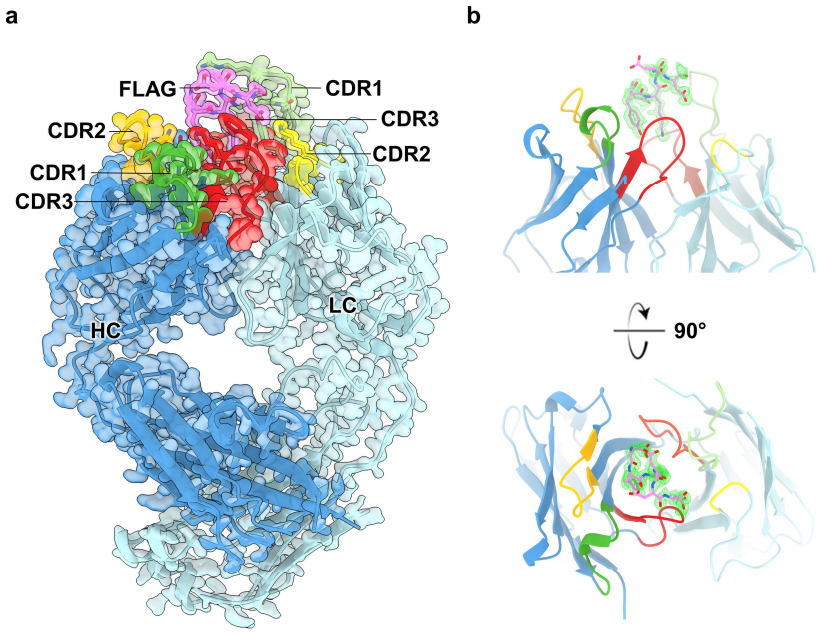
Lys3-Gly96 LC	Hydrogen bond (backbone carbonyl)	
Lys3-Glu99 HC	Ionic interaction	V
Lys3-Asn33 LC	Hydrogen bond (backbone carbonyl)	
Asp6-Asn35 LC	Hydrogen bond (backbone carbonyl)	
Asp4-Arg32 LC	Double salt bridge	V
Asp6-Lys55 LC	Salt bridge	
Asp6-Tyr37 LC	Salt bridge	

Table 3: Surface plasmon resonance data

	Anti-FLAG WT IgG		Anti-FLAG WT Fab	Anti-FLAG K100R IgG		Anti-FLAG K100R Fab
	High affinity (nM)	Low affinity (nM)	One site (µM)	High affinity (nM)	Low affinity (µM)	One site (µM)
GGSDYKDDDDK	2.95 ± 0.31	306 ± 8.9	1.91 ± 0.24	59.8 ± 7.1	4.30 ± 0.52	23.9 ± 3.0
GGSDYKDDDD	3.08 ± 0.40	289 ± 14	2.01 ± 0.43	61.3 ± 1.8	4.04 ± 6.0	25.3 ± 3.0
GGSDYKDDD	9.72 ± 2.6	512 ± 130	3.52 ± 0.72	197 ± 13	16.7 ± 2.8	118 ± 73
GGSDYKDD	785 ± 190	-	n.b.	5250 ± 1300	-	n.b.
GGSDYKD	288 ± 48	-	n.b.	1460 ± 220	-	n.b.
GGSDYKDEDDK	2.23 ± 0.25	274 ± 30	1.78 ± 0.13	45.9 ± 4.8	4.08 ± 0.20	18.3 ± 2.1
GGSDYKDED	10.2 ± 0.48	555 ± 32	3.54 ± 0.24	131 ± 15	6.99 ± 1.1	62.1 ± 29

289 **Figures:**

290 Figure 1: Structure of the anti-FLAG M2 Fab in complex with the FLAG peptide

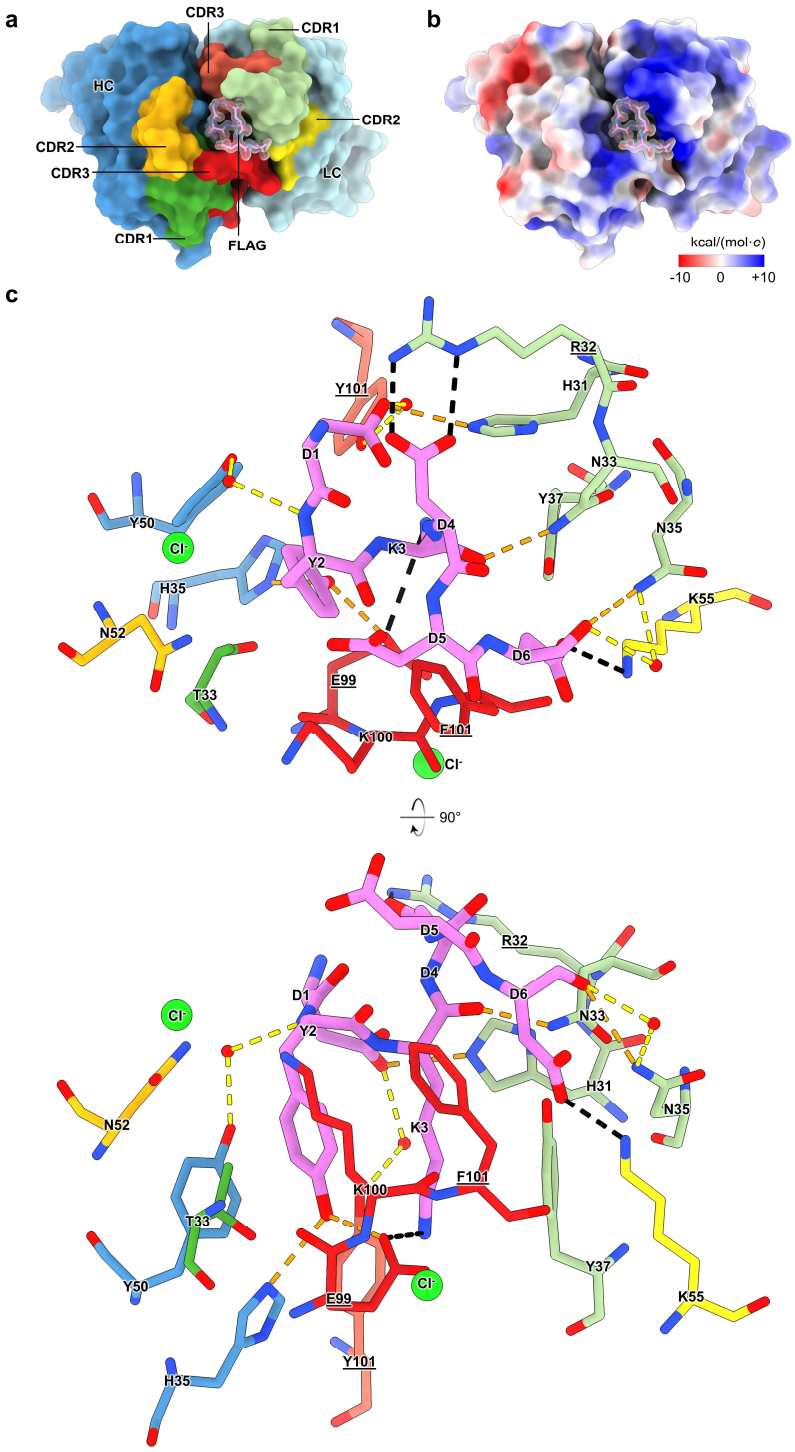


291 (A) Overview of the structure of the complex of the anti-FLAG M2 Fab and the FLAG peptide; CDR loops
292 are indicated in green (CDR1), yellow (CDR2) and red (CDR3). (B) mF_o-DF_c difference map at 3 r.m.s.d
293 based on the (complete) *apo* anti-FLAG M2 Fab structure (PDB 7BG1) reveals the density for the FLAG
294 peptide near the expected binding site.

295

296

Figure 2: An intricate network of mostly hydrophilic interactions stabilises the interaction of the FLAG peptide with the anti-FLAG M2 Fab.



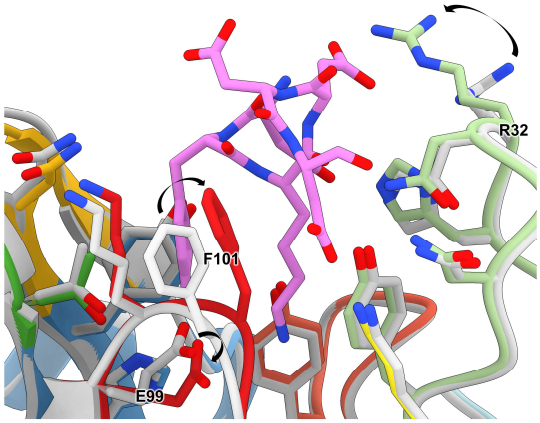
(A) Overview of the complex structure in surface (Fab) and sticks (FLAG peptide) representation, colouring as in figure 1 (heavy chain in dark blue, light chain in light blue, FLAG peptide in pink, CDR1 in green, CDR2 in yellow and CDR3 in red). (B) Electrostatic surface potential map of the Fab shows a mostly electropositive paratope, which would complement the net negatively charged FLAG peptide

304 (shown as pink sticks), same orientation as in figure 2A. (C) Direct hydrogen bonds between the FLAG
305 peptide and paratope residues are indicated as orange dashed lines, salt bridges *idem* in black, indirect
306 hydrogen bonds mediated by a single water molecule (red spheres) are shown as yellow dashed lines.
307 Paratope residues mutated from germline (most likely by somatic hypermutation) are underlined.
308 Protein and peptide are coloured as in figure 1 and 2A.

309

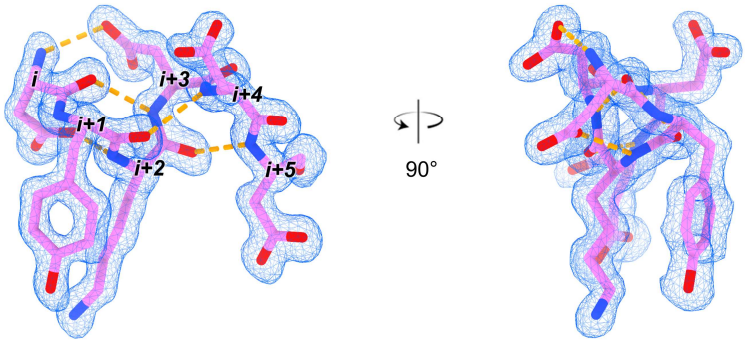
310

Figure 3: Conformational changes observed between the structure of the *apo* anti-FLAG M2 Fab and the complex with the FLAG peptide



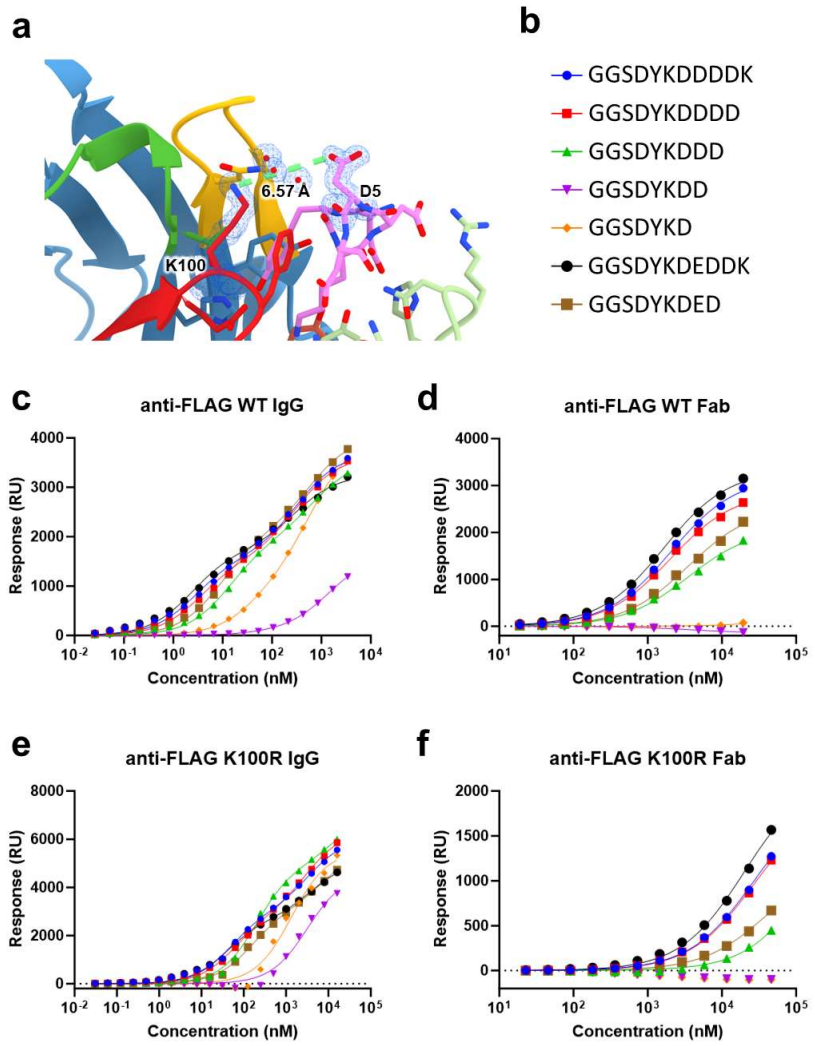
Overlay of the *apo* (grey) anti-FLAG M2 structure (PDB 7BG1) and the complexed structure with FLAG peptide presented here (coloured as in Figure 1).

Figure 4: The FLAG peptide adopts a 3_{10} helix conformation in complex with anti-FLAG M2.



Backbone hydrogen bonds within the FLAG peptide are indicated as orange dashed lines. 2mF_o-DF_c electron density at 1.3 r.m.s.d. shown as blue mesh.

Figure 5: Surface plasmon resonance quantifies binding affinities of the FLAG/anti-FLAG M2 interaction and variants.



(A) Location of Fab heavy chain K100 and FLAG peptide D5, displaying a distance of 6.6 Å. Selected part of 2mF_o-DF_c electron density at 1.3 r.m.s.d. shown as blue mesh. (B) Sequences of FLAG peptides used for affinity measurements. Coloured shapes correspond to curves in c-f. (C) Binding of anti-FLAG WT IgG to FLAG peptides. (D) Binding of anti-FLAG WT Fab to FLAG peptides. (E) Binding of anti-FLAG K100R IgG to FLAG peptides. (F) Binding of anti-FLAG K100R Fab to FLAG peptides.

333 **Supplementary figure legends:**

334 Supplementary figure 1: The N-terminus of the Fab heavy chain has a pyroglutamate modification.

335 2mF_o-DF_c electron density at 2.1 r.m.s.d. shown as blue mesh.

336 Supplementary figure 2: Two chloride ions and several water molecules are proximal to the FLAG
337 peptide binding site.

338 Hydrogen bonds formed by water molecules are indicated as blue dashed lines. Water and chloride
339 2mF_o-DF_c electron density at 0.9 r.m.s.d. shown as blue mesh.

340 Supplementary figure 3: Surface plasmon resonance sensograms show binding of antibodies to
341 peptides.

342 Response over time after injecting a concentration range of anti-FLAG WT IgG, anti-FLAG WT Fab, anti-
343 FLAG K100R IgG, and anti-FLAG K100R Fab as measured in a region of interest containing peptide with
344 the sequence (A) GGSDYDDDDK, (B) GGSDYDDDD, (C) GGSDYDDD, (D) GGSDYDD, (E) GGSDYD, (F)
345 GGSDYDEDDK, or (G) GGSDYDED. One replicate with a nominal spotting concentration of 200 nM is
346 shown per experiment. Values that were averaged and used for making response curves are indicated
347 with a red block.

348 Supplementary figure 4: Surface plasmon resonance response curves and fitting parameters used to
349 quantify binding affinities of the FLAG/anti-FLAG M2 interaction and variants.

350 Response curves for four technical replicates at nominal spotting concentrations of 500, 200, 50, and
351 20 nM, for injecting a concentration range of anti-FLAG WT IgG, anti-FLAG WT Fab, anti-FLAG K100R
352 IgG, and anti-FLAG K100R Fab as measured in a region of interest containing peptide with the sequence
353 (A) GGSDYDDDDK, (B) GGSDYDDDD, (C) GGSDYDDD, (D) GGSDYDD, (E) GGSDYD, (F) GGSDYDEDDK, or
354 (G) GGSDYDED. Inset tables for each graph show K_D and Bmax modelled for one or two binding modes
355 per curve.

356 **Supplementary video legends:**

357 Supplementary video 1:

358 Density and model for the structure of the complex of the anti-FLAG M2 Fab with the FLAG peptide

359 Supplementary video 2:

360 Conformational changes observed in the anti-FLAG M2 paratope upon binding of the FLAG peptide.

References:

[1] A. Einhauer, A. Jungbauer, The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins, J. Biochem. Biophys. Methods 49 (2001) 455–465. https://doi.org/10.1016/S0165-022X(01)00213-5.

[2] T.P. Hopp, K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D. Pat Cerretti, D.L. Urdal, P.J. Conlon, A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification, Bio/Technology 6 (1988) 1204–1210. https://doi.org/10.1038/nbt1088-1204.

[3] D.S. Waugh, Making the most of affinity tags, Trends Biotechnol. 23 (2005) 316–320. https://doi.org/10.1016/j.tibtech.2005.03.012.

[4] H. Nonaka, S. hei Fujishima, S. hei Uchinomiya, A. Ojida, I. Hamachi, FLAG-tag selective covalent protein labeling via a binding-induced acyl-transfer reaction, Bioorganic Med. Chem. Lett. 19 (2009) 6696–6699. https://doi.org/10.1016/j.bmcl.2009.09.122.

[5] B.L. Brizzard, R.G. Chubet, D.L. Vizard, Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution., Biotechniques 16 (1994) 730–5. http://www.ncbi.nlm.nih.gov/pubmed/8024796.

[6] F. Sasaki, T. Okuno, K. Saeki, L. Min, N. Onohara, H. Kato, T. Shimizu, T. Yokomizo, A high-affinity monoclonal antibody against the FLAG tag useful for G-protein-coupled receptor study, Anal. Biochem. 425 (2012) 157–165. https://doi.org/10.1016/j.ab.2012.03.014.

[7] S.H. Park, C. Cheong, J. Idoyaga, J.Y. Kim, J.H. Choi, Y. Do, H. Lee, J.H. Jo, Y.S. Oh, W. Im, R.M. Steinman, C.G. Park, Generation and application of new rat monoclonal antibodies against synthetic FLAG and OLLAS tags for improved immunodetection, J. Immunol. Methods 331 (2008) 27–38. https://doi.org/10.1016/j.jim.2007.10.012.

[8] A. Knappik, A. Plückthun, An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments., Biotechniques 17 (1994) 754–61. http://www.ncbi.nlm.nih.gov/pubmed/7530459.

[9] R.M. Miceli, M.E. DeGraaf, H.D. Fischer, Two-stage selection of sequences from a random phage display library delineates both core residues and permitted structural range within an epitope, J. Immunol. Methods 167 (1994) 279–287. https://doi.org/10.1016/0022-1759(94)90097-3.

[10] J.W. Slootstra, Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5, Mol. Divers. 2 (1996) 156–164. https://doi.org/10.1007/BF01682203.

[11] K. Ikeda, T. Koga, F. Sasaki, A. Ueno, K. Saeki, T. Okuno, T. Yokomizo, Generation and characterization of a human-mouse chimeric high-affinity antibody that detects the DYKDDDDK FLAG peptide, Biochem. Biophys. Res. Commun. 486 (2017) 1077–1082. https://doi.org/10.1016/j.bbrc.2017.03.165.

[12] K.C. Entzminger, J.M. Hyun, R.J. Pantazes, A.C. Patterson-Orazem, A.N. Qerqez, Z.P. Frye, R.A. Hughes, A.D. Ellington, R.L. Lieberman, C.D. Maranas, J.A. Maynard, De novo design of antibody complementarity determining regions binding a FLAG tetra-peptide, Sci. Rep. 7 (2017) 1–11. https://doi.org/10.1038/s41598-017-10737-9.

[13] J.M. Rini, U. Schulze-Gahmen, I.A. Wilson, Structural Evidence for Induced Fit as a Mechanism for Antibody-Antigen Recognition, Science (80-.). 255 (1992) 959–965. https://doi.org/10.1126/science.1546293.

[14] M.E.A. Churchill, E.A. Stura, C. Pinilla, J.R. Appel, R.A. Houghten, D.H. Kono, R.S. Balderas, G.G. Fieser, U. Schulze-Gahmen, I.A. Wilson, Crystal structure of a peptide complex of anti-influenza peptide antibody Fab 26/9: Comparison of two different antibodies bound to the same peptide antigen, J. Mol. Biol. 241 (1994) 534–556. https://doi.org/10.1006/jmbi.1994.1530.

[15] N. Krauß, H. Wessner, K. Welfle, H. Welfle, C. Scholz, M. Seifert, K. Zubow, J. Aÿ, M. Hahn, P. Scheerer, A. Skerra, W. Höhne, The structure of the anti-c-myc antibody 9E10 Fab

412 fragment/epitope peptide complex reveals a novel binding mode dominated by the heavy
413 chain hypervariable loops, Proteins Struct. Funct. Genet. 73 (2008) 552–565.
414 https://doi.org/10.1002/prot.22080.
415 [16] M. Kaufmann, P. Lindner, A. Honegger, K. Blank, M. Tschopp, G. Capitani, A. Plückthun, M.G.
416 Grütter, Crystal structure of the anti-His tag antibody 3D5 single-chain fragment complexed
417 to its antigen, J. Mol. Biol. 318 (2002) 135–147. https://doi.org/10.1016/S0022-
418 2836(02)00038-4.
419 [17] T.P. Roosild, S. Castronovo, S. Choe, Structure of anti-FLAG M2 Fab domain and its use in the
420 stabilization of engineered membrane proteins, Acta Crystallogr. Sect. F Struct. Biol. Cryst.
421 Commun. 62 (2006) 835–839. https://doi.org/10.1107/S1744309106029125.
422 [18] W. Peng, M.F. Pronker, J. Snijder, Mass Spectrometry-Based De Novo Sequencing of
423 Monoclonal Antibodies Using Multiple Proteases and a Dual Fragmentation Scheme, J.
424 Proteome Res. 20 (2021) 3559–3566. https://doi.org/10.1021/acs.jproteome.1c00169.
425 [19] F. Ehrenmann, M.-P. Lefranc, IMGT/DomainGapAlign: IMGT standardized analysis of amino
426 acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF)., Cold
427 Spring Harb. Protoc. 2011 (2011) 737–49. https://doi.org/10.1101/pdb.prot5636.
428 [20] F. Ehrenmann, Q. Kaas, M.-P. Lefranc, IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a
429 database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and
430 MhcSF., Nucleic Acids Res. 38 (2010) D301-7. https://doi.org/10.1093/nar/gkp946.
431 [21] L.W. Dick, C. Kim, D. Qiu, K. Cheng, Determination of the origin of the N-terminal pyro-
432 glutamate variation in monoclonal antibodies using model peptides, Biotechnol. Bioeng. 97
433 (2007) 544–553. https://doi.org/10.1002/bit.21260.
434 [22] Y.D. Liu, A.M. Goetze, R.B. Bass, G.C. Flynn, N-terminal glutamate to pyroglutamate
435 conversion in vivo for human IgG2 antibodies, J. Biol. Chem. 286 (2011) 11211–11217.
436 https://doi.org/10.1074/jbc.M110.185041.
437 [23] M.E. Karpen, de H.P. L., K.E. Neet, Differences in the amino acid distributions of 310 -helixes
438 and alpha- helixes, Protein Sci. 1 (1992) 1333–1342.
439 [24] J. García-Nafría, J.F. Watson, I.H. Greger, IVA cloning: A single-tube universal cloning system
440 exploiting bacterial In Vivo Assembly., Sci. Rep. 6 (2016) 27459.
441 https://doi.org/10.1038/srep27459.
442 [25] I.J. Tickle, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C. Vonrhein, G. Bricogne, STARANISO
443 (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi), (2018).
444 http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi.
445 [26] S. Delagenière, P. Brenchereau, L. Launer, A.W. Ashton, R. Leal, S. Veyrier, J. Gabadinho, E.J.
446 Gordon, S.D. Jones, K.E. Levik, S.M. Mcsweeney, S. Monaco, M. Nanao, D. Spruce, O.
447 Svensson, M.A. Walsh, G.A. Leonard, ISPyB: An information management system for
448 synchrotron macromolecular crystallography, Bioinformatics 27 (2011) 3186–3192.
449 https://doi.org/10.1093/bioinformatics/btr535.
450 [27] L. Potterton, J. Agirre, C. Ballard, K. Cowtan, E. Dodson, P.R. Evans, H.T. Jenkins, R. Keegan, E.
451 Krissinel, K. Stevenson, A. Lebedev, S.J. McNicholas, R.A. Nicholls, M. Noble, N.S. Pannu, C.
452 Roth, G. Sheldrick, P. Skubak, J. Turkenburg, V. Uski, F. Von Delft, D. Waterman, K. Wilson, M.
453 Winn, M. Wojdyr, CCP4i2: The new graphical user interface to the CCP4 program suite, Acta
454 Crystallogr. Sect. D Struct. Biol. 74 (2018) 68–84.
455 https://doi.org/10.1107/S2059798317016035.
456 [28] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser
457 crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
458 https://doi.org/10.1107/S0021889807021206.
459 [29] P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics, Acta Crystallogr.
460 Sect. D Biol. Crystallogr. 60 (2004) 2126–2132. https://doi.org/10.1107/S0907444904019158.
461 [30] G.N. Murshudov, P. Skubák, A. a. Lebedev, N.S. Pannu, R. a. Steiner, R. a. Nicholls, M.D. Winn,
462 F. Long, A. a. Vagin, REFMAC5 for the refinement of macromolecular crystal structures, Acta

Crystallogr. Sect. D Biol. Crystallogr. 67 (2011) 355–367. https://doi.org/10.1107/S0907444911001314.

[31] V.B. Chen, W.B. Arendall, J.J. Headd, D. a. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 12–21. https://doi.org/10.1107/S0907444909042073.