Electronic Supplementary Material (ESI) for Chemical Science.

Selectivity over coverage in *de novo* sequencing of IgGs

Maurits A. den Boer[1,2,#], Jean-Francois Greisch[1,2,#], Sem Tamara[1,2,#], Albert Bondt[1,2], Albert J.R. Heck[1,2,*]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute of Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

 # equal contributing first authors

** to whom correspondence should be addressed*

Supplementary Information

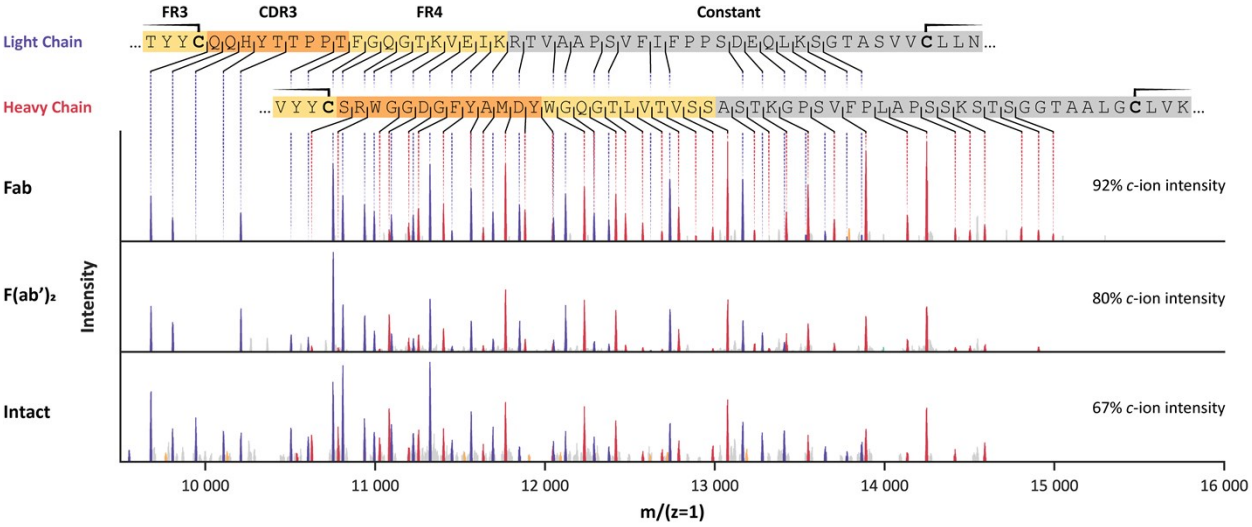

Figure S1. Color-annotated charge-deconvoluted ECD-MS spectra of the intact, F(ab')$_2$ and Fab molecules of Trastuzumab showing primarily c-ion ladders covering the CDR3s of the LC (purple) and HC (red). Fragment *z*-ions originating from the LC and HC are indicated in green and orange respectively. The protein sequences are indicated above with the cysteine residues involved in the disulfide bonds denoted with a solid black line. Ion intensity within the shown mass window that can be explained by c-ions is highest for the Fab preparation, while considerably lower for the intact IgG1.

Table S1. Some of the advantages and disadvantages of using native intact, F(ab')$_2$ or Fab molecules as precursor ions in the ECD-MS analysis.

Quality	Intact	F(ab')$_2$	Fab
Mass separation of *c*-ion fragments from precursor ions	Excellent	Excellent	Overlap with precursor ions
Spectral purity in terms of *c*-ions	Moderate due to interference from Fc *z*-ions	High	Highest
Ability for reproducible sample preparation	Excellent	Excellent	IgG1: Excellent IgG2-4: Poor due to use of nonspecific proteases
Sample homogeneity	Moderate due to glycosylation	Excellent	IgG1: Excellent IgG2-4: Poor due to use of nonspecific proteases

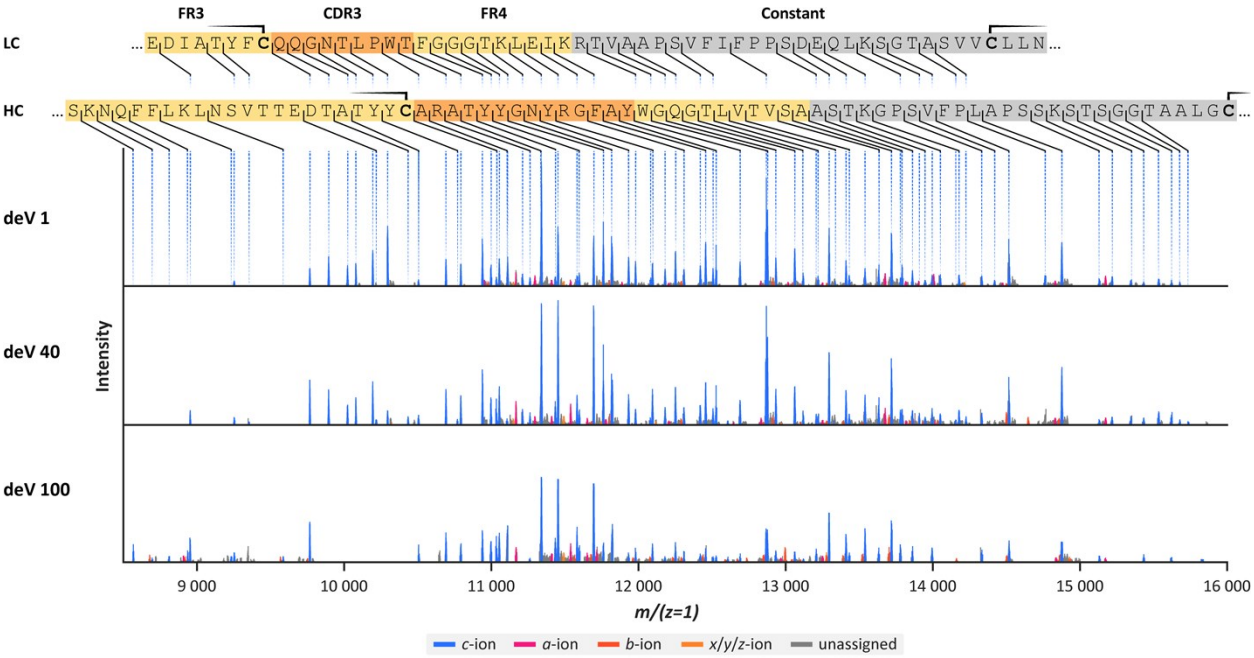

Figure S2. Color annotated EChcD-MS spectra of the F(ab')$_2$ molecule of anti-DNP IgG1 measured by using different levels of supplemental collisional activation. Fragment *c*-ions of the LC and HC are indicated in blue, while *a*- and *b*-ions are indicated in purple and red respectively. C-terminal fragment ions (*z*/*y*/*z*-ions) are shown in orange. The sequences are indicated above with the cysteine residues involved in the disulfide bonds denoted with a solid black line. Although *c*-ion sequence coverage is somewhat extended by supplemental collisional activation, this is in less informative parts of the IgG, such as the FR3 and constant regions. Moreover, with the increase in supplemental collisional activation, more fragments of additional ion series are introduced, crowding the spectrum. For this experiment, the fragment ion search was performed with a 5 ppm tolerance.

Table S2. Percentage of the total ion intensity within the 8 500 < Mw < 16 000 Da window of **Figure S2** that can be explained by searching for all different ion types with a 5 ppm tolerance. When minimizing collisional activation using a HCD direct eV setting of 1, about 70% of the total ion signal corresponds to *c*-ions. Increasing the level of collisional activation leads to a decrease in *c*-ion intensity, while *b*-ion intensity increases.

HCD deV	*a*-ions	*b*-ions	*c*-ions	*x* or *x*·-ions	*y*-ions	*z* or *z*·-ions
1	4.3	**0.7**	**70.8**	4.3	7.3*	1.6
40	3.2	**2.3**	**69.4**	3.0	6.0*	2.7
100	3.4	**6.1**	**66.8**	4.9	2.5*	1.7

* The relatively high *y*-ion intensity results from two LC *y*-ions that can also be explained as *c*-ions from the HC.

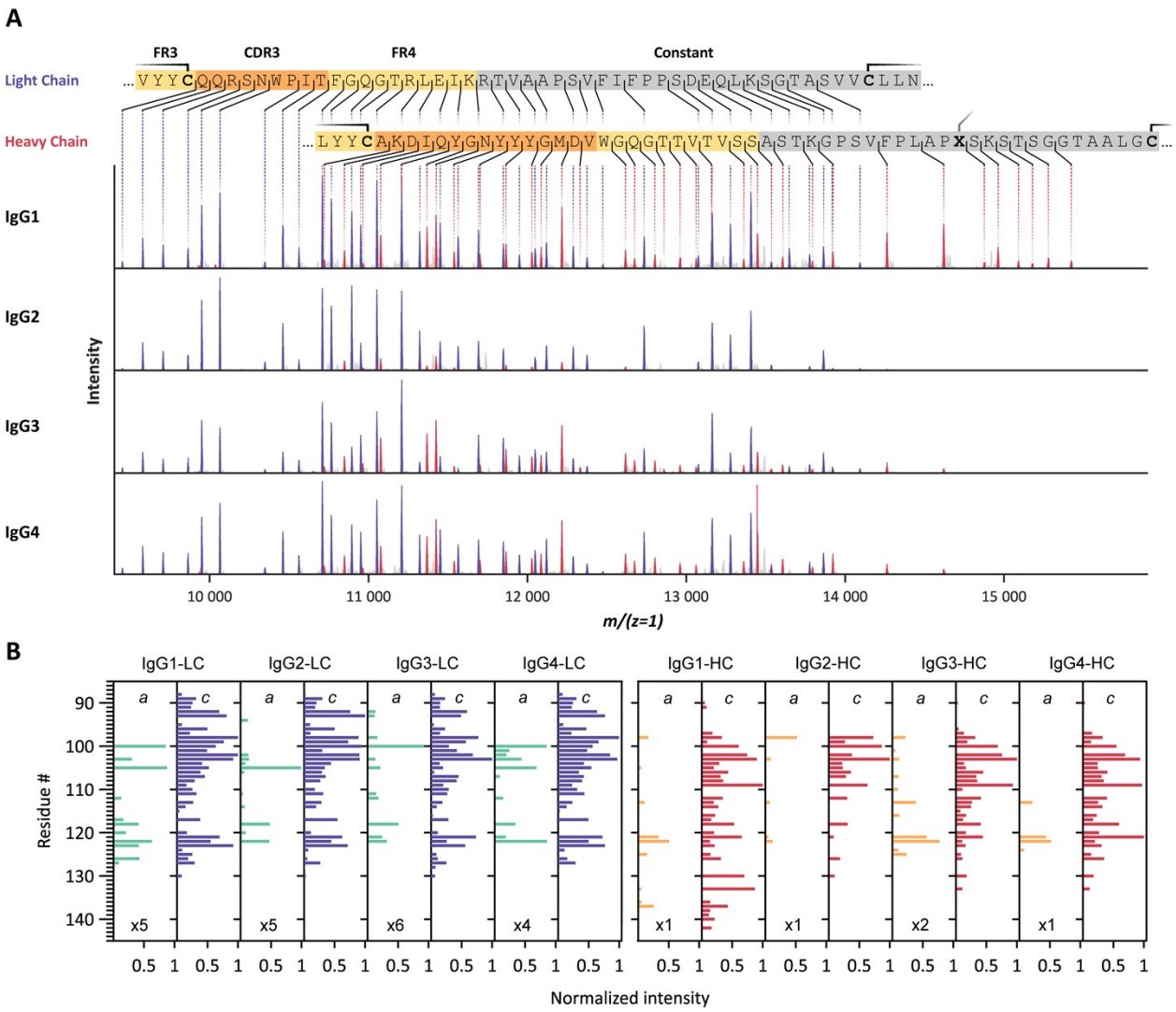

Figure S3. A. Charge-deconvoluted ECD-MS spectra of IgG1, IgG2, IgG3 and IgG4 F(ab')$_2$ molecules of anti-CD20 showing exclusively *c*-ion ladders covering the LC (purple) and HC (red) CDR3 regions. Cysteine residues involved in disulfide bonds are highlighted on the color-coded sequences by a solid line. Residue X, in …PLAPXS… in the heavy chain, represents a Ser for IgG1, and a Cys for IgG2-4. **B.** Bar-chart displaying normalized *c*- and *a*-ion fragment intensities for each residue of IgG1, IgG2, IgG3 and IgG4 LC and HC. Although all ECD spectra for a given mAb are very similar, IgG1 HC *c*-ion fragments further extend towards the C-terminus. For some cleavage sites, a-ions complement missing c-ions.

Table S3. Percentage of sequence-describing cleavages covered by *c*-ions in the ECD-MS spectra related to **Figure 2** and **Figure S3** resulting from a search with a 3 ppm tolerance. Near-complete coverage is obtained for the CDR3 and FR4 regions of both the LC and HC. The percentage of sequence-describing cleavages can be extended by using EChcD (ECD with supplemental collisional activation) as depicted in Figure S2. Opting for EChcD is, however, only recommended to complete sequence reads due to the formation of additional ion types (e.g. *b/y*), hindering de novo analysis, as discussed in the text.

mAb	LC CDR3	HC CDR3	LC CDR3-FR4	HC CDR3-FR4	LC	HC
IgG1 anti-CD20	**90.0**	**75.0**	95.0	77.8	21.0	15.4
IgG1 anti-DNP	**80.0**	**93.3**	85.0	92.3	16.4	16.7
IgG2 anti-CD20	**90.0**	**62.5**	95.0	48.1	18.7	6.3
IgG2 anti-DNP	**90.0**	**86.7**	95.0	80.8	17.3	11.0
IgG3 anti-CD20	**90.0**	**87.5**	95.0	88.9	21.5	10.8
IgG3 anti-DNP	**80.0**	**86.7**	85.0	84.6	14.5	11.1
IgG4 anti-CD20	**90.0**	**75.0**	95.0	77.8	20.6	11.3
IgG4 anti-DNP	**80.0**	**86.7**	85.0	84.6	15.9	12.2
Average	**86.3**	**81.7**	91.3	79.4	18.2	11.8

Table S4. Percentage of the total ion intensity within the 9 000 < Mw < 16 000 Da window of Figures 2A and S3 that can be explained by searching for all different ion types with a 3 ppm tolerance. Measurements were performed under minimal pressure and with an HCD direct eV setting of 1 to prevent supplemental collisional activation. Notably, in most ECD spectra typically quite a few of the ion signals cannot be explained, and therefore the fact that we can assign about above 80% of all fragment ion intensities is quite exceptional. The additional 20% is either chemical noise or other non-annotated ion signals (e.g. small neutral losses of c-ions).

mAb	*a*-ions	*b*-ions	*c*-ions	*x* or *x*·-ions	*y*-ions	*z* or *z*·-ions
IgG1 anti-CD20	**5.5**	1.4	**76.0**	-	0.2	0.3
IgG1 anti-DNP	**3.9**	1.0	**73.6**	0.3	6.2*	0.5
IgG2 anti-CD20	**2.2**	0.3	**83.9**	0.1	-	-
IgG2 anti-DNP	**2.8**	0.3	**73.9**	2.4	2.0	0.7
IgG3 anti-CD20	**4.2**	1.2	**78.1**	-	-	-
IgG3 anti-DNP	**1.9**	1.4	**73.2**	1.2	0.9	-
IgG4 anti-CD20	**5.3**	0.7	**80.0**	-	-	-
IgG4 anti-DNP	**1.8**	0.5	**70.3**	1.5	1.5	0.7
Average	**3.4**	0.9	**76.1**	0.7	1.3	0.2

* The relatively high *y*-ion intensity results from two LC *y*-ions that can also be explained as *c*-ions from the HC.

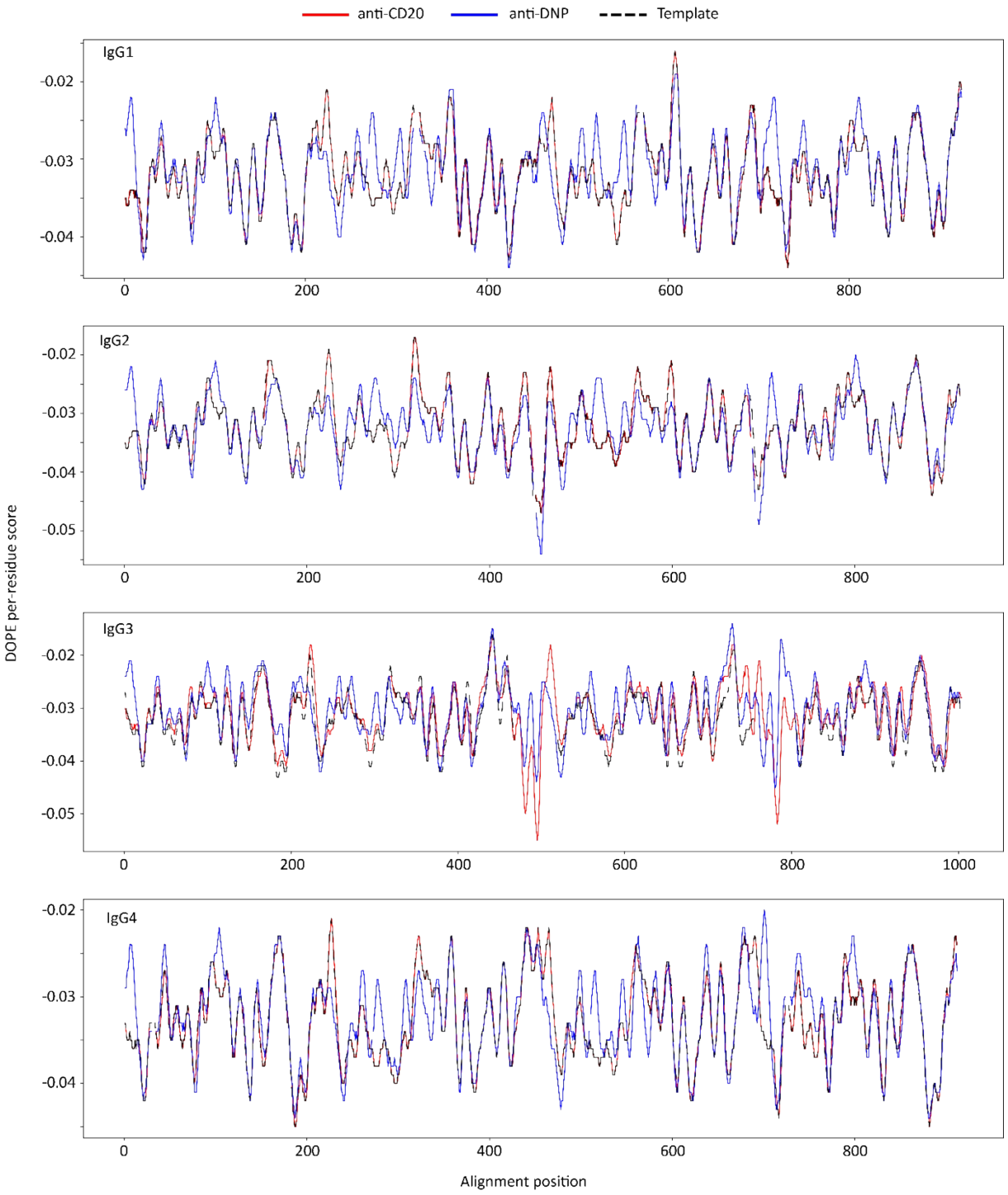

Figure S4. Discrete Optimized Protein Energy (DOPE) profiles showing the DOPE per-residue score as function of the residue number (alignment position) for F(ab')$_2$ subclasses (IgG 1-4) comparatively modelled for anti-CD20 and anti-DNP. While the total DOPE score displays an invert correlation with the native character of a structure – the total DOPE score is lower for more native like (in vacuo) structures[1] –, the DOPE score of each residue, displayed here, indicates how much a residue contributes to the protein energy as well as, upon comparison with reference values, how much the local structure potentially differs

from a template or analogue “ideal” structure. Comparison of the three curves sets (black dash, red, and blue) highlights the perfect match between anti-CD20 (red) and the templates (black dashes) while anti-DNP (blue) deviates slightly more. The rather small positive deviations of the blue curve compared to the template nevertheless indicate that we are close to optimal structures (following the optimization of flexible loops of residues). The templates used are IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3.

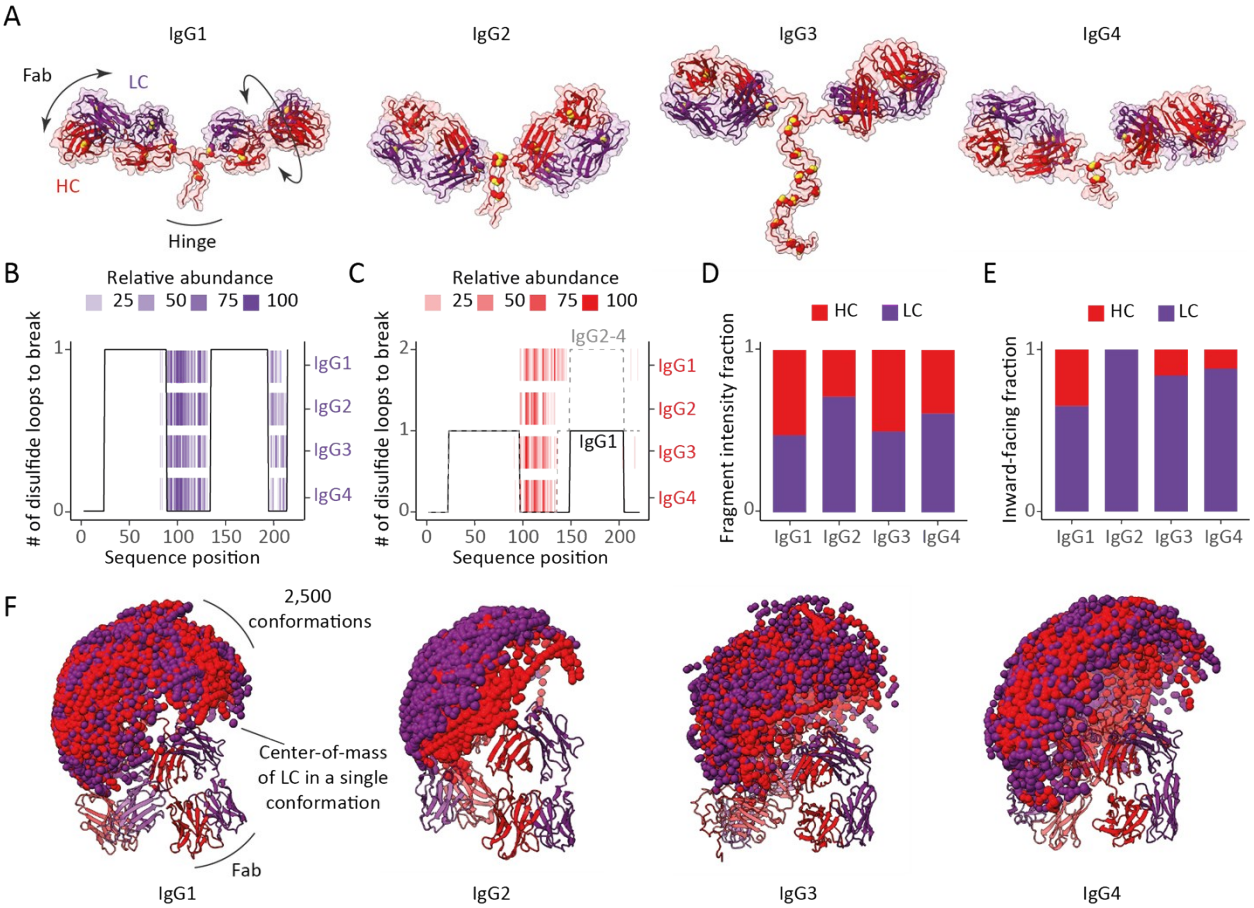

Figure S5. Influence of structural constraints on ECD fragmentation of F(ab')$_2$ from different subclasses of anti-DNP IgG. A) Structures of F(ab')$_2$ comparatively modelled against available structures of IgG subclasses (IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3). Disulfide bridges are displayed as paired yellow spheres. B-C) Number of disulfide bonds that span each sequence position of (B) the LC and (C) the HC in different subclasses of IgG anti-DNP overlaid with the corresponding ECD fragments. D) Fractions of ECD fragment intensities originating from the LC and HC for the different subclasses of IgG anti-DNP. E) Fractions of exposed LC and HC in the sampling of 500 most compact conformations. See Figure S7 for more detailed description of the estimated chain relative exposure based on the center-of-mass calculations. F) Sampling of F(ab') arm movements displayed by spheres representing the center-of-mass of (purple) LCs and (red) HCs, showing remarkable differences in between the different subclasses.

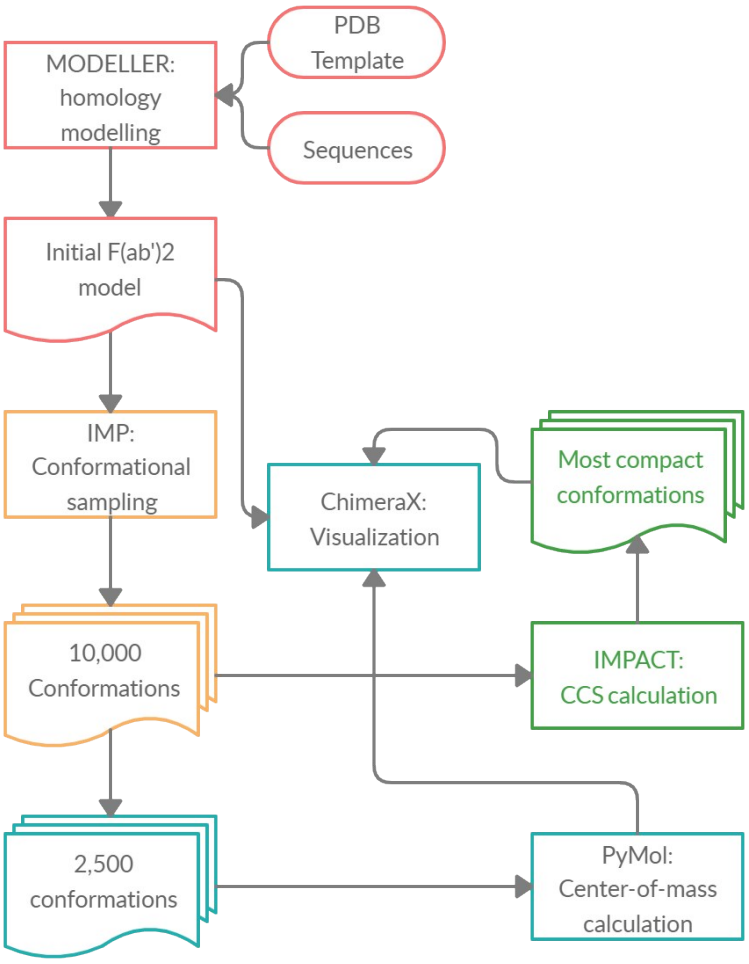

Figure S6. Modelling workflow used to obtain F(ab')$_2$ structural models for of anti-CD20 and anti-DNP IgG1, IgG2, IgG3 and IgG4. First, the structures were homology modelled against the available structures of IgGs from the different subclasses (taken were the following PDB structures; IgG1: 1HZH, IgG2: 1IGT, IgG3: 5A16, IgG4: 5DK3). Next, the F(ab') arm movement was sampled by using the Integrative Modelling Platform rapidly exploring the random tree algorithm,[2] where the F(ab') portions were set as a rigid body and amino acids in the hinge region were set to rotate in a random manner in line with their torsional space. From the 10 000 generated conformations 2 500 were used for the calculation of the center-of-mass (COM) with PyMol for both the entire F(ab')$_2$ and each chain separately. For COM calculations of the IgG3 F(ab')$_2$, we included only the first two hinge disulfides to avoid the offset created by the large hinge region. Subsequently the average distances from the heavy or light chain's COM to the full molecule's COM were used to estimate the relative exposure of the light and heavy chain within each conformation. In parallel, the entire population of F(ab')$_2$ conformations was subject to calculation of collisional cross-sections (CCS) making use of the Impact software.[3] All visualizations were performed in ChimeraX environment.[4]

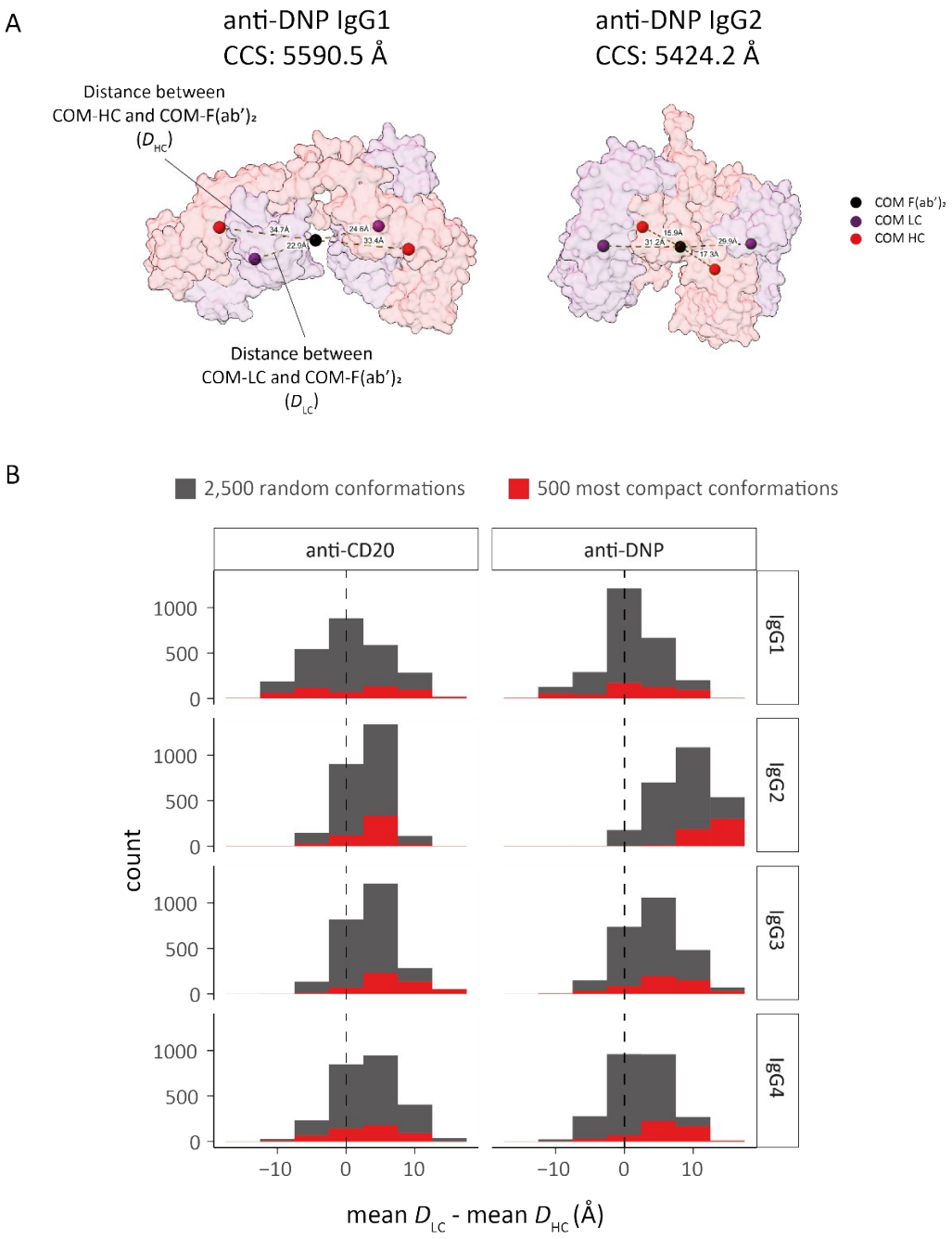

Figure S7. Estimation of the relative exposure in the F(ab')$_2$ molecules based on the Center-of-mass (COM) based calculations (Figure S6). A) Distances between the light chain COM and F(ab')$_2$ COM (D_{LC}) and the heavy chain COM and F(ab')$_2$ COM (D_{HC}) are visualized in compact conformations of anti-DNP IgG1 and IgG2. B) Distributions of differences between average Center-of-mass D_{LC} and D_{HC} in 2500 random and 500 compact conformations of anti-CD20 and anti-DNP IgG1, IgG2. IgG3 and IgG4.

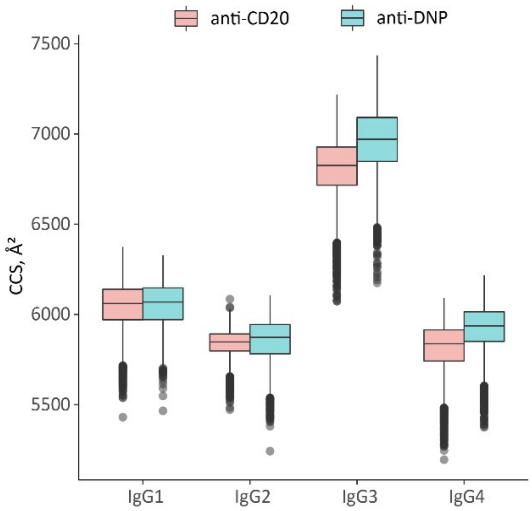

Figure S8. Distribution of collisional cross-sections of IgG anti-CD20 and IgG anti-DNP conformations generated by using the Integrative Modelling Platform[2]. Collisional cross-sections were calculated with Impact software[3].

Table S5. Amino acid sequences of the anti-CD20 and anti-DNP IgG mAbs used in our experiments. The cleavage site of the IdeS enzyme that produces the $F(ab')_2$ is indicated with a red forward slash in the heavy chain sequences.

IgG anti-CD20 (7D8)

IgG1 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRF
TISRDNAKNSLYLQMNSLRAEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPS
SKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYI
CNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDV
SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEK
TISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDG
SFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

IgG2 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRF
TISRDNAKNSLYLQMNSLRAEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPC
SRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYT
CNVDHKPSNTKVDKTVERKCCVECPPCPAPPVA/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHED
PEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISK
TKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFL
YSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQkSLSLSPGK

IgG3 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRF
TISRDNAKNSLYLQMNSLRAEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPC
SRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYT
CNVNHKPSNTKVDKRVELKTPLGDTTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPEPKSC
DTPPPCPRCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNA
KTKPREEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREPQVYTLPPSR
EEMTKNQVSLTCLVKGFYPSDIAVEWESSGQPENNYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIF
SCSVMHEALHNRFTQKSLSLSPGK

IgG4 Heavy Chain

EVQLVESGGGLVQPDRSLRLSCAASGFTFHDYAMHWVRQAPGKGLEWVSTISWNSGTIGYADSVKGRF
TISRDNAKNSLYLQMNSLRAEDTALYYCAKDIQYGNYYYGMDVWGQGTTVTVSSASTKGPSVFPLAPC
SRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYT
CNVDHKPSNTKVDKRVESKYGPPCPSCPAPEFLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQE
DPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTIS
KAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFF
LYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK

κ Light Chain

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSG
TDFTLTISSLEPEDFAVYYCQQRSNWPITFGQGTRLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLL
NNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSP
VTKSFNRGEC

IgG anti-DNP (G2a2)

IgG1 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRI
SITRDTSKNQFFLKLNSVTTEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPSS
KSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYIC
NVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVS
HEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT
ISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGS
FFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

IgG2 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRI
SITRDTSKNQFFLKLNSVTTEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCS
RSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYTC
NVDHKPSNTKVDKTVERKCCVECPPCPAPPVA/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDP
EVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKT
KGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLY
SKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

IgG3 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRI
SITRDTSKNQFFLKLNSVTTEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCS
RSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYTC
NVNHKPSNTKVDKRVELKTPLGDTTHTCPRCPEPKSCDTPPPCPRCPEPKSCDTPPPCPRCPEPKSCD
TPPPCPRCPAPELLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFKWYVDGVEVHNAK
TKPREEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKTKGQPREPQVYTLPPSRE
EMTKNQVSLTCLVKGFYPSDIAVEWESSGQPENNYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIFS
CSVMHEALHNRFTQKSLSLSPGK

IgG4 Heavy Chain

DVRLQESGPGLVKPSQSLSLTCSVTGYSITNSYYWNWIRQFPGNKLEWMVYIGYDGSNNYNPSLKNRI
SITRDTSKNQFFLKLNSVTTEDTATYYCARATYYGNYRGFAYWGQGTLVTVSAASTKGPSVFPLAPCS
RSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTKTYTC
NVDHKPSNTKVDKRVESKYGPPCPSCPAPEFLG/GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQED
PEVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISK
AKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFL
YSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLGK

κ Light Chain

DIRMTQTTSSLSASLGDRVTISCRASQDISNYLNWYQQKPDGTVKLLIYYTSRLHSGVPSRFSGSGSG
TDYSLTISNLEQEDIATYFCQQGNTLPWTFGGGTKLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLL
NNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSP
VTKSFNRGEC

Table S6. Overview of the molecular masses of the used intact IgGs, F(ab')$_2$ and Fab's. The theoretical mass is obtained from the provided protein sequences as shown in **Table S3**, the experimental mass is determined by mass spectrometry.

Species	$MW_{theoretical}$ (Da)	$MW_{experimental}$ (Da)	Mass difference (Da)
Trastuzumab G1F/G0F	148 476	148 220	-256.0*
Trastuzumab F(ab')$_2$	97 629	97 631	1.6
Trastuzumab Fab	47 500	47 502	1.9
IgG1 anti-DNP F(ab')$_2$	99 201	99 203	2.3
IgG2 anti-DNP F(ab')$_2$	98 630	98 632	1.7
IgG3 anti-DNP F(ab')$_2$	109 352	109 355	3.2
IgG4 anti-DNP F(ab')$_2$	98 782	98 783	1.2
IgG1 anti-CD20 F(ab')$_2$	98 430	98 432	2.3
IgG2 anti-CD20 F(ab')$_2$	97 859	97 862	2.3
IgG3 anti-CD20 F(ab')$_2$	108 582	108 588	6.0
IgG4 anti-CD20 F(ab')$_2$	98 011	98 014	2.7

* Corresponding to the mass shift induced by C-terminal lysine clipping of both HCs.

REFERENCES

1. M. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, *Protein Sci*, 2006, **15**, 2507–2524, doi:10.1110/ps.062416606.

2. D. Russel, K. Lasker, B. Webb, J. Velázquez-Muriel, E. Tjioe, D. Schneidman-Duhovny, B. Peterson, A. Sali, Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies, *PLOS Biology*, 2012, **10**, e1001244, doi:10.1371/journal.pbio.1001244.

3. E. G. Marklund, M. T. Degiacomi, C. V. Robinson, A. J. Baldwin, J. L. P. Benesch, Collision Cross Sections for Structural Proteomics, *Structure*, 2015, **23**, 791–799, doi:10.1016/j.str.2015.02.010.

4. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis, *Protein Sci*, 2018, **27**, 14–25, doi:10.1002/pro.3235.