Inhibition of KDM4D and Stabilisation of the PHF8 Plant Homeodomain's Transient Structural States Using Antibodies

DPhil Thesis *Finn Wolfreys* *Trinity College*

Trinity Term, 2017

Abstract

Though antibodies as therapeutics are limited to extracellular targets, their repertoire of molecular interactions has particular relevance to the many intracellular cellular proteins for which small molecule screening has reached impasse. For such proteins there is little recourse to theory, since molecular recognition is, in practical terms, still not well understood. Here I apply antibody discovery to the lysine demthylases KDM4D and PHF8, two proteins difficult to inhibit selectively due to the similarity of their binding pockets to those of the larger family. With a selective, picomolar affinity antibody, dependent on residues distal to the KDM4D active site, I present what is likely the first example of allosteric inhibition of a KDM4 lysine demethylase, demonstrating that there is opportunity outside active sites oversubscribed with pan inhibitors. Antibody discovery for PHF8, however, was plagued by a familiar problem: antibodies that bound when their antigen was immobilised directly to a surface, but barely bound at all when it was free in solution. The common explanation is that the partial denaturation that accompanies immobilisation reveals epitopes unavailable in solution, but examining the problem in detail for the Plant Homeodomain of PHF8 revealed a connection to its rarely sampled conformations. The prominence these antibodies in the immune responses to PHF8, and to some extent KDM4D, motivates two hypotheses on their origin: either the states are very immunogenic or there is a connection between states of irreversible damage and those sampled reversibly, but rarely, by a protein in solution.



In memory of John Arthur Wolfreys

Contents

Contents **4**

1 Introduction **7**

2 KDM4D **15**

2.1 Introduction 15

Lysine Demethylase 4D 15

Monoclonal Antibody Recovery From Single Cells 20

2.2 Results 26

Preparation of KDM4D 26

Activity Assay 26

Antibody Discovery 28

Inhibition Assay 34

The GDI Family 35

The GDI Family's Epitope 42

Further Crystal Trials and Remaining Antibodies 47

Summary of Main Results 48

2.3 Conclusions 48

3 PHF8 **52**

3.1 Introduction 52

PHD Finger Protein 8 52

3.2 Results 55

Antibody Discovery on PHF8 55

Solution Binding Behaviour 63

Solution Binding to PHF8 plant homeomain (PHF8-PHD) 69

Recognition of PHD finger containing protein 8 (PHF8)$_{1-447}$ is Possible 76

Summary of Main Results 78

3.3 Conclusions 78

4 PHF8-PHD **82**

4.1 Introduction 82

Phage Display 83

4.2 Results 86

Immune Phage Display 86
Preimmune Phage Display 102
Inhibition of the PHD-H3K4me3 Interaction 108
Epitope Mapping by NMR 112
Crystallography 135
Summary of Main Results 140
4.3 Conclusions 140

5 Problems in Solutions **144**
5.1 Introduction 144
5.2 Results 144
SPR Affinities 144
Binding Cooperativity 151
Time 158
EDTA 160
Zinc 166
Cross-Reactivity 168
Heat 172
5.3 Conclusions and Future Work 174

6 Conclusions and Future Work **179**

7 Materials and Methods **191**
7.1 Subcloning 191
Avi-tag 191
Cloning PHF8-PHD Constructs 192
7.2 Protein Purification 192
Tranformation by Heat Shock 192
Bacterial Culture 192
Culture for Biotinylation 193
Purification 193
7.3 Immunisation Constructs 194
KDM4D 194
PHF8 195
7.4 Antibody Recovery 195
B-Cell Culture 195
FMAT 195
Fluorescent Foci Assay 196
Reverse Transcriptase PCR 197
Antibody Recovery 197
Subcloning Antibody Variable Regions 198
Recombinant Antibody Expression 199
Antibody Purification 199

7.5 Analysis 200
ELISA 200
Analytical Size Exclusion Chromatography 201
Affinity Determination by SPR 201
PHF8-PHD SPR Binding Assay 202
Activity Assay 202
PAR/PMPS Assay 203
7.6 Phagemid Display Library Construction 203
Pull-Through PCR 205
Ligation and Transformation 205
Phage Rescue 206
Phage Panning (Plastic) 207
Phage Panning (Solution) 207
Chemical Biotinylation 208
Monoclonal Phage Rescue 208
Phage ELISA 209

1 Introduction

Screening is a technical solution to a scientific problem. The problem is that molecular recognition is not yet understood deeply enough to rationally design drugs *de novo*. The solution is to substitute rational design for a random sampling of interactions through chemical libraries. It is a compromise essential to drug discovery. A successful screening campaign will return compounds making the few interactions necessary to gain a foothold on a protein, and thus begin the process of drug discovery. But so small are chemical libraries in the shadow of drug-like chemical space that an unsuccessful screen admits few conclusions. For the many therapeutically relevant targets failed by small molecule drug discovery, there are few alternatives besides further screening, and as always, no guarantees.

Here I apply antibodies to breaking the stalemate that small-molecule drug discovery for lysine demethylase 4D (KDM4D) and PHF8 has reached. Despite extensive screening, and active sites seemingly amenable to inhibition, it has been difficult to find inhibitors for these lysine demethylases (KDMs) that are not co-factor mimetics with poor selectivity. Both proteins are intracelluar, so they would usually be outside the purview of antibodies and perhaps not be considered for antibody discovery. But the need to understand novel function modifying interactions for both KDM4D and PHF8 is great, and those available through the immune system's repertoire possibly some of the most diverse and readily generated. My thesis is thus: *What insights can antibodies provide about the targets that small-molecule drug discovery fails, and what reflection does it allow on molecular recognition?*

Molecular Recognition

The basis for much of drug discovery is engineering the specific non-covalent recognition of one molecular by another. It is a problem that is deceptively simple. Many of the phenomena that support molecular recognition are easily explained qualitatively, but without prior knowledge accurate prediction of binding affinity or even design of binding partners is not possible in any general way. Described in terms of thermodynamic potentials, molecular recognition is a compromise between enthalpy change, ΔH, and entropy change ΔS. Complexation is favoured when the corresponding Gibbs free energy, $\Delta G = \Delta H - T\Delta S$, where T is temperature, is lower for two molecules in complex than when apart. A straightforward interpretation is that specific interactions must overcome a systems tendency towards disorder, that the formation of hydrogen bonds, salt bridges, cation-π, and van der Waals interactions made familiar by structural biology, must release enough energy to account for the ligand being denied the multiplicity of states available to it when free in solution. Were molecular recognition fully described by this interpretation, a proper statistical mechanical treatment of molecules the size of proteins, over the number of states permitted by their flexibility and in an aqueous environment, would be challenging; but even this interpretation is too simple.

Perhaps what complicates molecular recognition most is that how enthalpy and entropy manifest in biological systems, both *in vitro* and *in vivo*, is for many phenomena poorly understood.[1] Taking the example of water, changes to the order and disorder of water molecules essentially contributes a silent entropic term to many non-covalent interactions. This contribution can be through an increase in entropy, when water is expelled from a binding site, or a decrease in entropy, when water is trapped in a binding site. Furthermore, enthalpic contributions of water through mediating hydrogen bonding can also compensate for the entropic penalty of it being trapped.[2] And this is a single complication to the inherently cooperative problem of molecular recognition. In addition, the behaviour of electrostatic networks

in proteins has likely not been fully appreciated,[3] nor has protein flexibility,[4] and the influence of enthalpy-entropy compensation remains an open, if somewhat controversial, area of investigation.[5,6] Resolving the problem of molecular recognition is an ambitious task, and one for which the appropriate experimental a computational techniques may not yet exist.[1] However, evolution has wrought in antibodies a system for efficient generation of novel non-covalent interactions whose design deserves the attention of any attempt to understand molecular recognition.

Antibodies

As therapeutics, antibodies have few desirable properties. They are subject to denaturation,[7,8] degradation,[9] and aggregation;[10] they can be immunogenic;[11] their sensitivity to pH and digestive enzymes prevents oral administration; without specific engineering they struggle to cross the blood brain barrier;[12] they cannot enter the cell, and even if they could would lose integral disulphide bonds; they are expensive; and making generic "biosimilars" is non-trivial.[13] Yet, in 2016, 5 of the top 10 selling drugs were monoclonal antibodies.[14] So valuable is expanding the repertoire of interactions available to drug discovery that all these shortcomings of therapeutic antibodies are tolerated.[15]

The immune system curates, in B-cells, a preimmune library of approximately 10^{12} antibodies.[16] They are built to serve molecular recognition: should they be unable to bind, their host B-cell is programmed to seek its own demise.[17] But even 10^{12} B-cells only cover a fraction of chemical space, so those saved by the affinity of their antibody elaborate on its sequence by somatic hypermutation.[17,16] When molecular interactions are modified by random mutation, just as when they are selected at random, failure is more likely than success, so selection is, again, by competition – a silent process of life-and-death efficiency.

To generate antibodies is to engineer novel protein-protein interactions. Function modulation is often an accompaniment. In fact, the success of

antibodies as inhibitors in their own right relies on it. But antibodies, by stabilising conformation relevant to function or structural studies, or more aggressively sampling epitopes outside pockets fond of small molecules, can offer more to drug discovery than an alternative modality.[15,18] Though, as the inspiration for synthetic small molecules, antibodies have been used only once.[19] It is perhaps because what separates antibodies from small molecules, besides their format, are the same poorly understood principles of molecular recognition that separate receptors from rationally designed inhibitors. However, for the targets that drug discovery leaves behind, the manipulation of molecular recognition by antibodies may offer enough to break the stalemate.

Epigenetics

Separately differentiated cells, though they share the same genome, produce progeny in their likeness. Epigenetics is the corollary: there is heritable genetic information besides a sequences of DNA base pairs. Actually, there is little in chromatin, the structure of protein and DNA that packages the genome, not exploited to encode this auxiliary information. Even the tail of a single subunit in the histone octomers that with DNA form nucleosomes, is subject to: methylation, acetylation, PARylation, ubiquitinylation, phosphorylation, sumoylation, ribosylation and citrullination.[20] Evolution's ruthless optimisation implies each modification in this "histone-code" carries its own meaning important to function,[21] and the reader, writer and eraser domains wrought for its manipulation, close as they are to gene expression, are attractive drug targets.[22]

KDM4D and PHF8 remove methyl groups from a specific lysine, on a specific histone tail, and in doing so influence gene expression. Their epigenetic eraser domain is the jumonji-C (JmjC) domain, and their family the JmjC domain containing family of KDMs, itself a subfamily of 2-oxoglutarate (2OG) oxygenases. Their mechanism is one turned toward diverse ends in biology, among them: hydroxylation, halogenation, and ring

formation.[23] Demethylation follows from hydroxylation.[24,25] A reactive oxyferryl species is generated, subsequently hydroxylating methyl-lysine and thus facilitating its release as formaldehyde. It is here, with this versatile mechanism, that problems of selectivity begin. Biology's habit of sharing small molecule co-factors puts little evolutionary pressure on the pockets they bind to be different. So synthetic small molecules, predisposed by thermodynamics to the same pockets, must appeal to some artefact of evolution that can differentiate one homologue from another. Though the specificity of KDM4D and PHF8 differentiate them from their closest relatives (KDM4D is specific H3K9me3/2 and PHF8 is specific for H3K4me3K9me2/1[a]) it has not been enough to turn hits that are almost entirely 2OG competitors into selective inhibitors.[26] And it may never be enough. However, the control of modulations of epigenetics demands selectivity.

To encode 20 amino acids requires the four base pairs and three nucleotides of the genetic code, but to encode the epigenome evidently requires an alphabet considerably larger. KDMs, a single class of epigenetic proteins, are sensitive to the specific methylation states, of specific lysine residues, on specific histone tails. Just the first three lysine residues of histone three, by being methylated 0, 1, 2, or three times encode as much as the genetic code. The implication is that methylation, in concert with the many other epigenetic modification, describes a system responsible for much more than up- and down- regulation. And it is a system too nuanced, managed by too few actors, to tolerate non-specific inhibition.

Thesis

Antibodies are unlikely therapeutics. That they are used at all demonstrates the value of novel molecular recognition to drug discovery. However, the weight of biology's lessons is more often in physical principles that constrain its solutions, than the solutions themselves. Antibodies solve im-

[a]The abbreviation H3K9me3/2 refers to histone 3, lysine 9, tri- or di-methylated. Similarly, H3K4me3K9me2/1 refers to histone 3, lysine 4 tri-methylated, lysine 9 di- or mono-methylated.

munological action at a distance by specific recognition of foreign antigens and recruitment of immune elements. When they also modulate function, antibodies reach ideals of small molecule drug discovery, that is, the potent, selective modulation of biological function. The principles of molecular recognition antibodies must appease are the same as those of small molecule drug design, while the sampling of molecular interactions performed in antibody generation is many orders of magnitude greater than that possible with a library of small molecules.

However, antibody binding and small molecular binding initially seem quite different. Whereas small molecules are directed towards binding pockets, antibodies may engage surfaces, linear peptides, or small molecules. It appears that the size of antibodies affords them a much larger diversity of potential binding partners, so how antibodies could be the inspiration for small molecule ihibitors may not be immediately obvious. But closer examination of many antibody-antigen complex structures reveals often less than 10 % of residues at the binding site are involved in non-covalent interactions.[27] Though small molecules inspired by antibodies are unlikely to resemble antibodies in structure, these few strong interactions, "hot-spots",[28] in combination with those from other antibodies could provide constraints to the design of new small molecules.

The size and diversity of chemical space available to antibodies at their choice of twenty amino acids positioned on six complementarity determining regions (CDRs) is significantly less than that available to synthetic molecules. But despite this seeming disadvantage they offer one of the most diverse examples of molecular recognition available. Furthermore, intentionally restricting the sequence diversity of antibody CDRs can still lead the generation of antibodies recognising diverse epitopes.[29] Implicit in the design of antibodies is, it seems, efficient manipulation of molecular recognition more refined than simply blind application of chemical diversity, and therefore, understanding the solutions antibodies provide to problems of small molecule drug discovery could be a practical means towards a practi-

cal understanding of molecular recognition.

I hoped to generate function modifying antibodies and then identify the few chance interactions that confer specific inhibition. Perhaps ones that could transform a non-selective small-molecule scaffold, or even constrain a search for novel scaffolds through its pharmacophores, thus turning antibodies towards intracellular proteins. Thus, chapters 1, 2, and 3 detail antibody discovery campaigns and the search for inhibitors of KDM4D and PHF8 and the plant homeodomain (PHD) of PHF8. However, antibodies can offer more than examples of molecular recognition to copy. Through a seemingly contradictory property: high affinity but low complex formation, these antibodies permit a reflection on both molecular recognition and protein dynamics, the topic of chapter 4, and permit two hypotheses on their origin: one immunological and one structural, the topic of chapter 5.

References

[1] G. M. Whitesides and V. M. Krishnamurthy. "Designing ligands to bind proteins". In: *Quarterly reviews of biophysics* 38.4 (2005), pp. 385–395.

[2] J. E. Ladbury. "Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design". In: *Chemistry & biology* 3.12 (1996), pp. 973–980.

[3] I. Gitlin, J. D. Carbeck, and G. M. Whitesides. "Why are proteins charged? Networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis". In: *Angewandte Chemie International Edition* 45.19 (2006), pp. 3022–3060.

[4] P. L. Kastritis and A. M. Bonvin. "On the binding affinity of macromolecular interactions: daring to ask why proteins interact". In: *Journal of The Royal Society Interface* 10.79 (2013), p. 20120835.

[5] B. Breiten et al. "Water networks contribute to enthalpy/entropy compensation in protein–ligand binding". In: *Journal of the American Chemical Society* 135.41 (2013), pp. 15579–15584.

[6] J. D. Chodera and D. L. Mobley. "Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design". In: *Annual Review of Biophysics* 42.1 (2013), pp. 121–142.

[7] A. W. Vermeer and W. Norde. "The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein". In: *Biophysical journal* 78.1 (2000), pp. 394–404.

[8] K. L. Lazar, T. W. Patapoff, and V. K. Sharma. "Cold denaturation of monoclonal antibodies". In: *MAbs*. Vol. 2. 1. Taylor & Francis. 2010, pp. 42–52.

[9] J. Vlasak and R. Ionescu. "Fragmentation of monoclonal antibodies". In: *MAbs* 3.3 (2011), pp. 253–263.

[10] W. Li et al. "Antibody aggregation: insights from sequence and structure". In: *Antibodies* 5.3 (2016), p. 19.

[11] W. Y. K. Hwang and J. Foote. "Immunogenicity of engineered antibodies". In: *Methods* 36.1 (2005), pp. 3–10.

[12] S.-U. Shin et al. "Transferrin-antibody fusion proteins are effective in brain targeting." In: *Proceedings of the National Academy of Sciences* 92.7 (1995), pp. 2820–2824.

[13] P. A. Scolnik. "mAbs: a business perspective". In: *MAbs* 1.2 (2009), pp. 179–184.

[14] A. Philippidis. *The Top 15 Best-Selling Drugs of 2016*. 2017. URL: `http://web.archive.org/web/20170805103835/http://genengnews.com/the-lists/the-top-15-best-selling-drugs-of-2016/77900868` (visited on 05/08/2017).

[15] A. D. Lawson. "Antibody-enabled small-molecule drug discovery". In: *Nature reviews. Drug discovery* 11.7 (2012), p. 519.

[16] B. Alberts. *Molecular biology of the cell*. Garland science, 2017.

[17] P. J. Delves et al. *Essential immunology*. John Wiley & Sons, 2017.

[18] C. J. Marsden et al. "The use of antibodies in small-molecule drug discovery". In: *Journal of biomolecular screening* 19.6 (2014), pp. 829–838.

[19] A. Raktabutr et al. "Design and synthesis of a mimetic from an antibody complementarity-determining region". In: *Science* 253 (1991), p. 762.

[20] B. Yao et al. "Epigenetic mechanisms in neurogenesis". In: *Nature reviews. Neuroscience* 17.9 (2016), p. 537.

[21] B. D. Strahl and C. D. Allis. "The language of covalent histone modifications". In: *Nature* 403.6765 (2000), p. 41.

[22] C. H. Arrowsmith et al. "Epigenetic protein families: a new frontier for drug discovery". In: *Nature reviews. Drug discovery* 11.5 (2012), p. 384.

[23] S. Martinez and R. P. Hausinger. "Catalytic mechanisms of Fe (II)-and 2-oxoglutarate-dependent oxygenases". In: *Journal of Biological Chemistry* 290.34 (2015), pp. 20702–20711.

[24] J. Schneider and A. Shilatifard. "Histone demethylation by hydroxylation: chemistry in action". In: *ACS chemical biology* 1.2 (2006), pp. 75–81.

[25] Y.-i. Tsukada. "Hydroxylation mediates chromatin demethylation". In: *The Journal of Biochemistry* 151.3 (2012), pp. 229–246.

[26] T. Maes et al. "Advances in the development of histone lysine demethylase inhibitors". In: *Current opinion in pharmacology* 23 (2015), pp. 52–60.

[27] A. D. Lawson, M. MacCoss, and J. P. Heer. "Importance of rigidity in designing small molecule drugs to tackle protein-protein interactions (PPIs) through Stabilization of Desired Conformers: Miniperspective". In: *Journal of medicinal chemistry* (2017).

[28] A. A. Bogan and K. S. Thorn. "Anatomy of hot spots in protein interfaces". In: *Journal of molecular biology* 280.1 (1998), pp. 1–9.

[29] F. A. Fellouse, C. Wiesmann, and S. S. Sidhu. "Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.34 (2004), pp. 12467–12472.

2 KDM4D

2.1 Introduction

Lysine Demethylase 4D

Even among its subfamily, the structure and activity of KDM4D is unique in its simplicity. The two PHDs and two tudor domains of lysine demethylase 4A (KDM4A), lysine demethylase 4B (KDM4B), and lysine demethylase 4C (KDM4C),[a] are in KDM4D absent. Also apparently absent is dual activity for H3K9me3 and H3K36me3. KDM4D's demethylation is limited to H3K9me3 and H3K9me2. And the only accompaniment to its catalytic JmjC domain is a jumonji-N (JmjN) domain[1] – a motif so common the two domains were once thought inseparable. The JmjN and JmjC domains, together, cover the first two-thirds of KDM4D's 523 residues, and leave only a C-terminal domain; the structure, if it has one, is still unknown.

[a] An additional lysine demethylase 4 (KDM4) family KDM, lysine demethylase 4E (KDM4E) has been reported, but is thought to be a pseudogene.

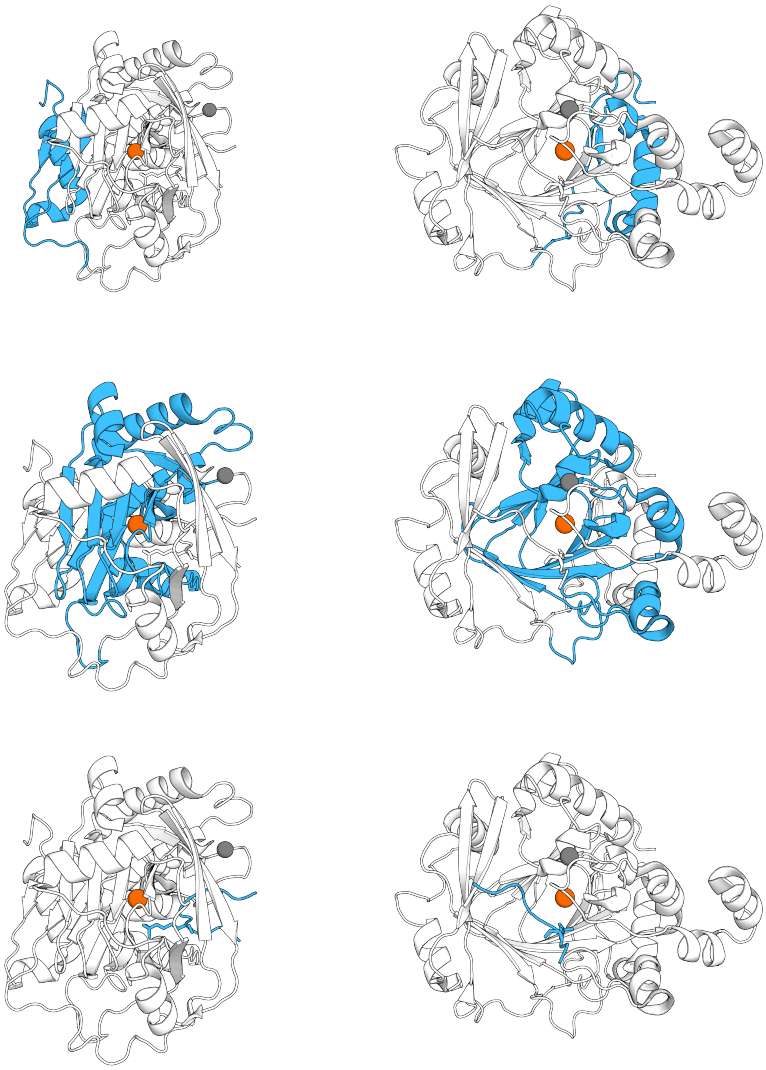


Figure 2.1: A crystal structure of KDM4D (PDB: 4HON).[1] *Highlighted in blue is, from top, the JmjN domain (domains are defined according to uniprot entry* `Q6B0I6`*), JmjC domain, and H3K9me3 peptide substrate with trimethyl-lysine shown as sticks. The left column depicts the KDM4D from the side of the active site and the right column an approximately* 90° *rotation towards the solvent-exposed active site with catalytic iron (nickel in the original structure) shown in orange and a structural zinc ion in grey.*

Expressed ectopically, in mouse embryos subjected to somatic cell nuclear transfer, KDM4D demethylation occurs in at least 200 regions of chromatin.[2] This seemingly unrestrained demethylation is an indication of both the wide application of KDM4D's function, and that, without auxiliary epigenetic domains, KDM4D's function is perhaps too general for the subtlety

of epigenetic regulation. Selectivity is mostly imparted on KDM4D's function by interaction partners – and almost always through the C-terminal domain. Poly [ADP-ribose] polymerase 1 (PARP1), p53, and androgen receptor (at least in prostate-specific antigen (PSA) activation) are all reported to require the KDM4D's C-terminal domain for recognition,[3,4,5,6] and PARylation of KDM4D by PARP1, originally reported to target the JmjN domain,[7] is significantly more efficient when directed at residues in the C-terminal region (E357, R450, R451, and R455).[4,8] The only exception to the pattern of C-terminal regulation – RNA binding to residues in the JmjC domain – is itself complemented by a second interaction with RNA through a non-canonical binding site in the C-terminal region.[9,8]

Minor DNA damage is repaired by cells, but above a threshold it appears best to activate apoptosis rather than risk propagating aberrant genes. Human osteosarcoma (U2OS) cells respond to laser microirradiation by recruitng KDM4D through interactions with PARP1 and RNA.[3,4,9,8] To these U2OS cells KDM4D appears to give some resistance – knockout resensitises them to laser microirradiation.[4] The clinical complement to these observation of KDM4D in DNA damage repair comes from patients with limited stage Hodgkin lymphoma. Those who have received radiotherapy, and whose tumours have strong KDM4D expression, are reported to have "dismal" relapse-free survival.[10] The suggestion is that KDM4D, through DNA repair, may confer resistance to radiotherapy in a mechanism conserved between different cancers.

In HCT116 colon cancer cells, knockdown of KDM4D prevents proliferation.[5] The underlying mechanism is reportedly recruitment of KDM4D by p53, and the complex's subsequent activation of p21 expression – a marker for metastasis.[5] Analogously, in patients with non-small-cell lung cancer, KDM4D expression correlated with metastasis to the lymph nodes.[11] To these interactions of KDM4D linked to cancer, two more, less assured in their KDM4D dependence, can be added: complexation with androgen receptor activates PSA expression in LNCaP prostate cancer cells,[6] and

KDM4D activation of vascular endothelial growth factor (VEGF) expression in ovine placentas.[12]

Reports of KDM4D's action in cancer, have also been accompanied by particularly specific roles in inflammation. In dendritic cells stimulated with lipopolysaccharide (LPS), levels of KDM4D messenger RNA (mRNA) increase up to 80-fold, while KDM4A, 4B, and 4C levels remain the same.[13] KDM4D's targets, the promoters of MDC and IL12b, are important in their cell-type-dependent protein expression in response to inflammatory stimuli.[13] The effect, at least for IL12b activation, is permitted by the deubiquitinase, Trabid.[14] KDM4D, spared from disassembly, can activate IL12b. That Trabid knockdown can protect mice from autoimmune disease suggests, given their relationship, perhaps a similar phenotype can be expected of KDM4D inhibition.[14]

Some cancerous cells rely on KDM4D's action to proliferate,[5] but normal cells are apparently unaffected by its absence: apart from higher global H3-K9me3 levels, mice unable to express KDM4D appear otherwise healthy.[15] KDM4D is therefore an attractive therapeutic target. However, there have not yet been any specific KDM4D inhibitors reported; and not for a lack of scaffolds: hydroxamic acids, pyridine dicarboxylates, 8-hydroxyquinolines, and bipyridines can all be KDM4D inhibitors.[16] They mimic 2OG by chelating the active site iron, but for them decarboxylation by KDM4D is impossible, and thus so too is H3K9me3/2 turnover. So similar are the poses of inhibitors derived from these scaffolds that the superposition of pyridine dicarboxylate, bipyridine, and 8-hydroxyquinolines has been suggested as a hypothetical tricyclic KDM inhibitor.[16] But there is very little in the active sites of the KDM4 family to differentiate family members, especially at the active site. There are also few reported alternatives to 2OG mimetics. Transition metal complexes have been presented, but lack selectivity.[17] As do inhibitors that eject structural zinc.[16] Therefore, there is reason to pursue alternative techniques orthogonal to small molecule drug discovery. If nothing else to inspire new constraints for the ideal small-molecule inhibitor.

```
KDM4D   METMKSKANCAQNPNCNIMIFHPTKEEFNDFDKYIAYMESQGAHRAGLAKIIPPKEWKAR
KDM4A   ----MASESETLNPSARIMTFYPTMEEFRNFSRYIAYIESQGAHRAGLAKVVPPKEWKPR
KDM4B   ---MGSEDHGAQNPSCKIMTFRPTMEEFKDFNKYVAYIESQGAHRAGLAKIIPPKEWKPR
KDM4C   --MEVAEVESPLNPSCKIMTFRPSMEEFREFNKYLAYMESKGAHRAGLAKVIPPKEWKPR
             :.     **...** * *: ***.:*.:*:**:**:*********::****** *

                                    A
KDM4D   ETYDNISEILIATPLQQVASGRAGVFTQYHKKKKAMTVGEYRHLANSKKYQTPPHQNFED
KDM4A   ASYDDIDDLVIPAPIQQLVTGQSGLFTQYNIQKKAMTVREFRKIANSDKYCTPRYSEFEE
KDM4B   QTYDDIDDVVIPAPIQQVVTGQSGLFTQYNIQKKAMTVGEYRRLANSEKYCTPRHQDFDD
KDM4C   QCYDDIDNLLIPAPIQQMVTGQSGLFTQYNIQKKAMTVKEFRQLANSGKYCTPRYLDYED
          **:*.:::* :*:**:.:*::*:****: :****** *:*::*** ** ** : ::::

                          A AA               A      A           AAA   A
KDM4D   LERKYWKNRIYNSPIYGADISGSLFDENTKQWNLGHLGTIQDLLEKECGVVIEGVNTPYL
KDM4A   LERKYWKNLTFNPPIYGADVNGTLYEKHVDEWNIGRLRTILDLVEKESGITIEGVNTPYL
KDM4B   LERKYWKNLTFVSPIYGADISGSLYDDDVAQWNIGSLRTILDMVERECGTIIEGVNTPYL
KDM4C   LERKYWKNLTFVAPIYGADINGSIYDEGVDEWNIARLNTVLDVVEEECGISIEGVNTPYL
        ********  :  ******:.*:::::. . :**:. * *: *::*.*.*  *********

        A      AAAAAAAAAAAAAAAA      AAAAAA       A
KDM4D   YFGMWKTTFAWHTEDMDLYSINYLHLGEPKTWYVVPPEHGQRLERLARELFPGSSRGCGA
KDM4A   YFGMWKTSFAWHTEDMDLYSINYLHFGEPKSWYSVPPEHGKRLERLAKGFFPGSAQSCEA
KDM4B   YFGMWKTTFAWHTEDMDLYSINYLHFGEPKSWYAIPPEHGKRLERLAIGFFPGSSQGCDA
KDM4C   YFGMWKTTFAWHTEDMDLYSINYLHFGEPKSWYAIPPEHGKRLERLAQGFFPSSSQGCDA
        *******:*****************:****:** :*****:******  :**.*::.* *

         AAAAAA                   A     AAAAAAAAAAAAA     AAAAAA
KDM4D   FLRHKVALISPTVLKENGIPFNRITQEAGEFMVTFPYGYHAGFNHGFNCAEAINFATPRW
KDM4A   FLRHKMTLISPLMLKKYGIPFDKVTQEAGEFMITFPYGYHAGFNHGFNCAESTNFATRRW
KDM4B   FLRHKMTLISPIILKKYGIPFSRITQEAGEFMITFPYGYHAGFNHGFNCAESTNFATLRW
KDM4C   FLRHKMTLISPSVLKKYGIPFDKITQEAGEFMITFPYGYHAGFNHGFNCAESTNFATVRW
        *****::**** :**: ****.::********:******************: **** **

                         A
KDM4D   IDYGKMASQCSCGEARVTFSMDAFVRILQPERYDLWKRGQDRAVVDHMEPRVPASQEL..
KDM4A   IEYGKQAVLCSCRKDMVKISMDVFVRKFQPERYKLWKAGKDNTVIDHTLPTPEAAEFL..
KDM4B   IDYGKVATQCTCRKDMVKISMDVFVRILQPERYELWKQGKDLTVLDHTRPTALTSPEL..
KDM4C   IDYGKVAKLCTCRKDMVKISMDIFVRKFQPDRYQLWKQGKDIYTIDHTKPTPASTPEV..
        *:*** *  *:* :  *.:*** *** :**:**.*** *:*  .:**  *   ::  :
```

Figure 2.2: An alignment of KDM4D, residues 1–358, to KDM4A, KDM4B, and KDM4C produced using clustal omega.[18,19] The annotation `A` *indicates residues within* 12 Å *of the active site iron – essentially defining the active site. The low sequence diversity amoung these residues, or each protein generally, highlights the difficulty in designing selective inhibitors. Below the residues at each position,* ∗ *indicates absolute conservation,* : *indicates strong similarity (defined as having a score calculated with the Gonnet 250 point accepted mutation matrix above* 0.5*),* . *indicates weak similarity (groups whose score is less than or equal to* 0.5*), and no symbol indicates no conservation.[20]*

Monoclonal Antibody Recovery From Single Cells

Recovering monoclonal antibodies would be a straightforward application of techniques already familiar to cell biology were the B-cell's lifetime, isolated from its surrounding tissue, not limited to only days. A search among billions of B-cells, for the source of single clones of antigen-specific antibodies, can be made into smaller searches by diluting and dividing cells among the wells of culture plates. Only a small minority of these cells secrete antigen-specific antibodies, and so only a few wells will react to the antigen in an enzyme-linked immunosorbent assay (ELISA) or similar. The cells from these positive wells, diluted, divided, and screened repeatedly, will eventually reach monoclonality, and so each positive ELISA will identify an antigen-reactive clone. In reality, most B-cells cannot survive sufficiently long in culture to be divided and screened more than once. Therefore techniques for monoclonal antibody recovery are developed to either make the search more efficient, or relax the constraint of the B-cell's short lifetime.

The fusion of B-cell and myeloma cell imparts on the resulting *hybridoma* the myeloma's immortality, and the B-cells antibody production.[21] It is the first and most enduring solution to the problem of monoclonal antibody recovery: immortalise the B-cells so that they survive long enough to be isolated, and then provide from their culture an indefinite supply of a monoclonal antibody. If the fusion of myeloma and B-cell was not so often fatal, monoclonal antibody recovery would be solved by hybridoma. But most B-cells perish during fusion; hybridomas capture only a tiny fraction of the antibody response. The success of the technique at all is owed largely to the B-cells responsible for antigen-specific antibodies being allowed by the immune system to proliferate, and thus be supplied to the electro-fusion chamber in large excess.

The single-cell antibody recovery used here is, in many ways, inspired by hybridoma, but by avoiding the often lethal fusion step, can sample a more larger, less biased subset of an antibody response.[22,23] It starts with dilution, but not dilution until monoclonality. For that the B-cell will not

survive long enough. Instead, a compromise is made. If B-cells are diluted and aliquoted into culture wells such that only 10 % of the wells will react to the antigen, then among the antibodies in each well, those responsible for the reaction are likely the same clone.[22] In other words, though securing the monoclonality of each clone is impractical, the monoclonality of hits can, at least empirically, be guaranteed. With such a guarantee monoclonal antibody recovery becomes a problem of identifying antigen-reactive wells and then isolating any antigen-reactive B-cells from thousands per well.

Culture plates of B-cells are diluted such that each well contains thousands of cells and cultured for five days. Those wells whose antibodies recognise the antigen are located by taking a little media from each well and subjecting it to the so-called "Fluorometric Microvolume Assay" (FMAT):[24] antigen-coated beads are left to incubate in B-cell supernatants, with the antibodies they contain tagged by fluorescently labelled anti-species antibodies. Those antibodies that recognise antigen-coated beads gather at their surface, along with their labels, and in doing so appear to increase the fluorescence of a given area relative to the background.

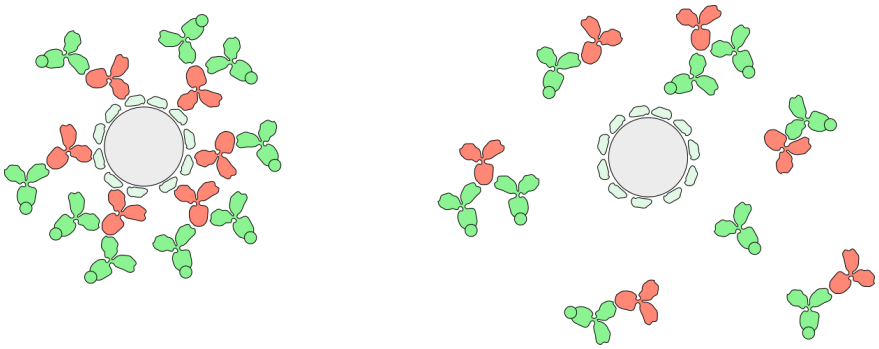


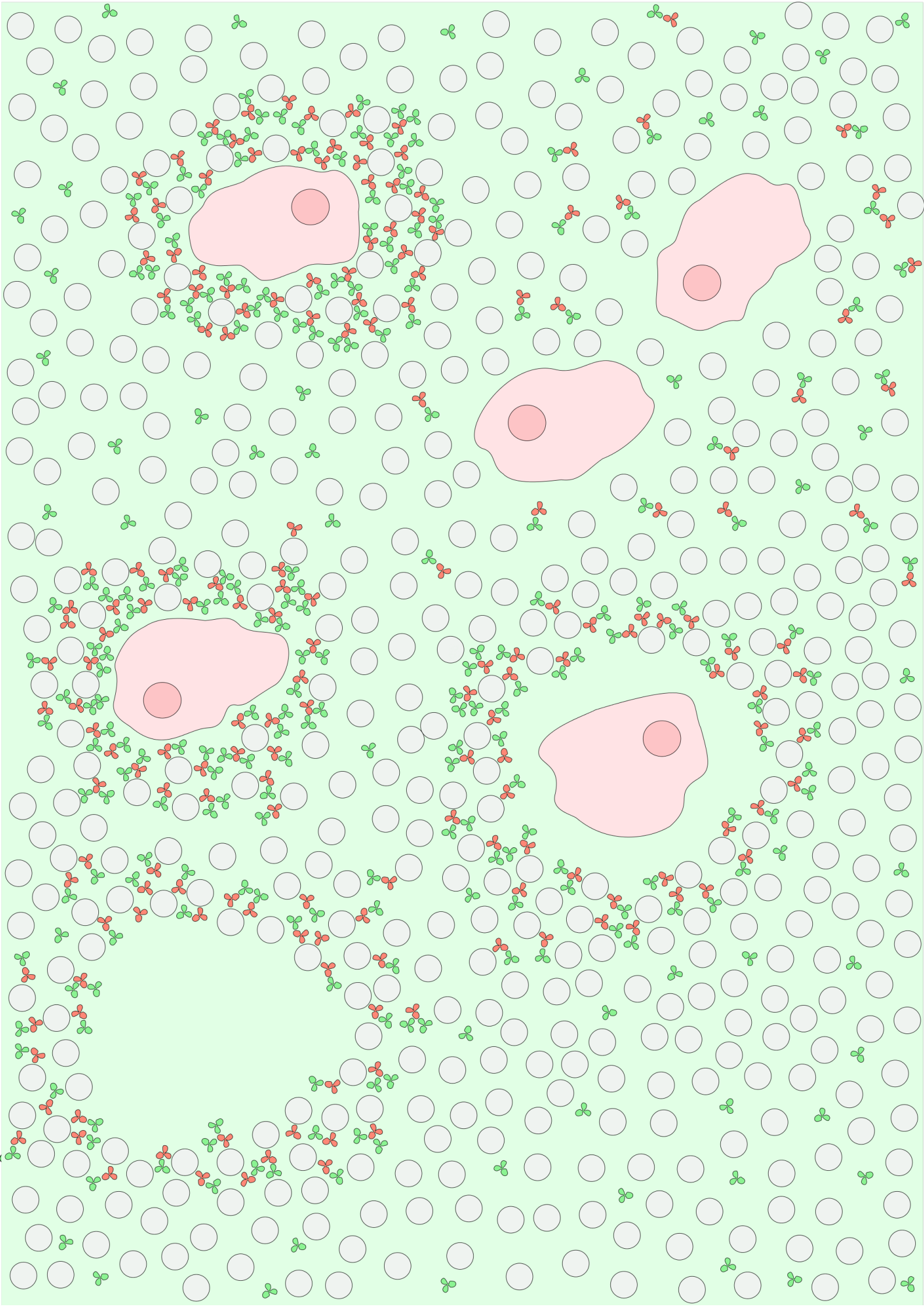
Figure 2.3: Antibodies specific for an antigen coated on beads, when labelled fluorescently, concentrate fluorescence at the bead, differentiating themselves from non-binding antibodies.

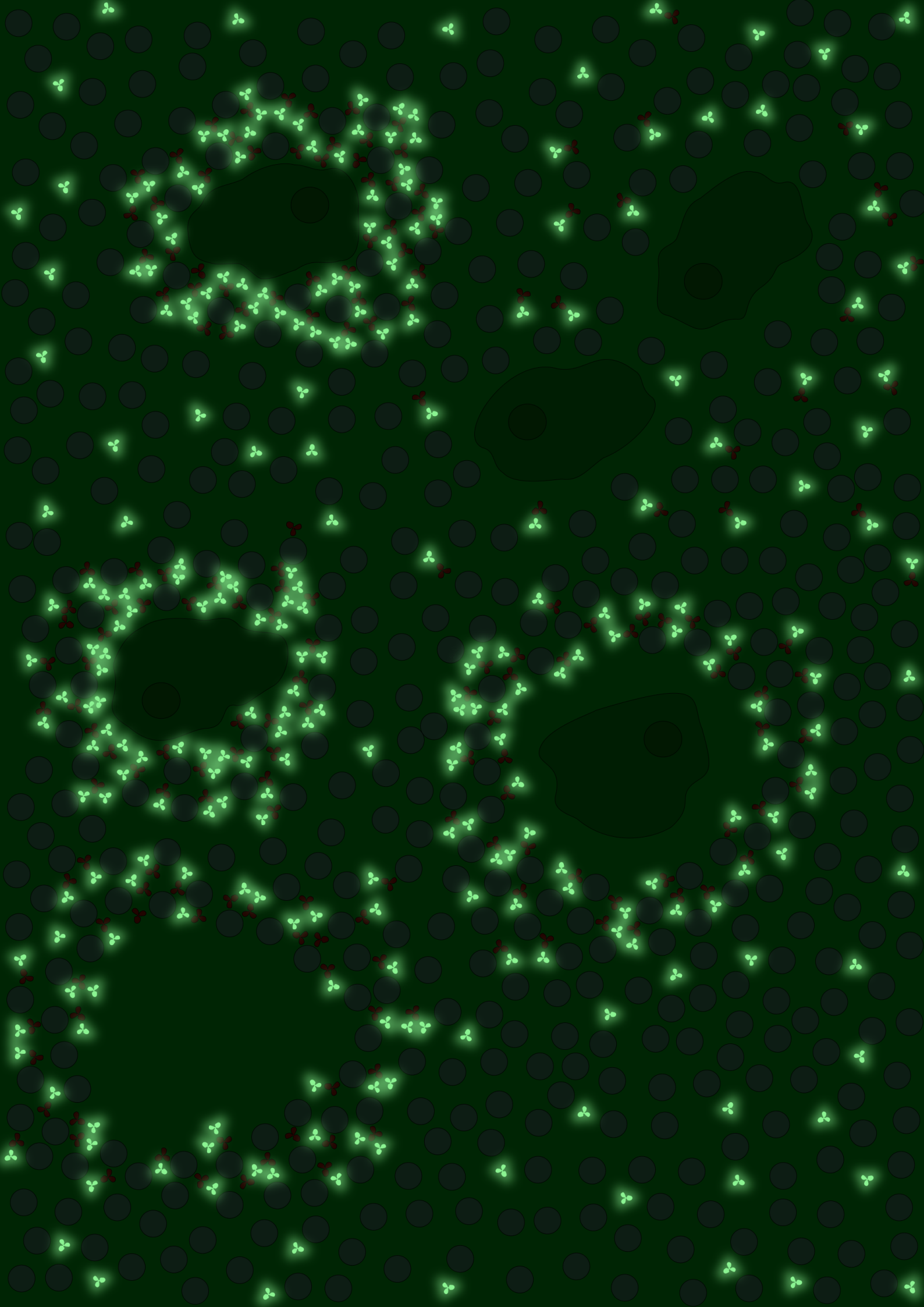
To isolate B-cells from antigen-reactive wells requires time, so the B-cells are preserved: antibody-containing supernatant is removed and stored for further screening, and the B-cells frozen awaiting isolation. Supernatants may be re-screened with the more stringent ELISA, or if it can tolerate

B-cell media, a functional assay. With a well chosen from the properties of its supernatant, the cells are thawed, washed, and given fresh media. Each again starts secreting its particular antibody clone. Some B-cells will secrete antibodies specific for the antigen, and some will not. But cultured on a microscope slide, in the presence of antigen-coated beads and fluorescently labelled anti-species antibodies the two are easily differentiated. Two reagents are needed. The first is the antigen itself. Biotinylated and coated on beads coating the microscope slide's surface, the diffusion of antigen-reactive antibodies from their source will be hindered proportional to their dissociation constant (K_{D}). Labelling all antibodies with a fluorescent-labelled species-specific antibody reveals this hindered diffusion as a halo around the source B-cell against the background fluorescence.

Once identified, an area is cleared around the cell by ejecting media from a micromanipulator, and the cell collected with the same device. The B-cell's fate is still to die – it is frozen directly in the media it is isolated in. But once isolated this is irrelevant. From a single B-cell, whether alive or dead, the reverse transcriptase PCR (RT-PCR) can convert mRNA to complementary DNA (cDNA), and from cDNA the right primers, by recognising the conserved leader sequences that demarcate antibody heavy and light chains in a B-cells genome, can amplify variable regions. Subcloned into an expression vector, the monoclonal antibody is thus recovered.

Figure 2.4: (Overpage) B-cells secreting antibodies (orange) specific for an antigen are identified with fluorescently labelled anti-species antibodies (green) and antigen-coated beads (grey). Their diffusion hindered by interactions with the antigen, specific antibodies outline their source B-cell, and when tagged with fluorescently-labelled anti-species antibodies, appear as a halo of fluorescence enclosing the cell.





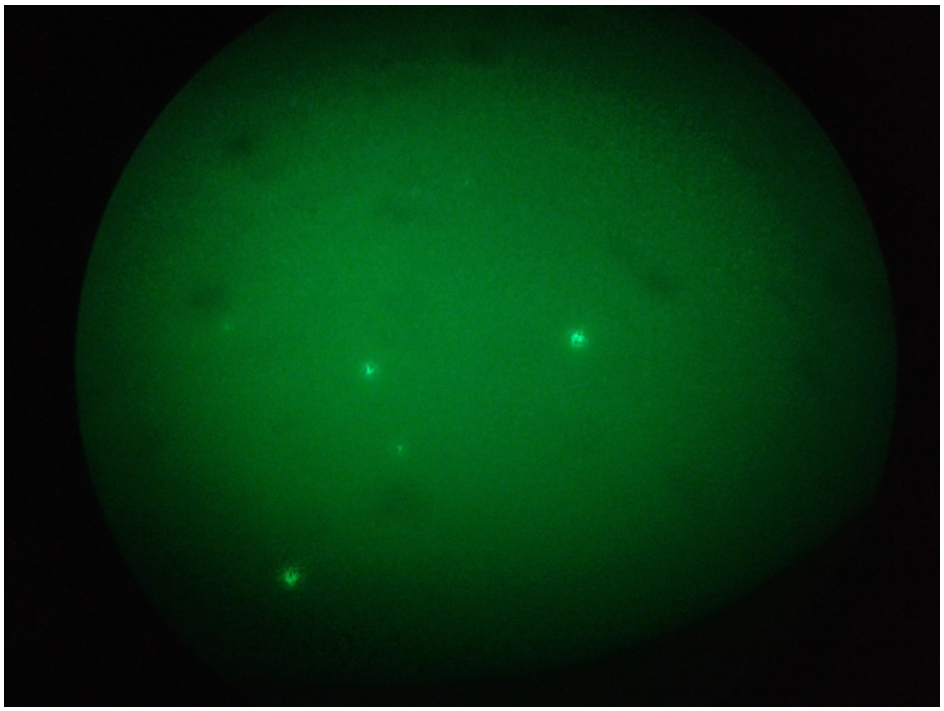


Figure 2.5: B-cells secreting antigen-reactive antibodies, incubating with antigen coated beads, and fluorescently labelled species-specific antibodies are identified by a halo of fluorescence.

2.2 Results

Preparation of KDM4D

Wildtype KDM4D has in total 523 residues. The catalytic region, containing the JmjN and JmjC domains, is covered in the first two-thirds of the sequence. It is the catalytic activity that ultimately was to be inhibited, so in order than antibodies be recruited to the responsible regions, and not distracted elsewhere, the KDM4D construct used for immunisation, $KDM4D_{1–358}$, ends soon after the JmjC domain. Expressed in *Escherichia coli* (E. coli) with a tobacco etch virus (TEV)-cleavable His_6-tag, the 20 mg required for immunisation of two rabbits came from a 6 L culture in lysogeny broth (LB). So that KDM4D could be immobilised on streptavidin-coated surfaces, the construct was also subcloned to append at the C-terminal an "avi-tag", and the biotinylated protein purified.

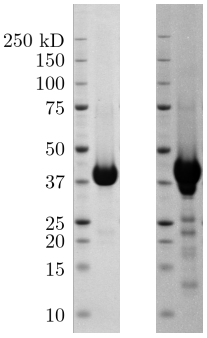


Figure 2.6: An SDS-PAGE gel showing (left) $KDM4D_{1–358}$ and (right) $KDM4D_{1–358}$ biotinylated at the C-terminal after recombinant expression in E. coli and purification.

Activity Assay

The $KDM4D_{1–358}$ purified here was enzymatically active, demethylating twice, at lysine 9, a 15mer derived from the histone 3 tail ($H3K9me3_{1–15}$) in the manner previously observed.[1] The loss of a methyl group from H3K9me3, and again from the product, H3K9me2, are reflected in mass spectra. So-called "RapidFire" mass spectrometry was used here. The main feature that differentiates it from conventional mass spectrometry is the replacement of a chromatography column with a solid phase extraction cartridge. Without the comparatively slow chromatography step, rapid analy-

sis of samples is permitted, and so it is attractive for small-molecule KDMs inhibitor discovery.[25,26] Using mass spectrometry also avoids alternative KDM activity assays which rely on coupled enzymes, such as formaldehyde dehydrogenase (FDH), or indirect readouts such as "AlphaScreen".[27]

The reaction of KDM4D with H3K9me3/2 proceeds satisfactorily at room temperature, with 2OG and H3K9me3 at 10 µM (their Michaelis constant (K_m)),[1,28] and iron (II) ($(NH_4)_2Fe(SO_4)_2 \cdot 6\,H_2O$) at 50 µM protected from oxidation by a two-fold molar excess of L-ascorbic acid (100 µM). KDM4D seemed particularly sensitive to the NaCl concentration – anything over 150 mM was enough to ablate demethylation completely. So besides the aforementioned co-factors and anti-oxidant the assay buffer contained only 2-(N-morpholino)ethanesulfonic acid (MES), buffered to pH 7.0. Although used here at the limit of its buffering range, MES is preferable to alternatives such as 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and tris(hydroxymethyl)aminomethane (TRIS) as it had little effect on substrate and product peptide signals in mass spectra whereas other agents can be strong ion suppressants due to solid phase extraction not efficiently removing the sample buffer before analysis. Under these conditions, that is, with cofactors at concentrations low enough to screen inhibitors, a KDM4D concentration of at least 300 nM was necessary for sufficient turnover.

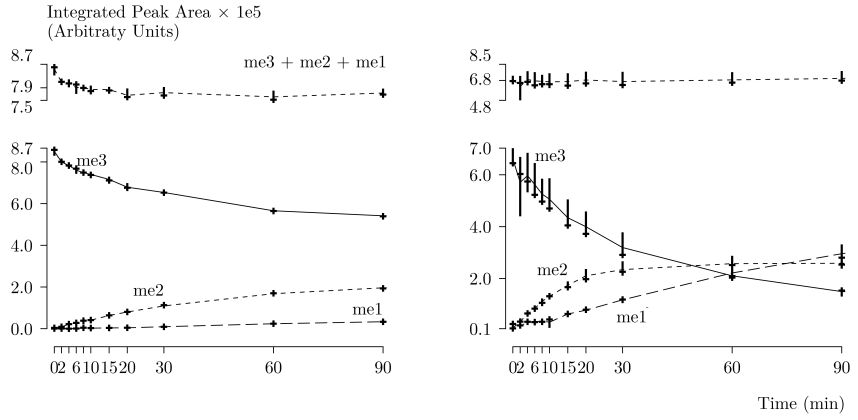


Figure 2.7: H3K9me3 demethylation by KDM4D at (left) 100 nM *and (right)* 300 nM*. Above each plot is the sum of all signals (the substrate and the two products) which is theoretically constant.*

Antibody Discovery

Two female New Zealand white (NZW) rabbits were immunised with KDM4D (residues 1–358), and B-cells from their lymph nodes mixed and cultured using the single-cell technique presented earlier. Approximately 5000 cells were aliquoted into each well of 300 96-well plates, and cultured for 5 days. Of these wells, the supernatants from 21 % reacted to KDM4D in an FMAT assay. This was far too many antibodies to isolate, so 1056 with the highest fluorescence, were selected to be frozen, and their supernatants preserved in 12 96-well plates (each with one blank column of controls) for further screening.

Ideally, function-modifying antibodies would be identified directly from the supernatants into which they are secreted. The supernatants would be assayed like any other potential KDM inhibitor, and specific antibodies from any that affected KDM4D's enzymatic activity recovered. Unfortunately, the concentration of antibodies, in the supernatant of B-cells cultured for five days, is extremely low, much lower than the 300 nM concentration of KDM4D needed to see sufficient demethylation. It is therefore difficult to detect antibody-induced changes in KDM4D's activity, even if the assay were carried out directly in the supernatant. This is itself difficult, since the chelating agents that prevent cells clumping also sequesters the iron (II) that must occupy KDM4D's active site for activity.

Rather than isolate antibodies without guidance, and risk repeatedly isolating dominant antibody families of little functional consequence, each well was categorised by its cross-reactivity toward a panel of KDMs. Each group represents at least one unique underlying epitope. Isolating antibodies of diverse specificity increases the chance of recovering diverse sequences, and so too the chance of identifying a function modifying antibodies. ELISA is best suited for this kind of assay: stringent washing steps prevent interference from B-cell media, and the enzyme turnover on which the assay relies amplifies the binding signal from relatively few antibodies. A panel of KDMs was prepared containing the remaining KDM4 family members:

4A, 4B, and 4C; PHF8 family members lysine demethylase 7A (KDM7A) and PHF8; lysine demethylase 2A (KDM2A); and lysine demethylase 5A (KDM5A).

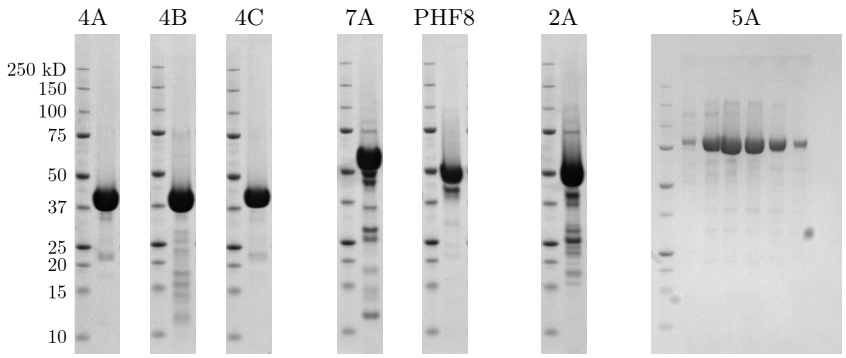


Figure 2.8: SDS-PAGE gels showing KDM family members purified to categorise antibodies by their cross-reactivity.

The bands immediately following the major, sometimes decreasing in discrete steps, are too intense to be a natively expressed protein and are a sign of degradation. Such cleavage (oxidative or proteolytic) has been reported for 2OG oxygenases previously.[29] However, minor degradation is unlikely to seriously affect selectivity screening.

Of the 1056 B-cell supernatants stored for further screening, ELISA confirmed 939 contained antibodies able to recognise KDM4D. The cross-reactivity of these 939 supernatants divided each into 51 categories of specificity. Those supernatants whose antibodies were specific only for KDM4D are the second-largest category, but comprise merely 11 % of the total hits – a reflection of just how homologous KDMs are, and why selectivity can be such an issue for inhibitors. Wells from different categories among the 51 previously identified were selected, thawed, and B-cells isolated using the fluorescent foci method introduced earlier. From each well 6 to 8 B-cells were isolated, and the antibody-encoding DNA from two of these cells cloned as a fragment antibody bindings (Fabs). In total, 23 antibodies were recovered. The antibody isolation process is summarised below.

1. **2** rabbits immunised with $KDM4D_{1–358}$

2. **300** plates (**28,800 wells**) incubated for **5** days
3. **21 %** of supernatants reacted to KDM4D in FMAT assay
4. **1056** wells preserved and stored
5. ELISA confirmed KDM4D specificity of **939** of these wells
6. Cross-reactivity towards **9** KDMs revealed **51** specificity categories
7. **304** B-cells isolated by micromanipulator
8. **23** Monoclonal antibodies cloned as Fab fragments[b]

[b]Fab fragments were chosen as the format for antibody expression due to their stability and high recombinant expression in mammalian cells, as well as their crystallographic suitability.

KDM Selectivity									
4D	4A	4B	4C	5A	2A	7A	PHF8	Bin Size	Percent
•		•	•					135	14.4
•								107	11.4
•			•					75	8.0
•	•	•	•				•	50	5.3
•		•						42	4.5
•							•	32	3.4
•	•	•						29	3.1
•					•		•	24	2.6
•	•							21	2.2
•		•	•				•	18	1.9
•			•				•	15	1.6
•		•					•	13	1.4
•		•	•		•		•	12	1.3
•	•		•					11	1.2
•	•	•	•			•		10	1.1
•		•			•		•	10	1.1
•	•	•					•	9	1.0
•		•	•		•			8	0.9
•	•	•	•	•				8	0.9
•	•						•	7	0.7
•	•	•	•		•		•	7	0.7
•					•			6	0.6
•	•				•		•	6	0.6
•	•	•	•			•	•	5	0.5
•	•		•		•		•	5	0.5
•		•	•	•				5	0.5
•	•	•	•		•			4	0.4
•		•	•			•		3	0.3
•	•	•	•	•			•	3	0.3
•	•	•	•	•	•	•	•	3	0.3
•	•		•				•	2	0.2
•			•			•		2	0.2
•		•			•			2	0.2
•	•	•			•			2	0.2
•	•	•			•		•	2	0.2
•			•		•		•	2	0.2
•				•				2	0.2
•				•			•	2	0.2
•			•	•			•	2	0.2
•	•	•	•	•	•		•	2	0.2
•						•		1	0.1
•	•				•			1	0.1
•			•		•			1	0.1
•	•	•	•		•	•		1	0.1
•	•				•	•	•	1	0.1
•		•			•	•	•	1	0.1
•		•		•				1	0.1
•		•		•			•	1	0.1
•	•	•	•	•		•	•	1	0.1
•				•	•		•	1	0.1
•		•		•	•		•	1	0.1

Table 2.1: The cross-reactivity of 939 *antibody-containing supernatants to a panel of eight demethylases determined by ELISA using a threshold five-times the background signal from a control well containing no antigen.*

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
J-16	SGFSLSSYSMN	LMHTGGSAYYAS	G--------------NI	QASQSVYKKNYLA	YASTLDS	QGGYDCSSADCN--A
J-25[a]	SGIDLSSYAMN	FIYTDGRTYYAS	G--------------NL	QSSQSFY-SGYLS	SASTLAS	AGRYSATEA-----T
J-26	SGIDLSSYAMN	FIYTDGRTYYAS	G--------------NL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN------T
J-25[b]	SGIDLSSYAMN	FIYTDGRTYYAS	G--------------NL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN------T
J-24	SGIDLSRYGVN	FIYTDGRAYYTN	G--------------DL	QASQSVYNGNYLS	YASTLDS	QGAYDCGSADCN--A
J-12	SGIDLSSYAMN	FIYTNGRTYYAS	G--------------DL	QSSESVYNNNYLA	YTSTLDS	QGAYSGTN------T
J-7	SGFSLSSYHIC	VISSSGTTYYAN	G--------------DL	QSSQSVV-NNNLA	KASTLAS	LGSYDCSSTDCN--A
J-13	SGFSLSSYHIC	VISSSGTTYYAN	G--------------DL	QSSQSVV-NNNLA	KASTLAS	LGSYDCSSTDCN--A
J-11	SGIDLSTNGVN	YIYTDRSTYYAT	G--------------DI	QASQSVYKNNYLA	YASTLDS	QGGYDCSRADCN--V
J-20	SGIDLSSYGMN	YIYTDGNTYYAY	G--------------DI	QASQSVYKNNYLA	YTSTLAS	QGAYDCSRADCN--V
J-17	SGIDLSSNGMN	YIYTDSSTYYAS	G--------------DI	QASQSVYNNNYLA	YASTLDS	QGGYDCSRADCN--V
J-23	SGFSLSIFAMN	IISNDGGTYYAS	RLAGYAGYGYATFYFNI	RASEDI--ERYLA	DASDLAS	QSGYYSNSADF---P
J-5	SGFSLSSYYMT	IINSGGSAYYAS	DRYNNIDSG----TYNF	QASQTIY--NFLS	GASNLES	QSNYGSSSNNYGN-A
J-14	SGFSLSSYSMI	IIGGSGSTYYAT	GYNAGVAGT----YFNI	QASQSVYKNNYLA	TASKLPS	LGSYDCTSADCH--A
J-8	SGIDLSSYLMS	GISSSGRTKYAT	GPGDSTGRA-----FDP	QASQSVYNNENLA	YASTLAS	QGEFSCSSADCT--A
J-3	SGFSLSSYYMT	IMGSSGNVYYAS	GM-VGIGVY----SLNI	QASQSVYNNKNLA	AASYLAS	QGEFSCSSGDCS--A
J-6	SGFSLNSYAMS	WINGGTSTYYAS	NIIATNSGG------GL	QASQSVYSNKNLA	AASNLAS	AGEFSCSTADCT--A
J-2	SGFSLNSYHMT	IISSSGNTYYAN	YIGWNFGS-------DL	QASQSI--NDYLS	KASDLAS	QSNYYSSSSDDGALA
J-1	SGFSLSSYSIY	IIRPSGRTYYAS	GY---PGII----SDNI	QSSQSVYKKNYLA	LASTLAS	LGGYDDDVDN----G
J-4	SGIDLSSYAMG	WIIPNNKVAYAS	-----CGAG----AFNL	QSSQSVYNNNNLA	AASTLAS	QGYYSGGIG-----A
J-22	SGFSLSNYDMS	SISAGGNTYYAS	G------AV----SPTI	RSSQSVYSDNLLS	QASSLAS	LGVYDDVDN-----A
J-19	SGFSLSTYAVS	AIGNSGNTYYTS	G------GY----AASI	KTSQSLYNNNCLA	EASKLAT	AGDFGAGIS-----A
J-18	SGFSLSRYGVT	YIHTDSNTYYAS	I------GS------GI	QASQSVYKNNYLS	SASNLEI	LGSYDCNSANCF--S

Table 2.2: Sequences for KDM4D antibodies arranged by heavy chain CDR3 sequence.

The 23 antibodies that were recovered as Fabs, were cloned with a C-terminal His_{10} tag, and so they could be purified, by nickel-affinity chromatography followed by gel filtration.

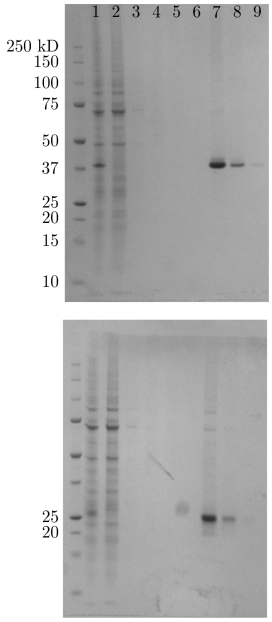


Figure 2.9: SDS-PAGE gels from the nickel-affinity purification of antibody J5 as a Fab fragment. Lane 1 is the input, 2 the flow through, 3 a wash with PBS, 4–6 washes with 30 mM *imidazole, and 7–9 elution with* 300 mM *imidazole. Non-reducing and reducing conditions are shown on top and bottom, respectively.*

Those antibodies whose sequences were unique, and expressed sufficiently as recombinant proteins in human embryonic kidney 293 (HEK293) cells, were purified and their cross-reactivity reconfirmed:

			KDM Selectivity							
	CDR-L3	**CDR-H3**	4D	4A	4B	4C	5B	2A	7A	PHF8
J-11	QGGYDCSRADCN-V	G----------DI	•					•		
J-24	QGAYDCGSADCN-A	G----------DL	•					•		
J-12	QGAYSGTN-----T	G----------DL	•					•		
J-25[b]	QGAYSGTN-----T	G----------NL	•					•		
J-26	QGAYSGTN-----T	G----------NL	•					•		
J-3	QGEFSCSSGDCS-A	GM-VGIGVY----SLNI	•							
J-4	QGYYSGGIG----A	-----CGAG----AFNL	•							
J-16	QGGYDCSSADCN-A	G----------NI	•							
J-23	QSGYYSNSADF--P	RLAGYAGYGYATFYFNI	•	•	•	•				
J-6	AGEFSCSTADCT-A	NIIATNSGG------GL	•	•	•	•				
J-18	LGSYDCNSANCF-S	I------GS------GI	•	•					•	
J-22	LGVYDDVDN----A	G------AV----SPTI	•		•	•				
J-1	LGGYDDDVDN---G	GY---PGII----SDNI	•	•						
J-5	QSNYGSSSNNYGNA	DRYNNIDSG----TYNF	•		•					

Table 2.3: CDR3 sequences and cross-reactivity of the 14 *KDM4D antibodies isolated from B-cells and expressed recombinantly.*

Among the 23 antibodies recovered for KDM4D there is a family[c] that has avoided even the categorisation of each supernatant's cross-reactivity. It is almost certainly dominant in the immune response. The specificity of this family (KDM4D and KDM2A) is counted only six times in the cross-reactivity of all supernatants (table 2.1), so their repeated isolation is likely due to the initial hit-rate being slightly above the empirically determined 10 % that makes monoclonality of hits likely. Nevertheless, this family of antibodies aside, the sequences of the other 12 antibodies are unique, and provided a satisfactory place to begin inhibitor screening.

Inhibition Assay

In the mass-spectrometry-based activity assay previously shown, a KDM4D concentration of 300 nM was the lowest that produced sufficient turnover for inhibitor screening. During the 15 min while the reaction velocity remains approximately linear, and thus can best detect inhibition, the monomethylated product is not abundant to add more than noise to any calculation, and so the turnover was calculated as me2/(me3 + me2). With only solid phase extraction, high protein concentrations can suppress the signal of smaller peptides, so antibodies were screened at a concentration of 1 µM, giving a three-fold molar excess on KDM4D.

[c] In the absence of any more than sequence information, here and in the following chapters, antibody "family" refers to groups of antibodies that share heavy chain CDR3 sequences in which the residues are functionally conserved. Though not always true, generally the heavy chain CDR3 is involved in binding, so is was used to provide a rough measure of potential binding diversity.

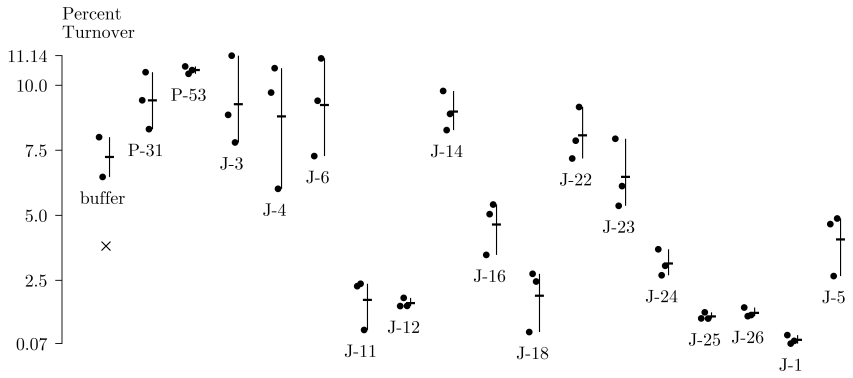


Figure 2.10: Turnover of H3K9me3$_{1-14}$ by KDM4D in the presence of antibodies. Triplicate measurements are shown as dots with vertical lines indicating their range, and horizontal lines their mean. Apparent outliers are marked by a cross.

Turnover without any antibodies present was approximately 7.5 %, and with P-31 and P-53, two antibodies that do not recognise KDM4D (chapter 3), it was approximately 11 %. In the presence of J-11, J-12, J-18, J-25, J-26, and J-1 turnover is reduced, the average not more than 2.5 %. J-18, probably mistaken for J-16 – an inhibitor – was later found not to inhibit, and so the final group of inhibiting antibodies comprised:

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
J-1	SGFSLSSYSIY	IIRPSGRTYYAS	GYPGIISDNI	QSSQSVYKKNYLA	LASTLAS	LGGYDDDVDN--G
J-12	SGIDLSSYAMN	FIYTNGRTYYAS	G-------DL	QSSESVYNNNYLA	YTSTLDS	QGAYSGTN----T
J-25[b]	SGIDLSSYAMN	FIYTDGRTYYAS	G-------NL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN----T
J-26	SGIDLSSYAMN	FIYTDGRTYYAS	G-------NL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN----T
J-11	SGIDLSTNGVN	YIYTDRSTYYAT	G-------DI	QASQSVYKNNYLA	YASTLDS	QGGYDCSRADCNV
J-16	SGFSLSSYSMN	LMHTGGSAYYAS	G-------NI	QASQSVYKKNYLA	YASTLDS	QGGYDCSSADCNA

The GDI Family

Five of six antibodies identified KDM4D inhibitors belong to a single family, characterised by a three-residue heavy-chain CDR3, one of which they derive their name from. The GDI family contains, in total, 11 members.

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
J-11	SGIDLSTNGVN	YIYTDRSTYYAT	GDI	QASQSVYKNNYLA	YASTLDS	QGGYDCSRADCNV
J-17	SGIDLSSNGMN	YIYTDSSTYYAS	GDI	QASQSVYNNNYLA	YASTLDS	QGGYDCSRADCNV
J-20	SGIDLSSYGMN	YIYTDGNTYYAY	GDI	QASQSVYKNNYLA	YTSTLAS	QGAYDCSRADCNV
J-16	SGFSLSSYSMN	LMHTGGSAYYAS	GNI	QASQSVYKKNYLA	YASTLDS	QGGYDCSSADCNA
J-25[b]	SGIDLSSYAMN	FIYTDGRTYYAS	GNL	QSSQSFY-SGYLS	SASTLAS	AGRYSATEA---T
J-26	SGIDLSSYAMN	FIYTDGRTYYAS	GNL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN----T
J-25[a]	SGIDLSSYAMN	FIYTDGRTYYAS	GNL	QSSESVYKNNYLA	YASTLDS	QGAYSGTN----T
J-12	SGIDLSSYAMN	FIYTNGRTYYAS	GDL	QSSESVYNNNYLA	YTSTLDS	QGAYSGTN----T
J-24	SGIDLSRYGVN	FIYTDGRAYYTN	GDL	QASQSVYNGNYLS	YASTLDS	QGAYDCGSADCNA
J-7	SGFSLSSYHIC	VISSSGTTYYAN	GDL	QSSQSVV-NNNLA	KASTLAS	LGSYDCSSTDCNA
J-13	SGFSLSSYHIC	VISSSGTTYYAN	GDL	QSSQSVV-NNNLA	KASTLAS	LGSYDCSSTDCNA

Besides substitution of analogous residues, the only difference between GDI family members is a deletion in the light chain of J-12, J-25[a], J-26, and J-25[b]. That said, a tyrosine residue, frequently used to donate a hydrogen bond,[30] is conserved in the light chain CDR3 of GDI family members. The selectivity of the family is primarily towards KDM4D, with what appears to be weak affinity for KDM2A (unless KDM2A immobilises on plastic significantly worse than KDM4D). 1 L-scale purification of representative GDI family members yielded high purity samples of recombinant Fab that could be used for activity and structural studies.

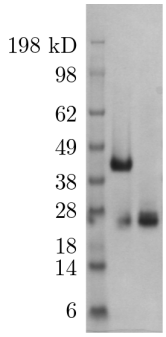


Figure 2.11: SDS-PAGE gel showing the antibody J11 as a Fab under (left lane) reducing and (right lane) non-reducing conditions.

pIC_{50} values were used here to confirm inhibition of KDM4D by selected antibodies. Again the "Rapidfire" mass-spectrometry assay was used, but the high antibody concentration required at the dose-response curve's extreme required the assay had to be left longer before quenching and so increase the substrate peptides signal. The effect, though not serious, is that IC_{50} values are potentially underestimated, as the approximation that reaction velocity is linear becomes less accurate with time. A ratio of integrated area corresponding to substrate and products was used to calculate turnover so that two samples suffering differing ion suppression are comparable. Rather than introduce controls to estimate the background captured by integrating ion chromatogram peaks, the measured values were used, and a four-parameter logistic regression fit to the data using the *scipy* python library.[31]

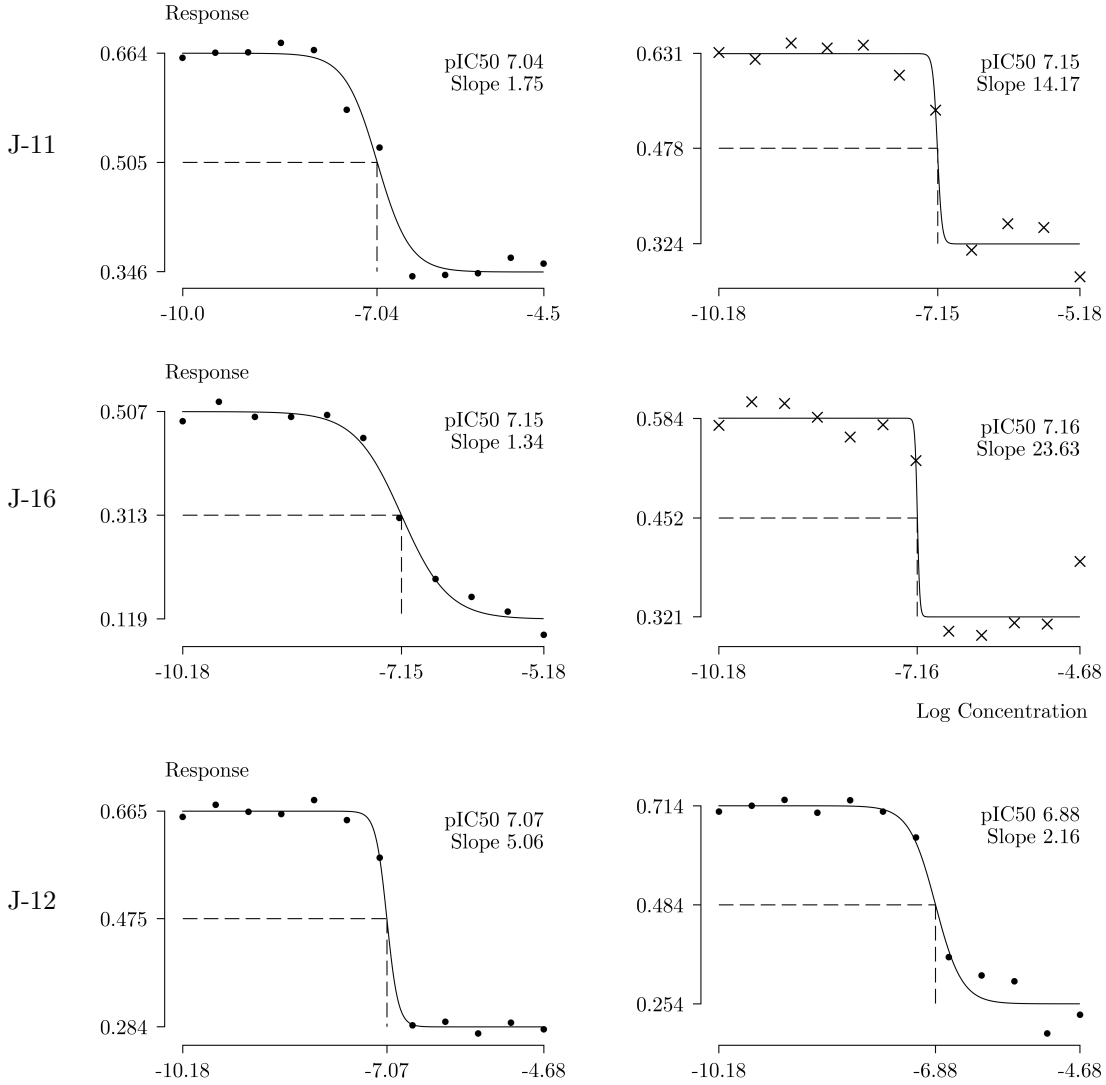


Figure 2.12: Dose-response curves for three members of the GDI antibody familiy. Datasets in which the transition was too steep to allow an accurate fit of the data to a four-parameter logistic regression are denoted with crosses. The concentration of KDM4D was 300 nM.

The transition between full activity and inhibition is swift for each of the antibodies. Although steep Hill slope's carry negative connotations, they can also indicate allosteric inhibition or very tight binding.[32] For J-11 and J-16 the average pIC_{50} values were 7.10 and 6.97, respectively. And for J-12, whose light chain CDR3 sequence differs slightly the pIC_{50} value was 7.16. Perhaps most significant is that they correspond to concentrations in the range 69 nM to 107 nM – less than half of the 300 nM KDM4D concentration used in the assay. That these values are below the tight binding limit may suggest that not all molecules of KDM4D are (at least concurrently) enzymatically active.

The argument that only a subpopulation of recombinant KDM4D samples is active gains more support when the family's binding kinetics are considered. Measured by surface plasmon resonance (SPR), J-11 and J-12, whose light chain CDR3 is slightly longer than that of J-16, had the highest affinity for KDM4D, though all retain a subnanomolar K_D.

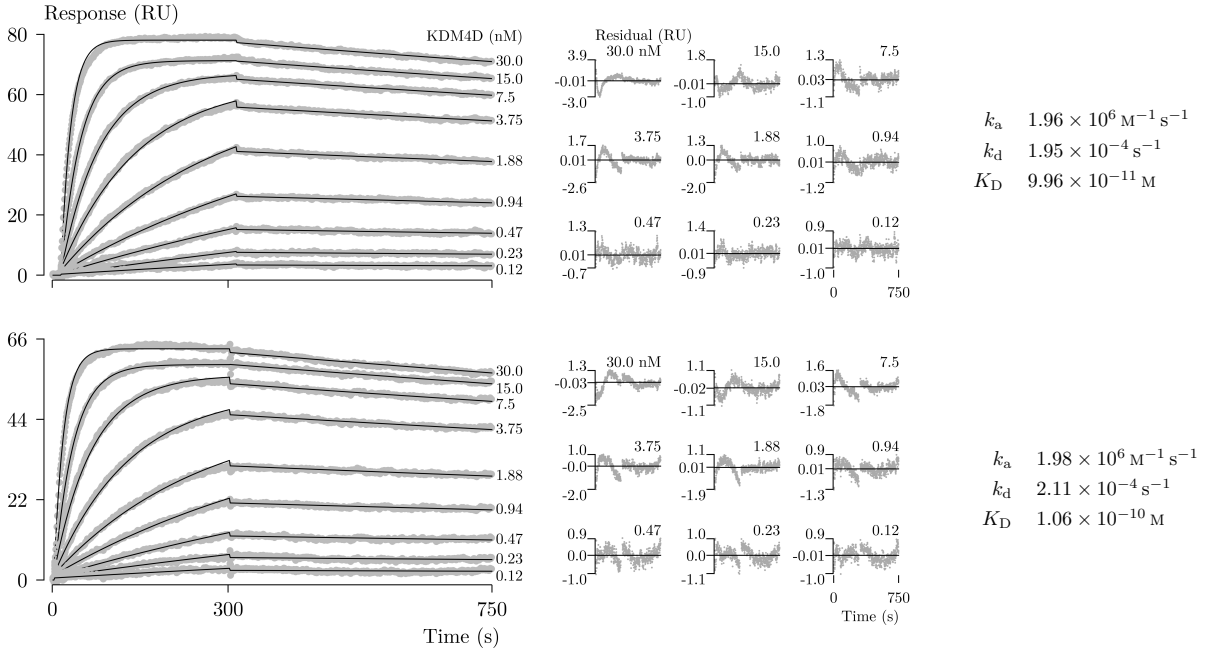


Figure 2.13: Duplicate determinations of the affinity of antibody J-11 for $KDM4D_{1–358}$ by SPR. Fabs were captured by surface-immobilised anti-rabbit Fab antibodies, and $KDM4D_{1–358}$ was injected at varying concentrations as the mobile phase.

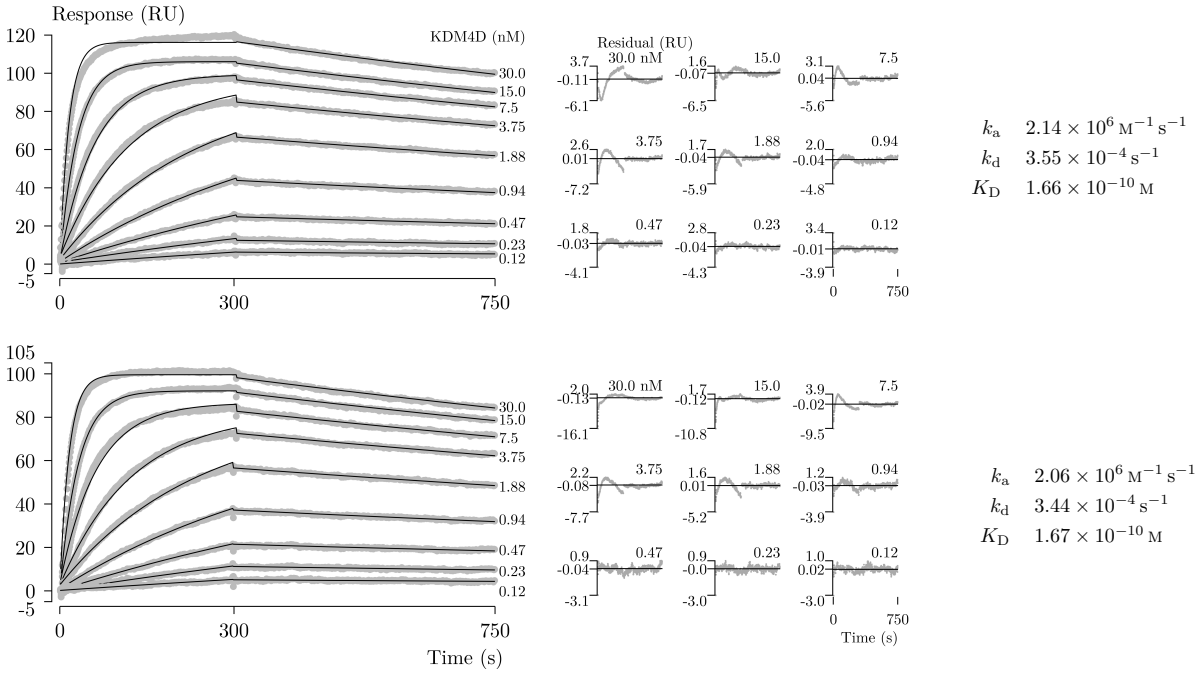


Figure 2.14: Duplicate determinations of the affinity of antibody J-16 for KDM4D$_{1–358}$. Fabs were captured by surface-immobilised anti-rabbit Fab antibodies, and KDM4D$_{1–358}$ was injected at varying concentrations as the mobile phase.

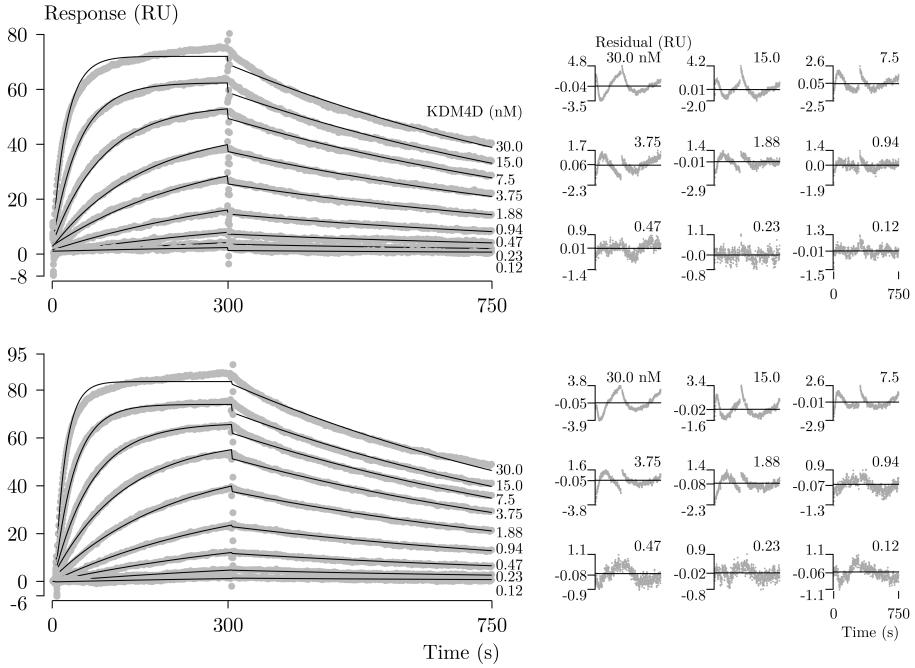


Figure 2.15: Duplicate determinations of the affinity of antibody J-12 for $KDM4D_{1–358}$. Fabs were captured by surface-immobilised anti-rabbit Fab antibodies, and $KDM4D_{1–358}$ was injected at varying concentrations as the mobile phase.

It appears from these SPR traces that the effect of light chain CDR3 deletions is most pronounced in the off-rate. It is increased by almost an order of magnitude, and reflected in a K_D that increases approximately seven-times.

Antibody	k_a ($M^{-1}\,s^{-1}$)	k_d (s^{-1})	K_D (M)
J-11	$1.97 \times 10^6 \divideontimes 1.01$	$2.03 \times 10^{-4} \divideontimes 1.04$	$1.03 \times 10^{-10} \divideontimes 1.03$
J-16	$2.10 \times 10^6 \divideontimes 1.02$	$3.49 \times 10^{-4} \divideontimes 1.02$	$1.66 \times 10^{-10} \divideontimes 1.00$
J-12	$1.58 \times 10^6 \divideontimes 1.16$	$1.29 \times 10^{-3} \divideontimes 1.01$	$8.18 \times 10^{-10} \divideontimes 1.15$

Table 2.4: Kinetic characterisation of GDI family antibodies towards $KDM4D_{1–358}$ by SPR. Values are indicated with their log-normal standard deviation.

Given the above kinetics, at the concentrations under which dose-response curves were recorded, there would be few unbound antibodies. Hence, the transition from activity to inhibition is sharp: over the antibody concentrations 31.3, 99, and 313 nM KDM4D transitions from essentially active to totally bound by the inhibitor. Since the pIC_{50} values are below the expected tight binding limit, and their engagement by antibodies is almost guaranteed, it implies KDM4D activity is produced by a subset of molecules. It could be that some are purified inactive, or deteriorate towards inactivity. But for several weeks the same sample of KDM4D was frozen and thawed repeatedly for activity testing with no obvious change to its activity. Thus, the results could support the existence of at least two different populations of enzymatically active and inactive KDM4D.

The GDI Family's Epitope

The value of these antibodies that bind intracellular proteins is in what constraints they can give the search for small molecule inhibitors. Extensive crystal screening, however, failed to return any hits. J-11, J-12, and J-16 were screened at both 4 °C and 18 °C in complex with the immunisation construct without result. The majority of KDM4D crystal structures contain a metal ion, and 2OG at the active site – a sign they may stabilise the protein. The addition of $NiSO_4$ to J-11 and KDM4D complexes, however,

caused considerable precipitation. Crystal trials on what remained after removing the precipitated material yielded little, as did the the addition of 2OG to samples of the complex before crystal trials.

The system that came closest to crystals was a GDI family member reformatted as a single-chain variable fragment (scFv). In complex with KDM4D there were no hits, but alone it sometimes formed a gel. A grid screen on the gel condition, produced small needles, but when crushed and used as seeds in new crystal trials they failed to produce more than crystalline precipitate.

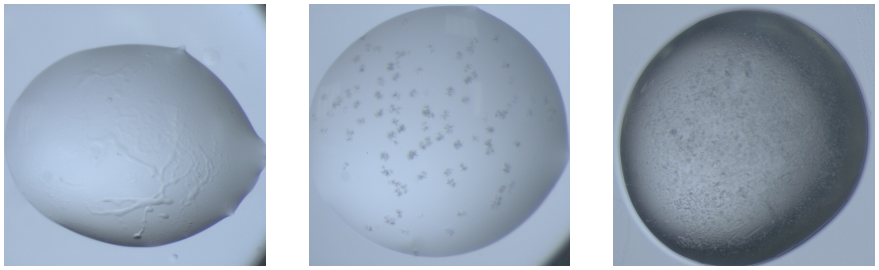
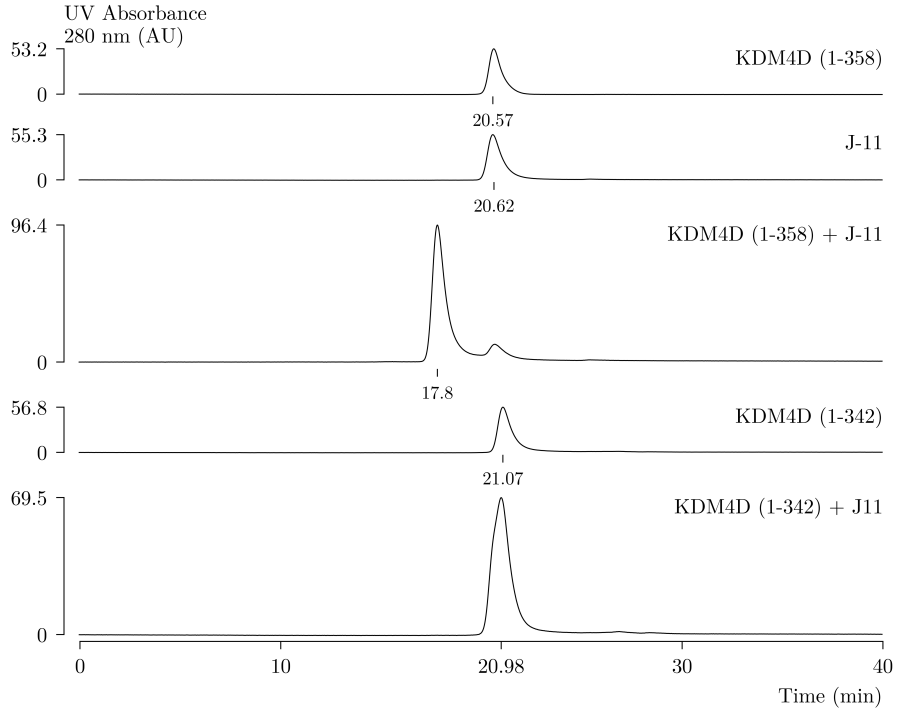
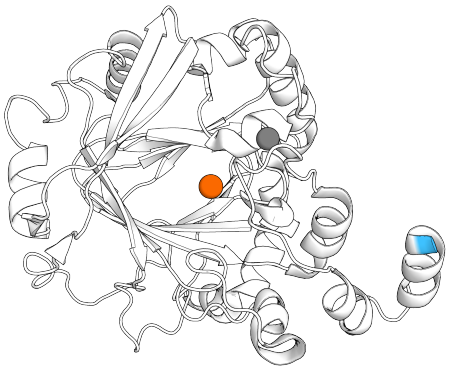


Figure 2.16: (Left) A gel, subsequently optimised by grid screen produced small needles (middle), but when crushed and used to seed a third optimisation did not help crystal formation.

Despite the aversion of GDI-family antibodies to crystallisation either alone or in complex with KDM4D, their binding properties still permit a unique reflection on the nature of KDM4D's enzymatic activity: the antibody J-11 will only bind to KDM4D when residues 342–358 are present. The construct used for immunisation was $KDM4D_{1–358}$, though that used for crystallisation in the literature only contained residues 1–342. Without these 16 residues, J-11 does not bind even at concentrations exceeding 10 µM used in size exclusion chromatography.



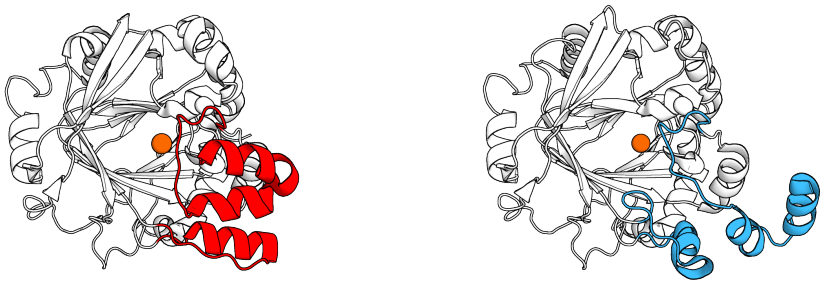


	342	358
KDM4D	drAVVDHMEPRVPASQEL	
KDM4A	dnTVIDHTLPTPEAAEFL	
KDM4B	dlTVLDHTRPTALTSPEL	
KDM4C	diYTIDHTKPTPASTPEV	

Figure 2.17: Size exclusion chromatograms, using an S200 column, show $KDM4D_{1–358}$ can form a complex with J-11, and induce the expected shift in retention time, but $KDM_{1–342}$ and J-11 cannot. The concentration of J-11 was 11 µM*, and the concentration of $KDM4D_{1–358}$ and $KDM4D_{1–342}$,* 10 µM*. The running buffer was HBS with NaCl at* 250 mM*. A crystal structure of KDM4D (PDB: 4HON) highlights residue 340, the last residue resolved at the C-terminal, in blue. The sequence from residue of the 16 residues removed between 342 and 358 and a comparison with the aligned regions in KDM4A, KDM4B, and KDM4C suggests that relatively little similarity in these region explains the selectivity of the GDI family.*

These size exclusion chromatograms are, without a crystal structure, perhaps the strongest evidence that the GDI family's epitope lies between residues 342–358. J-11, an antibody whose affinity was determined by SPR to be sub-nanomolar, appears to not have even residual affinity for KDM1–342 in a size exclusion chromatography (SEC) assay at micromolar concentrations. Residues 342–358 of KDM4D, the GDI family's putative epitope, is distal to the active site. In fact, it is as far from the active site as the immunisation construct will allow. And yet, it is an inhibitor.

Reported crystal structures for KDM4D, which currently reveal only the catalytic N-terminal region, capture two conformations. The first (*PDB: 3DXU*)[d], and some 260 others, show KDM4D with the C-terminal pointing toward the active site. But two structures, *4HON* and *4HOO*, have the same region pointing outwards.[1]



inward *outward*

Figure 2.18: Two conformations of KDM4D (left) 3DXU with the C-terminal oriented toward the active site (indicated by Fe(II) in orange), and (right) 4HON with same the region directed outward.

The biochemical implications of these two KDM4D conformation are not clear. Whether the peptide substrate is present or not, the "outward" conformation can be adopted (*4HON* and *4HOO* were crystallised with and without the substrate peptide respectively).[1] And the with or without Fe(II) and 2OG KDM4D can crystallise in the closed conformation. Whether KDM4D can adopt the closed conformation while bound to H3K9me3 peptides is uncertain, as there is only one co-crystal structure including the H3K9me3 substrate peptide. However, the preference of the GDI antibody

[d]Structure unpublished.

family for a sequence immediately following two alpha helices that are moving suggests the states may have functional relevance.

KDM4D, along with KDM4A, 4B, and 4C comprise the KDM4 subfamily. In all of the available crystal structures for KDM4A, 4B, and 4C they adopt a conformation analogous to KDM4D's less common, outward conformation:

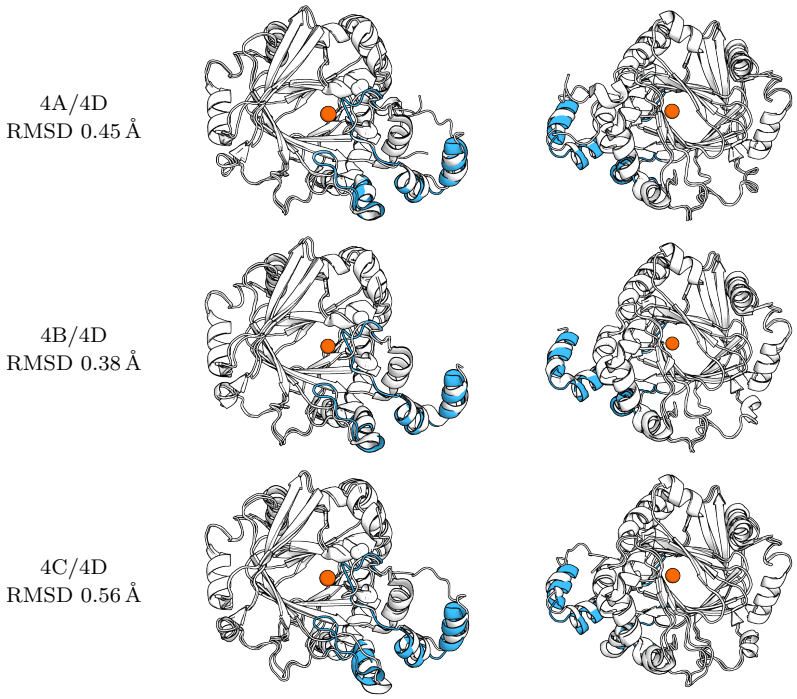


Figure 2.19: Crystal structures for KDM4A (PDB: 2GP3), KDM4B (4LXL), and KDM4C (4XDO) overlayed with KDM4D (4HON) and their rotation through 180°. *The C-terminal of KDM4D is coloured blue, and iron (at the active site) shown in orange for reference. RMSD was calculated using PyMOL.*[33]

These two conformations of KDM4D's C-terminal are unrepeated in the structures currently available for JmjC KDMs. Whether they are the results of crystal packing forces, or (more likely) KDM4D adopting at least two different conformation in solution, that they can be seen at all signals the regions propensity for dynamics. The GDI family of antibodies could be inducing a similar change in the region, or simply capture a particular state, and in doing so allosterically inhibit KDM4D enzymatic activity. Since the C-terminal domain is essential for most recruitment of KDM4D,

any allostery induced from interaction in the region may even be used to modulate KDM4D activity *in vivo*.

Further Crystal Trials and Remaining Antibodies

Unfortunately crystallisation trials for J-1, the inhibitor outside the GDI family, did not fare much better than those with J-11, J-12, or J-16. Crystallisation did not occur at 4 °C nor at 18 °C, and not for complexes of J-1 and KDM4D$_{1–358}$, nor J-1 and KDM4D$_{1–342}$. A ternary complex of J-1, J-11, and KDM4D$_{1–358}$ was also put into crystal trials as an additional system with potentially different crystal contacts, but also did not crystallise.

As crystallographic chaperones, the non-inhibiting antibodies isolated here had potential use. A different antibody added to the crystallisation system is a chance for new crystal contacts, and potentially crystals. However, many of these antibodies struggled to form complexes in size exclusion chromatography assays.

Antibody	J-1	J-3	J-4	J-5	J-11/J-12/J-16	J-19	J-23
Solution Binding	•	•			•		

This effect seems most likely due to antibodies that prefer an epitope only available in denatured protein. For each of these antibodies, positive ELISA binding and specific patterns of cross-reactivity towards a panel of KDMs essentially rules out non-specific binding. But if no binding was observed in a solution-based binding assay even at concentration of 10 µM, it is unlikely such a weak inhibitor could be detected by ELISA in the first place. This seeming contradiction implies a difference in antigen presentation: when KDM4D was immobilised on plastic for the ELISA, and was necessarily partially denatured, an epitope not normally available in solution was exposed. Such a binding preference prevents the application of these antibodies to crystallography with intact KDM4D.

Summary of Main Results

23 Monoclonal antibodies for the KDM4D were recovered from the lymph nodes of two NZW rabbits. A family of inhibiting antibodies, the "GDI" family, bound $KDM4D_{1–358}$ in an SPR assay with sub-nanomolar affinity, but didn't bind to $KDM4D_{1–342}$ in a SEC assay. The 16 residues that comprise this antibody family's putative epitope, 342–358, are located distal to the active site, in an area that can adopt at least two conformations. These results suggest that the GDI family of antibodies could inhibit KDM4D by an allosteric mechanism, and that the C-terminal region of KDM4D may influence activity of the N-terminal catalytic region.

2.3 Conclusions

To bind, the GDI family of inhibiting antibodies relies exclusively on residues distal to KDM4D's active site, contained in a region known to adopt more than one conformation, and preceding a domain required for recruitment by diverse binding partners. Even without a co-crystal structure, these antibodies make it difficult not to confront that KDM4D adopts at least two conformations and that they likely influence its function. They also suggest the selective inhibition so hard to achieve at KDM4D's active site, may be better achieved outside it.

Allosteric inhibition by the GDI family is best supported by their putative epitope being as far displaced linearly from KDM4D's active site as the immunised construct allows. It is also, perhaps more subtly, supported by IC_{50} values below the tight binding limit. These values imply not all molecules of KDM4D are simultaneously active. If this inactivity was not due to permanent damage, which seems unlikely given KDM4D's robustness at least to repeated freeze-thaw cycles, it could a symptom of KDM4D adopting both active and inactive conformations under assay conditions.

Although without a co-crystal structure it is difficult to say with certainty that the GDI family are allosteric inhibitors, probability weighs heav-

ily in its favour. The GDI family of antibodies occluding KDM4D's binding site despite binding elsewhere, is possible due to their relative sizes. But the region these antibodies need to bind is clearly flexible, and therefore likely to allow the significantly smaller substrate and cofactors access to the active site. Wild-type KDM4D also continues for an extra 163 residues past the end of the construct immunised here, but remains catalytically active. A more detailed kinetic analysis may help resolve the mechanism behind the inhibition observed. However KDM4D having four substrates: H3K9me3, 2OG, Fe(II), and O_2, complicates kinetic analysis. Though they may be approximated by supplying each co-factor in excess, it is difficult to prove more than non-competitive inhibition.

A crystal structure remains tantamount to resolving questions of KDM4D's allostery, as well as inspiring small-molecule inhibitors. GDI family antibodies in the scFv format offer, perhaps, the most potential for structural studies. Alone it came close to crystallising, producing a gel and, after optimisation, small needles. Any flexibility that previously prevented crystallisation may be stabilised by the addition of a peptide representing the 16 residues that contain the GDI family's probably linear epitope. Such a co-crystal structure, combined with molecular dynamics should give a reasonable insight into conformations of KDM4D stabilised by antibody binding, and how they may affect function. Though too time consuming to have been used here, an scFv-peptide complex is within the molecular weight accessible to NMR structural studies.

Whether the GDI family of antibodies binds to wild-type KDM4D is still unknown. But should the GDI family bind to wild-type KDM4D, since it is stable as an scFv, expression as an intrabody could yield a selective KDM4D inhibitor for cellular studies. Regardless, the GDI family has lent experimental support to an underappreciated region of KDM that likely has significance for function. The C-terminal of many KDMs serves as a recruitment domain, and would be surprising if evolution had not exploited these interactions to modulate the sensitive function of KDMs.

References

[1] S. Krishnan and R. C. Trievel. "Structural and functional analysis of JMJD2D reveals molecular basis for site-specific demethylation among JMJD2 demethylases". In: *Structure* 21.1 (2013), pp. 98–108.

[2] S. Matoba et al. "Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation". In: *Cell* 159.4 (2014), pp. 884–895.

[3] L. C. Young, D. W. McDonald, and M. J. Hendzel. "Kdm4b histone demethylase is a DNA damage response protein and confers a survival advantage following γ-irradiation". In: *Journal of Biological Chemistry* 288.29 (2013), pp. 21376–21388.

[4] H. Khoury-Haddad et al. "PARP1-dependent recruitment of KDM4D histone demethylase to DNA damage sites promotes double-strand break repair". In: *Proceedings of the National Academy of Sciences* 111.7 (2014), E728–E737.

[5] T.-D. Kim et al. "Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D". In: *PloS one* 7.4 (2012), e34618.

[6] S. Shin and R. Janknecht. "Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D". In: *Biochemical and biophysical research communications* 359.3 (2007), pp. 742–746.

[7] N. Le May et al. "Poly (ADP-ribose) glycohydrolase regulates retinoic acid receptor-mediated gene expression". In: *Molecular cell* 48.5 (2012), pp. 785–798.

[8] H. Khoury-Haddad et al. "The emerging role of lysine demethylases in DNA damage response: dissecting the recruitment mode of KDM4D/JMJD2D to DNA damage sites". In: *Cell Cycle* 14.7 (2015), pp. 950–958.

[9] M. Zoabi et al. "RNA-dependent chromatin localization of KDM4D lysine demethylase promotes H3K9me3 demethylation". In: *Nucleic acids research* (2014), gku1021.

[10] H. Bur et al. "Strong KDM4B and KDM4D Expression Associates with Radioresistance and Aggressive Phenotype in Classical Hodgkin Lymphoma". In: *Anticancer Research* 36.9 (2016), pp. 4677–4683.

[11] Y. Soini, V.-M. Kosma, and R. Pirinen. "KDM4A, KDM4B and KDM4C in non-small cell lung cancer". In: *International journal of clinical and experimental pathology* 8.10 (2015), p. 12922.

[12] E. R. Cleys et al. "Androgen receptor and histone lysine demethylases in ovine placenta". In: *PloS one* 10.2 (2015), e0117472.

[13] Y. Zhu, D. van Essen, and S. Saccani. "Cell-type-specific control of enhancer activity by H3K9 trimethylation". In: *Molecular cell* 46.4 (2012), pp. 408–423.

[14] J. Jin et al. "Epigenetic regulation of the expression of Il12 and Il23 and autoimmune inflammation by the deubiquitinase Trabid". In: *Nature immunology* (2016).

[15] N. Iwamori et al. "The testis-enriched histone demethylase, KDM4D, regulates methylation of histone H3 lysine 9 during spermatogenesis in the mouse but is dispensable for fertility". In: *Biology of reproduction* 84.6 (2011), pp. 1225–1234.

[16] C. C. Thinnes et al. "Targeting histone lysine demethylases–progress, challenges, and the future". In: *Biochimica et Biophysica Acta (BBA)–Gene Regulatory Mechanisms* 1839.12 (2014), pp. 1416–1432.

[17] L.-J. Liu et al. "An iridium (III) complex inhibits JMJD2 activities and acts as a potential epigenetic modulator". In: *Journal of medicinal chemistry* 58.16 (2015), pp. 6697–6703.

[18] F. Sievers et al. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega". In: *Molecular systems biology* 7.1 (2011), p. 539.

[19] M. Goujon et al. "A new bioinformatics analysis tools framework at EMBL–EBI". In: *Nucleic acids research* 38.suppl_2 (2010), W695–W699.

[20] *ClustalW FAQ*. 2017. URL: `http://web.archive.org/web/20171205143549/https://www.ebi.ac.uk/Tools/msa/clustalw2/help/faq.html` (visited on 12/05/2017).

[21] G. Köhler and C. Milstein. "Continuous cultures of fused cells secreting antibody of predefined specificity". In: *Nature* 256.5517 (1975), pp. 495–497.

[22] S. Tickle et al. "High-throughput screening for high affinity antibodies". In: *JALA: Journal of the Association for Laboratory Automation* 14.5 (2009), pp. 303–307.

[23] A. M. Clargo et al. "The rapid generation of recombinant functional monoclonal antibodies from individual, antigen-specific bone marrow-derived plasma cells isolated using a novel fluorescence-based method". In: *MAbs* 6.1 (2014), pp. 143–159.

[24] W. Bowen. "High-throughput antibody screening using high-sensitivity microplate cytometry". In: *Nature Methods— Application Notes* (2010).

[25] S. E. Hutchinson et al. "Enabling lead discovery for histone lysine demethylases by high-throughput RapidFire mass spectrometry". In: *Journal of biomolecular screening* 17.1 (2012), pp. 39–48.

[26] M. Leveridge et al. "Demonstrating enhanced throughput of RapidFire mass spectrometry through multiplexing using the JmjD2d demethylase as a model system". In: *Journal of biomolecular screening* (2013), p. 1087057113496276.

[27] M. Gale and Q. Yan. "High-throughput screening to identify inhibitors of lysine demethylases". In: *Epigenomics* 7.1 (2015), pp. 57–65.

[28] A. Kawamura. personal communication. 2015.

[29] M. Mantri et al. "Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases". In: *The FEBS journal* 279.9 (2012), pp. 1563–1575.

[30] S. Koide and S. S. Sidhu. "The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins". In: *ACS chemical biology* 4.5 (2009), pp. 325–334.

[31] E. Jones, T. Oliphant, and P. Peterson. "SciPy: open source scientific tools for Python". In: (2014).

[32] B. K. Shoichet. "Interpreting steep dose-response curves in early inhibitor discovery". In: *Journal of medicinal chemistry* 49.25 (2006), pp. 7274–7277.

[33] W. DeLano. *PyMOL molecular viewer*. 2002.

3 PHF8

3.1 Introduction

A prerequisite to realising the goal of translating an inhibiting antibody into an inhibiting small molecule is structural inspiration. Molecular weight makes X-ray crystallography the most appropriate technique, so an additional KDM, PHF8, was investigated in part because crystallography for KDM4D could be unsuccessful. PHF8 was selected because it relies on two independent domains: a PHD and a JmjC domain. Both domains are potential targets for inhibition as catalytic turnover relies on recognition of the substrate by both, therefore PHF8 offers, in one protein, two different systems for structural studies.

PHD Finger Protein 8

The effect, in PHF8 and its relative KDM7A, of combining PHD and JmjC, epigenetic reader and eraser, is distinct among KDMs. Alone, PHF8's JmjC domain is inactive, but preceded by its PHD, PHF8 becomes a H3K4me3-K9me2/1 demethylase.[1,2,3,4] The histone 3 tail, trapped at H3K4me3 by the PHD, positions K9me2/1 towards the JmjC, which is afforded, at least in part by avidity, the conditions to convert H3K9me2 to H3K9me1, and H3K9me1 to H3K9me0.[5] The PHD-JmjC domain relationship also occurs in KDM7A, where a greater distance between PHD and JmjC moves specificity from H3K9me2/1 to H3K27me2/1.[5] However, KDM7A retains some activity without its PHD.[5,6] PHF8's activity, by depending on both H3K9me2 and H3K4me3, makes explicit that these epigenetic marks have distinct meaning, that there is reason for them to not occur together and, by extension, reason for them to be preserved when separate. In PHF8's structure evolution has made from two domains one function, demonstrating that the language of

epigenetics is combinatorial, that is, epigenetic modifications have meaning in combination perhaps above what they mean alone. By extension the regulation of gene expression may be as nuanced as epigentic codes are, in combination, many.

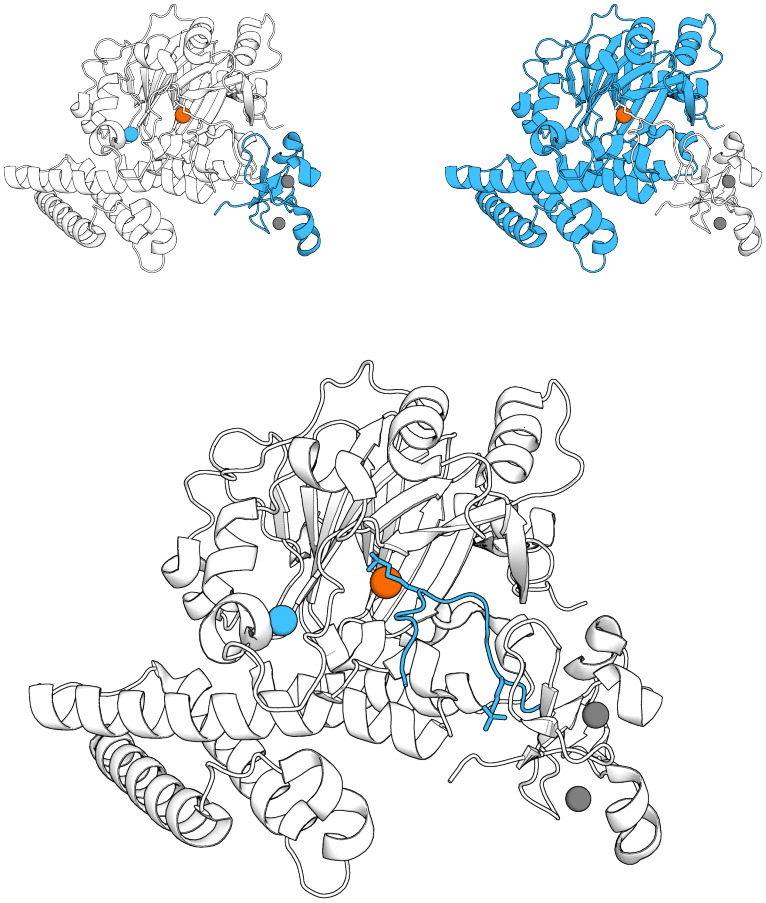


Figure 3.1: A crystal structure of the PHF8 catalytic region (rendered from PDB: 4HON)[5] showing in blue the PHD (top left), JmjC domain (top right), and substrate peptide (bottom). Zinc ions are shown are grey spheres, iron as an orange sphere, and nickel as a blue sphere. K9me2, represented as sticks, is positioned towards the active site iron in the JmjC domain, and K4me3, also represented as sticks is bound by the PHD.

PHF8's initial identification was as a gene whose deletetion, mutation, and truncation can accompany X-linked mental retardation (XLMR) with a cleft lip/palate phenotype.[7,8,9,10,11,12,13] Underlying the phenotype is demethy-

lation. In humans a single point mutation, F279S, causes XLMR with cleft lip/palate,[14] and in recombinant PHF8 the mutation is enough to ablate its catalytic activity.[15,2,3,1,4,14] This craniofacial and neuronal development phenotype is perhaps PHF8's the well known, present in humans, as well as zebrafish, whose jaws cannot develop without PHF8,[15] and *C. elegans*, whose locomotion is impaired by mutation in PHF8.[14] PHF8 is particularly well expressed in neurons, where its knockdown can impair neuronal differentiation, and its recruitment by the acetyltransferase TIP60 for activation of Arc, follows increased synaptic activity.[3,16,17]

To the proteins that recruit it, PHF8 can offer gene activation. The histone modifications removed, H3K9me2 and H3K9me1, are repressive. In contrast, the H3K4me3 modification PHF8 requires, but does not remove, activates gene expression. In cell-cycle regulation, PHF8 apparently also removes the repressive H4K20me2/1 modification by a mechanism only reproducible *in vitro* with histone octamers.[18,19,15] Both activities assist progression through the cell cycle, activating gene expression before the G_1/S transition, the G_2/M transition, and kinetochore assembly.[18,20,19] What regulates PHF8 is not understood in detail, though inactivating phosphorylation by cyclin E/CDK2 has been repored.[21] Beyond the cell cycle, PHF8's action extends to the activation of ribosomal RNA (rRNA) gene promoters,[22,4] activation of the gene expression connecting inflammation and T-cell activation,[23] and the differentiation of mouse embryonic stem cells into cardiomyocytes.[24,25]

It is almost inevitable that a protein with its function as close as PHF8's is to gene expression, and as widely employed, will be exploited by cancer cells. PHF8's involvement has been noted in prostate cancer,[26,27] glioma,[28] breast cancer,[29,30] non-small cell lung cancer,[28] esophageal squamous cell carcinoma,[31] and leukaemia.[32,33] Beneath each pathology is an apparently diverse application of PHF8's demethylation. Many of the interaction partners, targets, and regulators are familiar to cancer biology: NOTCH1 recruits PHF8 in leukaemia,[33] PHF8 can activate LMO3 expression in glioma

cells,[34] MYC can upregulate PHF8 in breast cancer cells,[29] and there is evidence for PHF8 being an androgen receptor co-activator in prostate cancer cells.[27] Although knockdown of PHF8 appears to make acute promyelocytic leukaemia cells resistant to all-trans retinoic acid acid treatment, in prostate, esophageal, breast, and lung cancer cells PHF8 knockdown inhibits proliferation.[26,31,29,28] Thus the majority of clinical indications call for inhibition of PHF8.

So far small-molecule screening has not explored regions beyond the PHF8 active site. The most successful PHF8/KDM7A inhibitor, daminozide, is roughly selective for the KDM2 and KDM7 subfamily, chelating the active site iron analogously to 2OG.[35] Moving from small molecules to cyclic peptides has brought improved selectivity, but at the cost of poorer affinity.[36] PHF8 presents in its PHD and JmjC two targets for competitive inhibitors, and in their spacial arrangement, on which catalysis relies, many more opportunities for allosteric inhibition. Were it not an intracellular protein, PHF8 could be considered an almost ideal therapeutic antibody target. Instead, it must be an attractive target for antibody-inspired small molecule drug discovery.

3.2 Results

Antibody Discovery on PHF8

Protein Expression and Immunisation

The construct of PHF8 that was used for immunisation begins at the PHD and extends 447 residues to the end of the JmjC domain – residues 1–447 of the 1024 residue wild-type protein.[a] By containing only the two domains needed for enzymatic activity, antibodies should be forced towards areas relevant to function, and hopefully modulate it. $PHF8_{1–447}$ was expressed in *E. coli* as an N-terminal fusion with thioredoxin cleavable through a TEV

[a] Uniprot accession code: `Q9UPP1-2`.

site. Though its expression was lower than that of KDM4D (chapter 2), a 12 L culture in LB was enough to yield the 20 mg need to immunise two rabbits. In addition, a C-terminal avi-tag construct was sub-cloned and expressed for screening that required biotinylation. A construct of PHF8 without a PHD (encoding residues 79–447) was also expressed to screen the selectivity of antibodies for the PHD or JmjC domain.

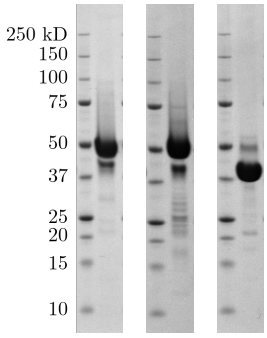


Figure 3.2: (From left) an SDS-PAGE gel for $PHF8_{1–447}$, $PHF8_{1-447}$ biotinylated using an avi-tag at the C-terminus, and $PHF8_{79–447}$.

The Antibody Response

Antibodies for PHF8 were isolated directly from single B-cells in the same manner as for KDM4D. Approximately 2×10^8 B-cells isolated from the lymph nodes of two NZW rabbits immunised with PHF8 were divided among the wells of 300 96-well plates and cultured for one week. Around 7 % of these wells contained antibodies that bound to $PHF8_{1–447}$ in the FMAT assay, of which the best 1050 were preserved for further screening and potentially antibody recovery. Since the antibody concentration in B-cell supernatants is too low to screen for PHF8 inhibitors, each well was instead categorised by its cross-reactivity to a panel of five KDMs using ELISA: KDM7A, KDM2A, KDM4A, KDM4B, and KDM4D; as well as the immunisation construct, $PHF8_{1–447}$, containing the PHD and JmjC, and $PHF8_{79–447}$, the JmjC domain alone.

PHF8 Selectivity		KDM Selectivity						
PHD-JmjC	JmjC	7A	2A	4A	4B	4D	Bin Size	Percent
•	•						489	53.7
•	•	•					154	18.0
•							103	12.1
•		•					47	5.5
•	•		•				19	2.2
•	•	•	•				11	1.3
•	•			•	•		5	0.6
•			•				5	0.6
•	•					•	4	0.5
•	•			•			3	0.4
•	•				•		3	0.4
•	•	•		•	•		3	0.4
•		•	•				3	0.4
•	•			•	•	•	2	0.2
•					•		1	0.1
•				•	•		1	0.1
•	•	•		•			1	0.1

Table 3.1: Cross-reactivity, determined by ELISA, of the 854 *antibody containing supernants selected for further screening from the initial B-cell cultures divides them into* 17 *categories.*

Cross-reactivity divides the antibody-containing supernatants into 17 unique categories. 90 % (793/854) of supernatants were specific for only $PHF8_{1–447}$ or PHF8 and KDM7A. An approximately uniform distribution of these epitopes across PHF8 is suggested, although not guaranteed, by the 81 % (64/793) of supernatants specific for the JmjC domain being approximately equal to the fraction of $PHF8_{1–447}$ residues occupied by the JmjC domain, 82 % (368/447). B-cells were isolated from wells of diverse cross-reactivity to give the best chances of recovering diverse antibodies.

Antibodies Isolated

Among the heavy and light chains of the 47 antibodies isolated here were 13 families of two or more antibodies, and a further 14 with unique heavy-chain CDR3 sequences.

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-12	SGIDLSSYEMG	IISGSGTTYYAS	-D--VGSSGAGNYFNI--	QASEDI-----ER--YLG	DASKLA	QDGGHTTTG---ID
P-14	SGIDLSSYEMG	IISGSGTTYYAS	-D--VGSSGAGNYFNI--	QASEDI-----ER--YLG	DASKLA	QDGGHTTTG---ID
P-29	SGIDLSSYEMG	IISGSGTTYYAS	-D--VGSSGAGNYFNI--	QSSESV-----YNNNRLS	KASTLE	AGDA-SGYI---LA
P-17	SGIDLSWYEMG	IISGSGTTYYAS	-D--VGSSGAGNYFNV--	RASESI-----ES--YLA	DASTLA	QQHY-SITNV-DNA
P-15	SGIDLSWYEMG	IISGSGTTYYAS	-D--VGSSGAGNYFNV--	QASEDI-----ER--YLG	DASKLA	QDGGHTTIA---ID
P-9	SGFSLSNNAMT	TIYSGGNTYYAN	-------FGVTVDKGI--	QSSQSV-----YNKNNLA	LASTLA	AGRY-SDNI---YD
P-38	SGFSLSNNAMT	TIYTGGNTYYAS	-------FGVTVDKGI--	QSSQSV-----YNKNNLA	LASTLA	AGRY-SDNI---YD
P-4	SGFSLSNNAMT	TIYSGGNTYYAN	-------FGVTVDKGI--	QSSQSV-----YNKNNLA	LASTLA	AGRY-SDNI---YD
P-31	SGFSLSYNAMT	TIYTGGNTYYAT	-------FGVTVDKGI--	QSSQNV-----YNNNNLA	LASTLA	AGRY-SDNI---YD
P-50	SGIDLSSYHMT	IISTGGTTYYAT	--------DDGSGSSI--	QSSQSV-----YSN-WLA	AASTLD	EGGYSSGDG---RA
P-23	SGIDLSSYHMT	IISTGGTTYYAT	--------DDGSGSSI--	QSSQSV-----YSN-WLA	AASTLD	EGGYSSGDG---RA
P-19	SGIDLSSYHMT	IISTGGTTYYAT	--------DDGSGSSI--	QSSQSV-----YSN-WLA	AASTLD	EGGYSSGDG---RA
P-26	SGIDLSTYAIP	IIGSSGSTYYAT	----------GPTTDL--	QASEDI-----ER--YLG	DASKLA	QDGGHTTIA---ID
P-28	SGIDLSTYAIP	IIGSSGSTYYAT	----------GPTTDL--	QASEDI-----ER--YLG	DASKLA	QDGGHTTTG---ID
P-18	SGIDLSTYAIP	IIGSSGSTYYAT	----------GPTTDL--	QSSESV-----YNNNRLS	KASTLE	AGGD-SGNI---YA
P-35	SGFSLSNYAMS	TIGTGGNTYYAN	----------GNPGNL--	QASQSL-----GRN-YLS	KASTLA	QSNYYTTSTNYRNT
P-49	SGIDLSSYAMG	TIGTGGTTYYAS	----------GNPGNL--	QASQSL-----DSK-YLS	RASTLE	QSNYYTISTNYKNT
P-3	SGFDISSYHMG	WISSSGNTYYAS	----------GFPSTI--	QASQSI-----GD---LA	RASTLA	QNNN-GGSG---YA
P-20	SGFSLSSYHMG	WISSSGNTYYAS	----------GFPSTI--	QASQSI-----GS---LA	DASDLA	QNNN-GGSA---YA
P-42	SGFSLSTYAMS	IIGSSGSPYYAS	AAYPAYSTTTGYYYGMDP	QASQSI-----GT--ALA	DASKLA	LGGY-ASNSG-TFA
P-43	SGFSLSTYAMS	IIGSSGSPYYAS	AAYPAYSTTTGYYYGMDP	QSSQSV-----YNNNLLS	AVSYLA	LGGY-DDDG--DCA
P-30	SGFSLSSYAMS	IISSRGSTYYAS	-SSSSGYYILEYYFNL--	QASESI-----YS--NLA	GASTLD	AGWY-SDNI---LT
P-34	SGFSLSSYAMS	IISSRGSTYYAS	-SSSSGYYILEYYFNL--	QASESV-----YNNNRLA	SASSLA	AGYK-SSNTD-GNA
P-8	SGIDLSSYAMG	IIEISGATGYAN	-V--NSPGYTTNTLKI--	QASQSV-----YESNSLA	DASTLP	LGEFSCSSAD-CLA
P-21	SGIDLSSYAMG	IIEISGATGYAN	-V--NSPGYTTNTLKI--	QASESI-----SS--YLA	RASTLA	QSYYGSSSTG-SNV
P-44	SGFSLSSNAMS	IISSSGSTYYAN	-----LIVVDHYYFNL--	QASQSI-----SS--YLN	RASTLA	QQGY-TSTNI-DNA
P-1	SGFSLSRYAMS	IISSRGSTYYAS	-----LVVMDHYYFNL--	QASQSI-----SS--YLS	GASNLA	QQGY-TSSNI-DNA
P-7	SGFSLSTYAMN	IIGSSGSTYYAS	-----WDDDYGDAFDP--	QATQII-----NS--WLS	RASTLA	QSNYYSTITD----
P-13	SGFSLSTYAMN	IIGSSGSTYYAS	-----WDDDYGDAFDP--	QATQII-----NS--WLS	RASTLA	QSNYYSTITD----
P-32	SGFSLSDYAVG	TISDSGSTFYAN	----------GWSL--	QASQSI-----STA--LA	RASTLA	HCTA-DGSAY-GNA
P-52	SGIDLSSYAMI	TTTSSGSTYYAS	----------AFSL--	QASESI-----STA--LA	KASTLA	QCSA-DGSSY-GNP
P-36	SGIDLSSYAMS	TISYRGNPYYAS	-R--GYPTDNTAMDDL--	QSSESV-----HRNNYLS	STSTLA	QGGYSSGDA---TA
P-24	SGIDLSSYVMG	IISFNGNTYYAS	-D--RAYASSSDYYNI--	QASQSI-----SS--YLA	SASTLE	QSTYDSSSN---NV
P-45	SGFSLSNYDMS	IIYGSDNIYYAN	-G--DAGNGGYTAFNI--	QASESI-------SNWLA	EASKLA	QQTA-SNSNV-HNV
P-22	SGIDLSSYAMG	IITGAANTYYAN	-G---SDGGSAPHYPT--	QSSESV-----YKNNYLS	SASTLA	GGVY-SGNI---LA
P-48	SGFSLSSYYMS	IIVASGSTYYAS	-T---GYAGDGYPMDI--	QASQSL-----SN--LLA	RASTLA	QGAD-YMTTT-ENT
P-11	SGFTISRYDMT	IIWSGDNTDYAN	-E----IYSYNSDGGL--	QASQSV-----YNNKNLA	SASILA	LGEFSCSSAD-CFA
P-40	SGFSLSSYRVS	IIYASGYTYYAS	-----SYSISSLYYDI--	QASQSI-----GS--DFS	RASTLA	QQAY-SDFNV-ENP
P-10	SGIDLSSYNMG	IIEISGATGYAN	--------VWGYYYGIDP	KTSQSV-----YNNNFLS	GASTLT	GGAY-SDNI---YG
P-2	SGIDLSRYAVI	IISRSATTYYAS	-------DHPGYSFNI--	QASQSV-----YNDNTLA	DSSNLA	QGYY-SGAI---YA
P-6	SGIDLSTYAMT	IIGTGGQTYYAN	-------GNAGGGADV--	QSSQSVYSKNRYSKNRLA	AASTLE	AGLY-SDNT---NT
P-39	SGFDISRYAMS	IISDSPNTAYAS	-------ITARTYPNL--	QASQSL-----YNNKNLA	AASKLA	LGEFNCDSAD-CLT
P-41	SGIDLSRYAMT	VIGSSGVTYYAT	--------GVVSAADI--	QSSQSV-----YENNRLA	RASKLE	AGAY-VDNL---YP
P-5	SGFSLSSYAMI	FIGDTGTTYYAN	----------GNPANI--	QASQSM-------RNYLA	DPSTLE	QTYYYTKTIN-GGA
P-16	SGFSLSSNAIN	YTSST-DTYYAT	------------ELNI--	QASQSI-----STA--LA	KASTLA	QCTGVGGSAD--MT
P-53	SGFSLTSAHMC	FIWSDGSTDYAN	------------NFNL--	QASQSV-----YDNDDLA	EASKLE	AGGFSDSNTD-KNA
P-25	SGFSLSSNAIS	VIGVSGTTYHAS	-------------ANI--	QSSESV-----YSNNRLA	SASTLD	QGGY-SSNI---LA

Figure 3.3: CDR sequences of the 47 isolated PHF8 antibodies arranged by heavy chain CDR3 sequence.

From these antibodies 28 were selected based on their sequence diversity and cross-reactivity for recombinant expression as Fabs. Their cross-reactivity was reconfirmed with KDMs: KDM4A, KDM4B, KDM2A, KDM7A; PHF8 as immunised, and the JmjC domain alone; and KDM4C replacing KDM4D.

			PHF8 Selectivity		KDM Selectivity				
	CDR-L3	CDR-H3	PHD-JmjC	JmjC	7A	2A	4A	4B	4C
P-35	QSNYYTTSTNYRNT	----------GNPGNL--	•	•					
P-49	QSNYYTISTNYKNT	----------GNPGNL--	•	•					
P-5	QTYYYTKTIN-GGA	----------GNPANI--	•	•					
P-12	QDGGHTTTG---ID	-D--VGSSGAGNYFNI--	•	•					
P-14	QDGGHTTTG---ID	-D--VGSSGAGNYFNI--	•	•					
P-44	QQGY-TSTNI-DNA	-----LIVVDHYYFNL--	•	•					
P-1	QQGY-TSSNI-DNA	-----LVVMDHYYFNL--	•	•					
P-45	QQTA-SNSNV-HNV	-G--DAGNGGYTAFNI--	•	•					
P-24	QSTYDSSSN---NV	-D--RAYASSSDYYNI--	•	•					
P-48	QGAD-YMTTT-ENT	-T---GYAGDGYPMDI--	•	•					
P-11	LGEFSCSSAD-CFA	-E----IYSYNSDGGL--	•	•					
P-53	AGGFSDSNTD-KNA	------------NFNL--	•	•					
P-9	AGRY-SDNI---YD	-------FGVTVDKGI--	•	•	•	•			•
P-4	AGRY-SDNI---YD	-------FGVTVDKGI--	•	•	•	•			•
P-38	AGRY-SDNI---YD	-------FGVTVDKGI--	•	•	•	•			•
P-31	AGRY-SDNI---YD	-------FGVTVDKGI--	•	•	•	•			•
P-6	AGLY-SDNT---NT	-------GNAGGGADV--	•	•	•	•			•
P-23	EGGYSSGDG---RA	--------DDGSGSSI--	•		•				
P-3	QNNN-GGSG---YA	----------GFPSTI--	•		•				
P-20	QNNN-GGSA---YA	----------GFPSTI--	•		•				
P-16	QCTGVGGSAD--MT	------------ELNI--	•	•	•				•
P-32	HCTA-DGSAY-GNA	------------GWSL--	•	•	•				•
P-36	QGGYSSGDA---TA	-R--GYPTDNTAMDDL--	•		•				•
P-13	QSNYYSTITD----	-----WDDDYGDAFDP--	•	•	•				
P-2	QGYY-SGAI---YA	-------DHPGYSFNI--	•		•		•	•	•
P-39	LGEFNCDSAD-CLT	-------ITARTYPNL--	•	•	•		•	•	•
P-7'	QSNYYSTITD----	-----WDDDYGDAFDP--	•	•					
P-43'	LGGY-DDDG--DCA	AAYPAYSTTTGYYYGMDP	•	•					

Table 3.2: Of 28 *antibodies expressed and purified,* 21 *had unrelated sequences. Cross-reactivity placed the purified antibodies into one of eight categories. P-7' and P-43' appeared to be specific for the immunisation construct and not the JmjC domain of PHF8, but were later found to recognise the JmjC domain.*

The majority of antibodies were chosen for their selectivity for only PHF8, as selectivity is important for KDM inhibitors, and difficult to achieve with small molecules. Antibodies with similar CDR3s were purified in some cases to provide redundancy in small-scale antibody expression PHF8. Summarised, the steps of antibody isolation were:

1. **2** rabbits immunised with $KDM4D_{1–358}$

2. **300** plates (**28,800 wells**) incubated for **5** days
3. **7 %** of supernatants reacted to KDM4D (FMAT assay)
4. **1056** wells preserved and stored
5. ELISA confirmed KDM4D specificity of **854** of these wells
6. Cross-reactivity towards **7** KDM constructs revealed **17** categories of specificity
7. **616** B-cells isolated by micromanipulator
8. **47** Monoclonal antibodies cloned as Fab fragments

Activity Assay

In vitro, $PHF8_{1–447}$ will demethylate a H3K4me3K9me$2_{1–21}$ to produce H3K4me3K9me1, which is again demethylated to yield H3K4me3K9me0.[5] This loss of a methyl group from lysine 9 of the histone peptide was once again monitored using a "RapidFire" mass spectrometry assay previously described for KDM4D (chapter 2). Incubated overnight at room temperature, demethylation was only obvious in reactions with PHF8 at concentrations exceeding 1 µM. Such a high enzyme concentration would have required that antibodies be assayed at an even higher concentration, something which significantly suppressed peptide signals in the mass spectra. The inhibition assay was therefore developed at 37 °C, which enabled $PHF8_{1–447}$ concentrations as low as 100 nM. Besides $PHF8_{1–447}$, the substrates 2OG and H3K4me3K9me$3_{1–21}$ were present at 8 µM – below their respective K_m.[37] Fe(II) ($(NH_4)_2Fe(SO_4)_2 \cdot 6\,H_2O$) was used at a concentration of 50 µM with 100 µM of L-ascorbic acid added to impede the oxidation of Fe(II). Turnover appeared more efficient with high NaCl concentrations (here 250 mM was used), and the reaction was buffered with 50 mM MES at pH 7.0.

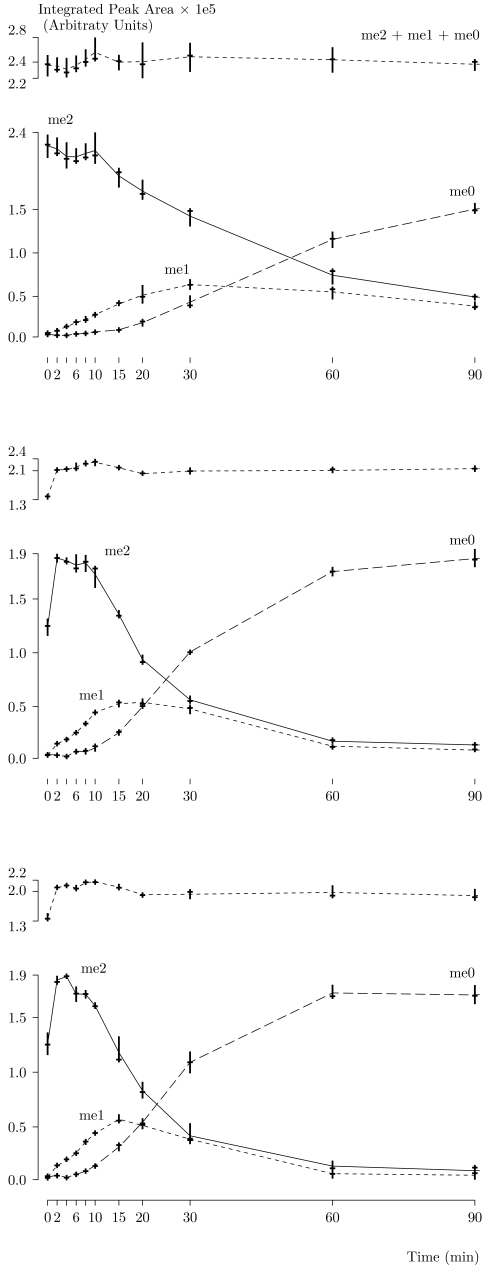


Figure 3.4: Turnover of H3K4me3K9me2 by $PHF8_{1–447}$ measured by "RapidFire" mass-spectrometry at enzyme concentrations of (from top) 100 nM, 300 nM, *and* 500 nM. *Above each timecourse is the sum of product and substrate signals, theoretically a constant. Signal variation at the start of runs is likely due to insufficient blank injects preceding the first sample rather than a genuine biochemical phenomena. Data points show the median of triplicate reactions with vertical bars indicating the maximum and minimum of these three measurements; lines trace the mean value through each time point.*

The plots in figure 3.4 detail the expected depletion of the substrate, H3K4me3K9me2; generation and subsequent depletion of the intermediate product, H3K4me3K9me1; and accumulation of the final product, H3K4me3K9me0. The signal sometimes varied at the start of runs, and was avoided in activity testing by increasing the number of blank runs before the first measurement. The lowest concentration of PHF8 that gave reasonable curves was 100 nM, and based on the timecourse, 20 min was chosen as a reasonable endpoint for activity assays that balanced the signal-to-noise while turnover still remained approximately linear. Antibodies were tested in small panels at a 10-fold molar-excess (1 µM) over $PHF8_{1–447}$. Those antibodies that visually appeared to inhibit were then retested in triplicate.

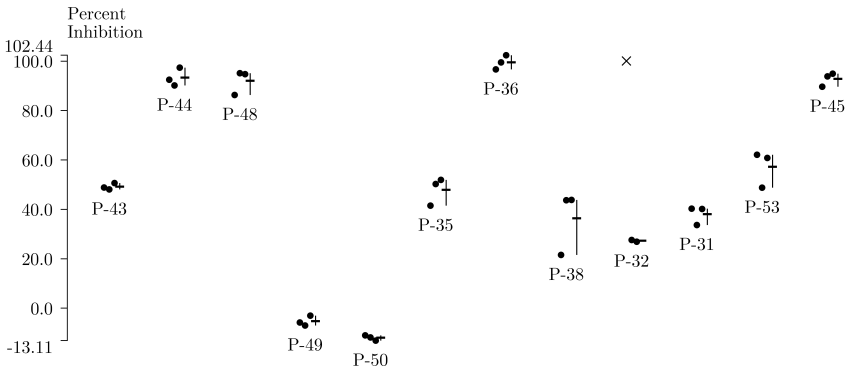


Figure 3.5: Percent inhibition of demethylation by $PHF8_{1–447}$ in the presence of different antibodies. Inhibition for triplicate reactions is shown as a dot, with a vertical line indicating their range, and horizontal line their mean. Apparent outliers are marked with a cross. Percent inhibition was calculated by normalising all points to a sample containing no antibodies, and one containing formic acid to ablate PHF8's activity.

Four antibodies essentially completely inhibited PHF8: P-44, P-48, P-36, and P-45. Others also show some degree of inhibition but weren't prioritised because at an inhibitor concentration of 1 µM, partial inhibition suggests the antibody has low affinity for its antigen, or it binds in a way ineffective for inhibition. That no inhibition was seen for P-49 or P-50 was reassuring for the assay's robustness. Signs of antibody-induced activation of PHF8 were also examined, but no increase in either H3K4me3K9me1 or H3K4me3K9me0 was observed relative to the controls. The inhibiting antibodies were

relatively diverse in their sequences, but conservative in their selectivity:

			PHF8 Selectivity		KDM Selectivity				
	CDR-L3	CDR-H3	PHD-JmjC	JmjC	7A	2A	4A	4B	4C
P-44	QQGY-TSTNI-DNA	----LIVVDHYYFNL	•	•					
P-48	QGAD-YMTTT-ENT	T---GYAGDGYPMDI	•	•					
P-45	QQTA-SNSNV-HNV	G--DAGNGGYTAFNI	•	•					
P-36	QGGYSSGDA---TA	R--GYPTDNTAMDDL	•		•				•

This was a positive result. Inhibitors of diverse sequence are interesting for what they can say about different interactions that can inhibit function, as well as giving options for crystallography. In addition, P-36 not recognising the JmjC domain of PHF8 implies its specificity is for the PHD, inhibition of which has received less attention that the JmjC but is theoretically just as effective.

Solution Binding Behaviour

Selected antibodies were expressed on a larger scale (1 L HEK293 cultures) and purified so that they could be combined with PHF8 and the complexes put into crystallisation trials.

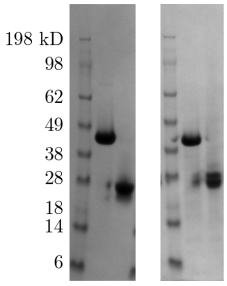


Figure 3.6: Representative SDS-PAGE gels of P36 and P13 purified on a 1 L *scale from HEK293 cell supernatant. Non-reducing conditions are shown in the left lane, and reducing conditions in the right lane.*

However, even at concentrations in excess of 100 µM, little, if any, complex formed. Gel filtration of the antibody P-45 and PHF8 exemplifies the problem:

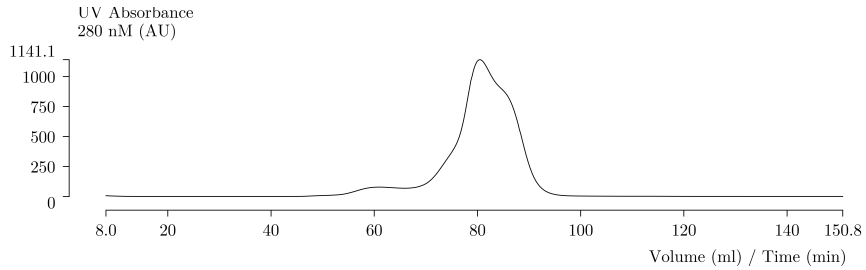


Figure 3.7: Gel filtration of PHF8$_{1–447}$ mixed with P-45 at approximately 100 µM*, and at a molar ratio of 1:1.2.*

Unbound PHF8 eluted from the column just before 80 min, followed by P-45, its molecular weight slightly lower. There is evidence of the complex eluting around 60 min, but not enough for structural studies.

It is difficult for size exclusion chromatography to produce false negatives unless an interaction's affinity is particularly low, or, less likely, the conditions of the assay interfere with complex formation. At concentrations exceeding 100 µM even interactions of modest affinity should form a complex. An interaction as weak as that between P-45 and PHF8 would have to be to produce the UV trace above is unlikely have been detected by even ELISA. But binding was detected by ELISA, and the recognition was specific for only the JmjC domain of PHF8, and therefore not the result of weak, nonspecific interactions.

Whether what prevents complex formation is due to the antibody or antigen can be determined from size exclusion chromatography. If chromatograms are obtained for different ratios of antibody and antigen, but with a constant total protein concentration, plots of the result should take one of the following forms:

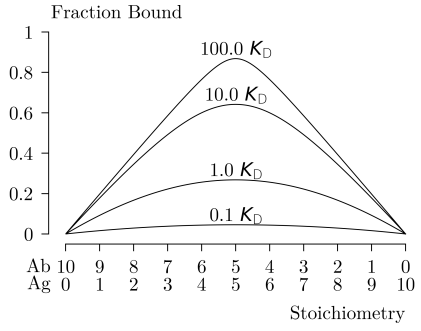


Figure 3.8: Fraction of the total protein bound when antibody (Ab) and antigen (ag) are mixed at differing ratios. The total concentration is shown in terms of K_D above each trace.

At concentrations far above the K_D, when the antibody or antigen is in a large excess, all that is available to bind will bind, and so complex formation will appear to increase linearly with the concentration of the minor species. As their stoichiometry approaches one-to-one, the complex formation reaches its maximum, since the number of molecules available to form a complex is at its maximum. But when equimolar, the concentration of each species, antibody and antigen, are also both the closest they will be (together) to the K_D. There is less pressure on the equilibrium of association and dissociation to favour complex formation, and so the previously linear ascent to maximum complex formation is slowed, and the curves take their characteristic parabolic shape. The closer the total protein concentration is to the K_D, more favourable dissociation is, and thus the less linear ascent to an flatter maximum becomes.

If the antibody or antigen is changed such that it can no longer bind (for instance by denaturation, or aggregation), their effective concentration is reduced. The effect is for their complex formation curves to shift towards higher concentrations of the species whose ability to bind is hindered.

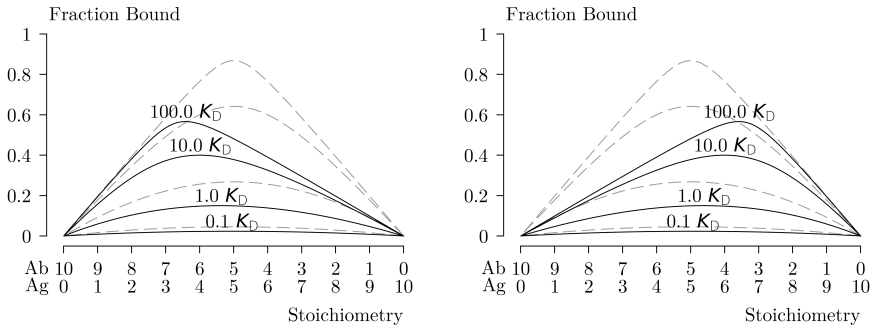


Figure 3.9: Fraction of the total protein bound when antibody (Ab) and antigen (ag) are mixed at differing ratios when (left) 50 % *of antibody and (right)* 50 % *of antigen molecules cannot bind.*

A titration of P-45 against PHF8, keeping the total protein concentration fixed, implied the problem was with $PHF8_{1–447}$, the antigen.

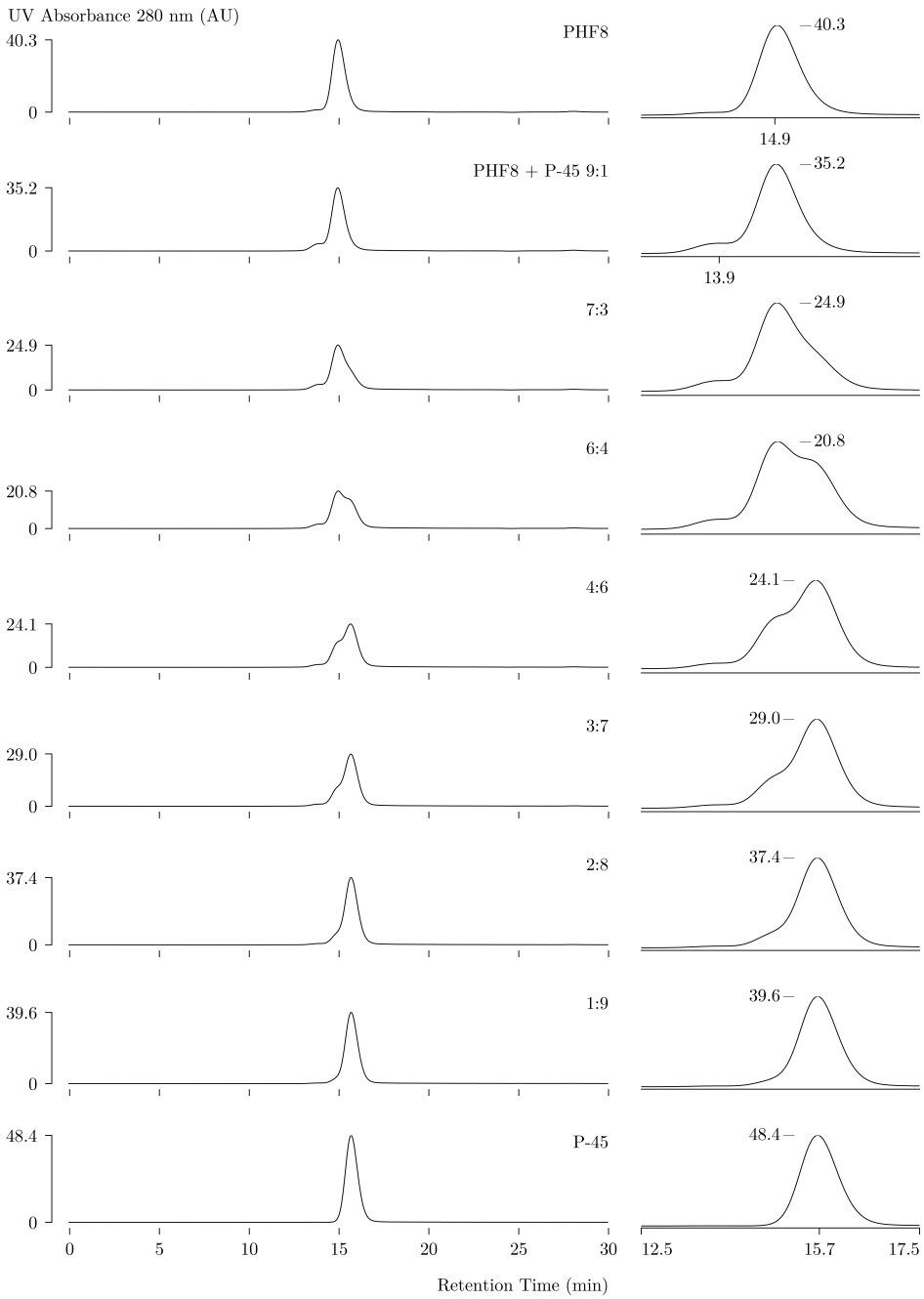


Figure 3.10: A titration of PHF8 against P-45 at a fixed total protein concentration of 5 µM*. Size exclusion chromatograms were obtained using an S200 column, with HBS as the running buffer.*

A complex between P-45 and $PHF8_{1–447}$, with a retention time of 13.9 min, seems to be forming. However it is not symmetric: as the antigen concentration decreases so does the amount of complex.

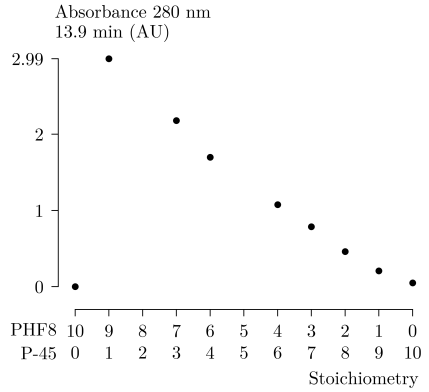


Figure 3.11: Complex formation between $PHF8_{1–447}$ and P-45 measured by UV 280 nM absorbance after 13.9 min. Complex formation appears to depend primarily on the concentration of PHF8.

If the relatively poor complex formation was due only to weak affinity, the plot above, figure 3.11, should be symmetrical. To an equilibrium of association and dissociation between antibody and antigen, and excess of either should have the same effect: pushing the equilibrium towards complex formation. Instead, when the ratio of $PHF8_{1–447}$ to P-45 is reversed, almost no complex forms. The result indicates not all of the antigen molecules can form a complex, since complex formation is only obvious at the extremes of antigen excess.

What is the source of this apparent heterogeneity in the antigen sample? Biologically, there are many possible origins, including: aggregates carried through purification, recombinant antigen that has unfolded, or even a legitimate antigen conformation whose energy is unfavourable enough to ensure only a small fraction of molecules ever adopt it simultaneously.

The PHD of PHF8

In anticipation of structral studies using PHF8-PHD, eight constructs of the PHD were sub-cloned:

```
   .MASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskghdthk...
1  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKK
2  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR
3  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gs
4  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskg
5  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskgh
6  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskghdt
7  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskghdth
8  SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR gsskghdthk
```

Figure 3.12: Constructs of PHF8-PHD below the a part of the wildtype sequence. The predicted unstructured region between the PHD and JmjC domain is indicated in lower case. An N-terminal serine is left by TEV protease.

As fusions with thioredoxin almost all of these constructs expressed in E. coli:[b]

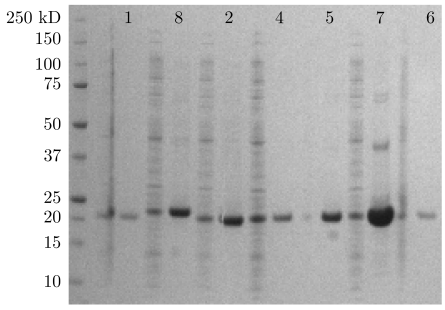


Figure 3.13: A 50 mL *scale expression trial in E. coli for each sub-cloned PHF8-PHD construct show in figure 3.12.*

Though not the most highly expressed, construct 2 was selected for subsequent experiments because it lacks the unstructured region connecting the PHD and JmjC, and therefore potentially minimises the flexibility that can undermine crystallisation.

[b] Construct 5 was ommitted in error, and its expression never tested.

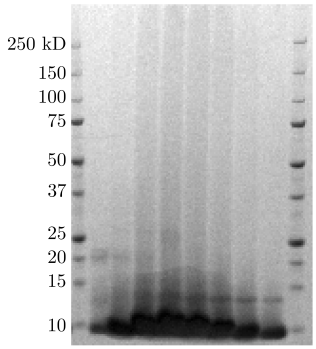


Figure 3.14: SDS-PAGE analysis following gel filtration of a 3 L *expression of PHF8-PHD construct 2.*

Solution Binding to PHF8-PHD

From the panel of antibodies already purified, five were found to be bind to the PHD, rather than JmjC domain of PHF8.

			PHF8 Selectivity			KDM Selectivity				
	CDR-L3	CDR-H3	PHD-JmjC	PHD	JmjC	7A	2A	4A	4B	4C
P-3	QNNN-GGSG---YA	----------GFPSTI--	•	•		•				
P-20	QNNN-GGSA---YA	----------GFPSTI--	•	•		•				
P-36	QGGYSSGDA---TA	-R--GYPTDNTAMDDL--	•	•		•				•
P-2	QGYY-SGAI---YA	-------DHPGYSFNI--	•	•		•		•	•	•
P-23	EGGYSSGDG---RA	--------DDGSGSSI--	•	•		•				

Table 3.3: Selectivity of antibodies binding PHF8-PHD as measured by ELISA, and for PHD/JmjC specificity SPR.

Omitting P-20 due to its sequence similarity with P-23 left four antibodies. When P-36 and PHF8-PHD were subjected to the fixed total protein concentration titrations as P-45 the 7 kDa mass shift separating the antibody and antibody-antigen complex seemed too subtle for detection, but at a total protein concentration of 20 µM, a putative complex, was observed as a shoulder to the P-36 peak.

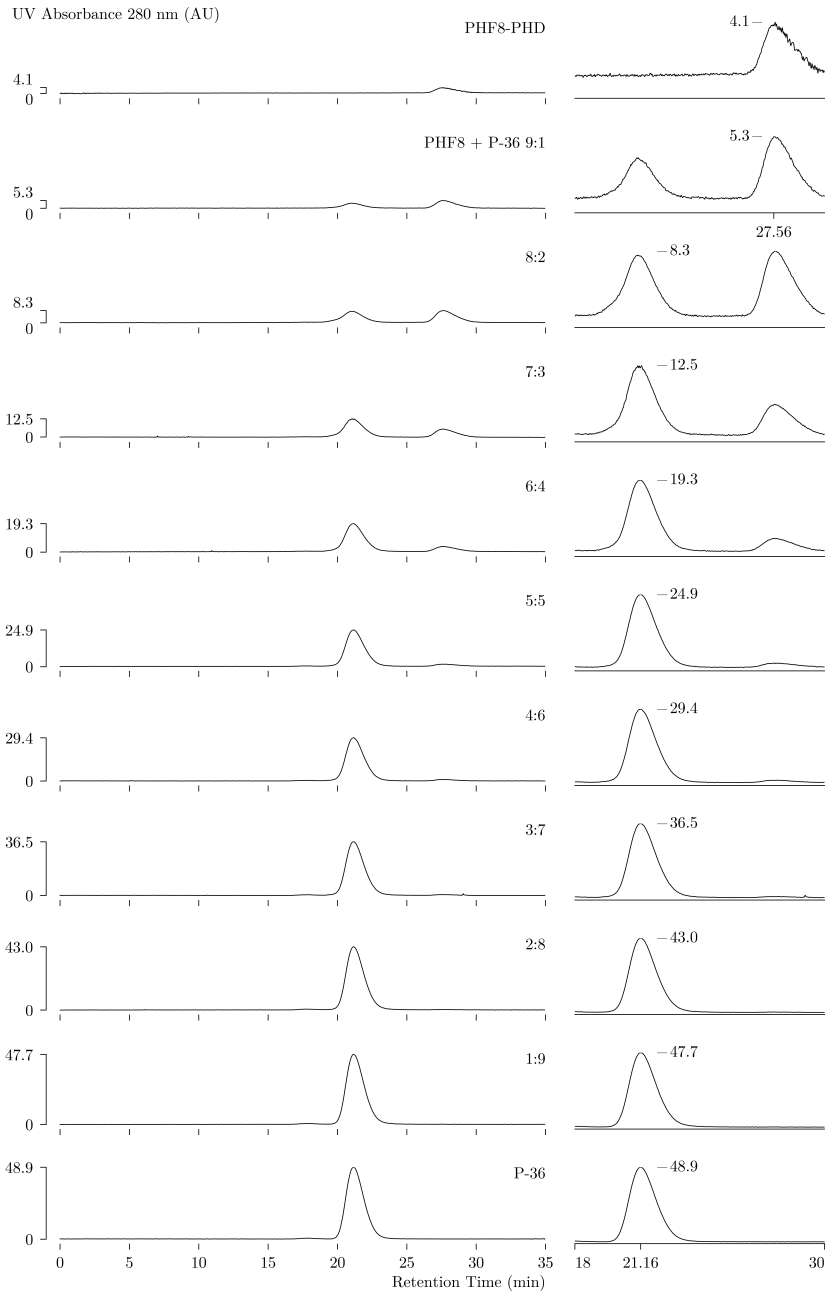


Figure 3.15: A titration of PHF8-PHD against P-36 at a fixed total protein concentration of 5 µM*. The column used for SEC was an S200, and the running buffer was PBS.*

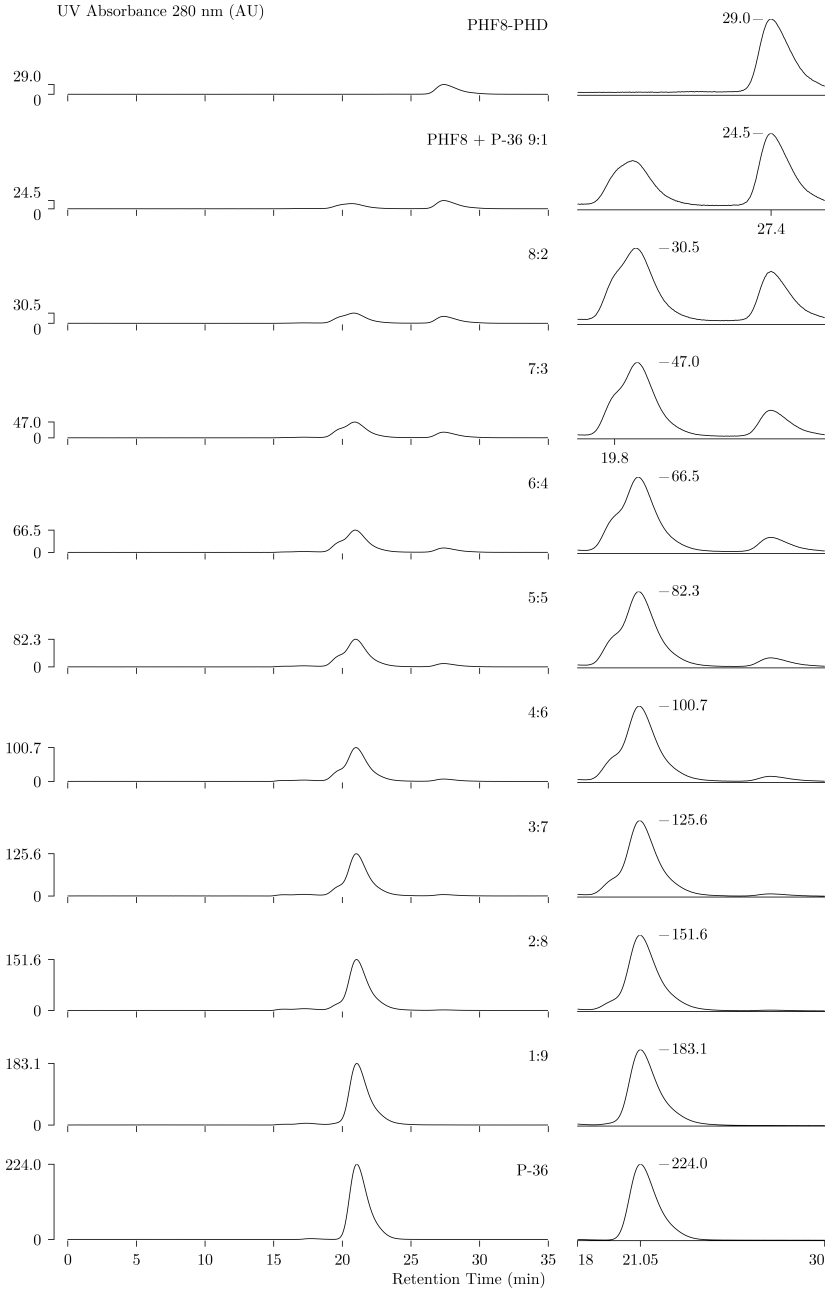


Figure 3.16: A titration of PHF8-PHD against P-36 at a fixed total protein concentration of 20 µM*. The column used for SEC was an S200, and the running buffer was PBS.*

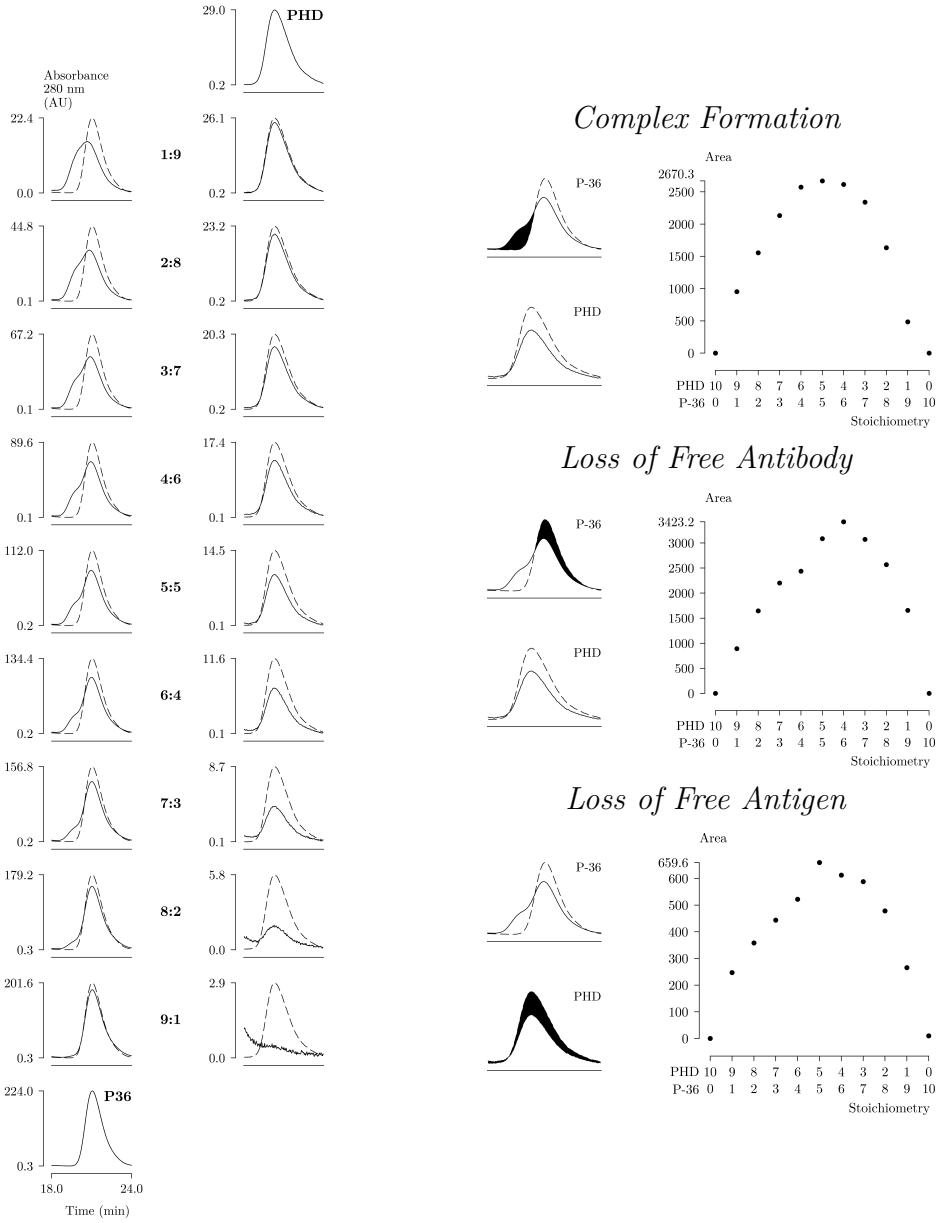


Figure 3.17: (Left) Antibody and antigen peaks extracted from the previous titration plotted with their theoretical UV traces assuming no binding. The molar ratio of antibody to antigen is indicated between the plots. (Right, from top) The difference in area between the theoretical peak (assuming no binding) and the measured peak for the complex of P-36 and PHD, P-36, and PHD. Theoretical peaks were calculated by scaling controls containing only PHF8-PHD or P-36 to the relevant stoichiometry.

At first glance the complex formation between P-36 and PHF8-PHD appears to be like that of two molecules interacting at a concentration above their K_D. But these plots cannot be solely explained by K_D. When the concentration of PHF8 is nine times that of P-36 (18 µM : 2 µM), approximately half of the antibody is in complex. From its definition, the K_D can be calculated as:

$$\begin{aligned} K_\mathrm{D} &= \frac{[A_b][A_g]}{[A_g A_b]} \\ &= \frac{([A_b^T] - [A_b A_g])([A_g^T] - [A_b A_g])}{[A_b A_g]} \\ &= \frac{(18-1)(17-1)}{18} \\ &= 0.9\,\mu\mathrm{M}\ (1\ \mathrm{sf.}). \end{aligned}$$

Where $[A_b]$ and $[A_g]$ are the concentration of unbound antibody and antigen, respectively; $[A_b^T]$ and $[A_g^T]$ are, respectively, the total antibody and antigen concentrations; and $[A_b A_g]$ is the concentration of the antibody-antigen complex. This is an apparent K_D – it takes into account dilution of the sample as it move through the size-exclusion chromatography column, and any fraction of antibody or antigen that is unable to bind, by representing an upper bound for their true K_D. The "true" K_D can be no lower than 0.9 µM. If this K_D is a faithful representation of the underlying affinity then when the antibody and antigen are equimolar at a concentration of 10 µM, over ten-times the K_D, essentially all molecules should be in complex; they are not.

The relationship between P-36 and PHF8-PHD is not seem as straightforward as that between P-45 and $\mathrm{PHF8}_{1\text{–}447}$, as it is not simply proportional to the antigen concentration (figure 3.11). The problem appears to be PHF8-PHD. When PHF8 and P-36 were mixed one-to-one at a concentration of 20 µM – 20 times the supposed K_D, little complex forms. Since the concentration exceeds the K_D so much, essentially all molecules that can

be in complex should be. But when the antigen concentration is increased relative to the antibody concentration, more complex forms. The antibody is therefore available for binding, but the appropriate antigen, presumably a fraction of the total antigen sample, is limited.

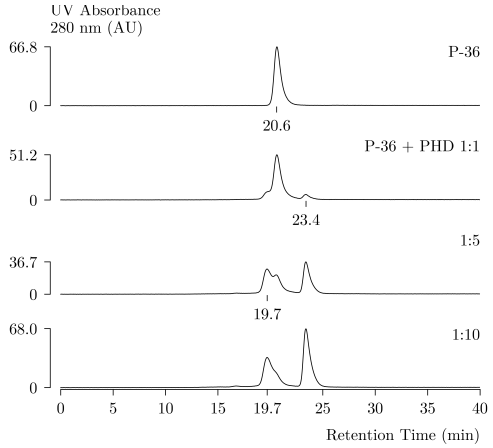


Figure 3.18: Size exclusion chromatograms for PHF8-PHD and P-36. The concentration of P-36 was 20 µM *with PHF8-PHD added at a 1-, 5-, and 10-fold molar excess. The column used for analysis was an S200, and the running buffer PBS.*

If the problem was once again the antigen being heterogeneous, and not all able to form a complex, then an important question remains: why does complex formation not follow the antigen concentration as it did for P-45 and $PHF8_{1–447}$? Perhaps the most straightforward resolution would be that the antibody-antigen complex was not in equilibrium when it was measured. Samples were injected in order, from highest PHF8 concentration to lowest, after over an hour of incubation. With each run taking over 40 min, the first sample is separated from the last by more than 6 h. If these samples had not reached equilibrium, earlier samples would appear to bind less than later samples. But as attractive as this explanation is, it requires accepting an extremely slow rate of complex formation. The search for further evidence in support of this hypothesis is, in part, the subject of chapter 5. Even without an explanation why, the results show antigen concentration appears to drive complex formation between P-36 and PHF8-PHD. This property has now been observed for two domains: the JmjC domain with P-45, and the PHD with P-36, suggesting that behind these observations is a general property.

P-23, P-2, and P-3

The remaining PHF8-PHD-specific antibodies though not forming complexes as readily as P-36, can be encouraged with an excess of antigen in a similar manner as P-36. After P-36, the next antibody studied was P-23:

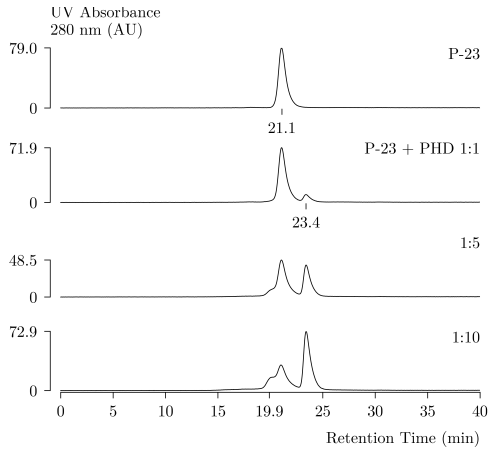


Figure 3.19: Size exclusion chromatograms for PHF8-PHD and P-23. The concentration of P-23 was 20 µM *with PHF8-PHD added at a 1-, 5-, and 10-fold molar excess. The column used for analysis was an S200, and the running buffer PBS.*

P-3 was one of the poorest binders:

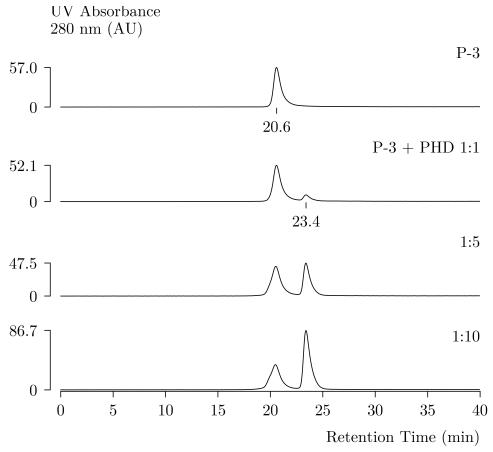


Figure 3.20: Size exclusion chromatograms for PHF8-PHD and P-3. The concentration of P-33 was 20 µM *with PHF8-PHD added at a 1-, 5-, and 10-fold molar excess. The column used for analysis was an S200, and the running buffer PBS.*

And although it was not investigated until later, with excess PHF8-PHD, P-2 formed a reasonable amount of complex.

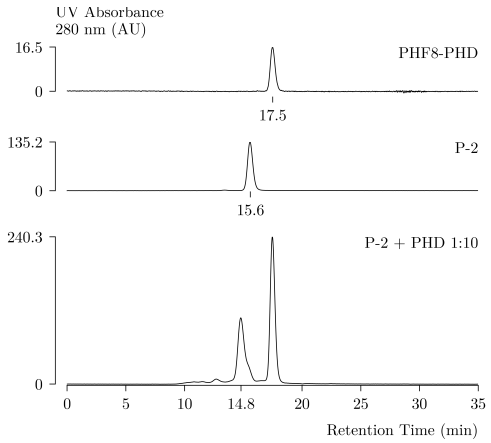


Figure 3.21: Size exclusion chromatograms for PHF8-PHD and P-2. The concentration of P-2 was 20 µM *with PHF8-PHD added at a 10-fold molar excess. The column used for analysis was an S200, and the running buffer PBS.*

Recognition of PHF8$_{1–447}$ is Possible

The antibody P-13, unlike the other antibodies for PHF8, easily formed a complex. Demonstrating that it is possible, and that at least one antibody capable of complex formation with PHF8$_{1–447}$ is contained within antibody response.

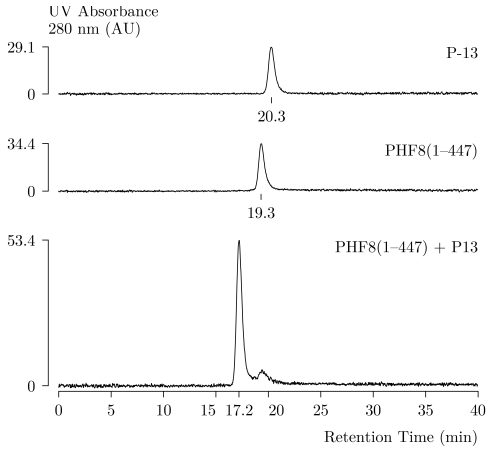
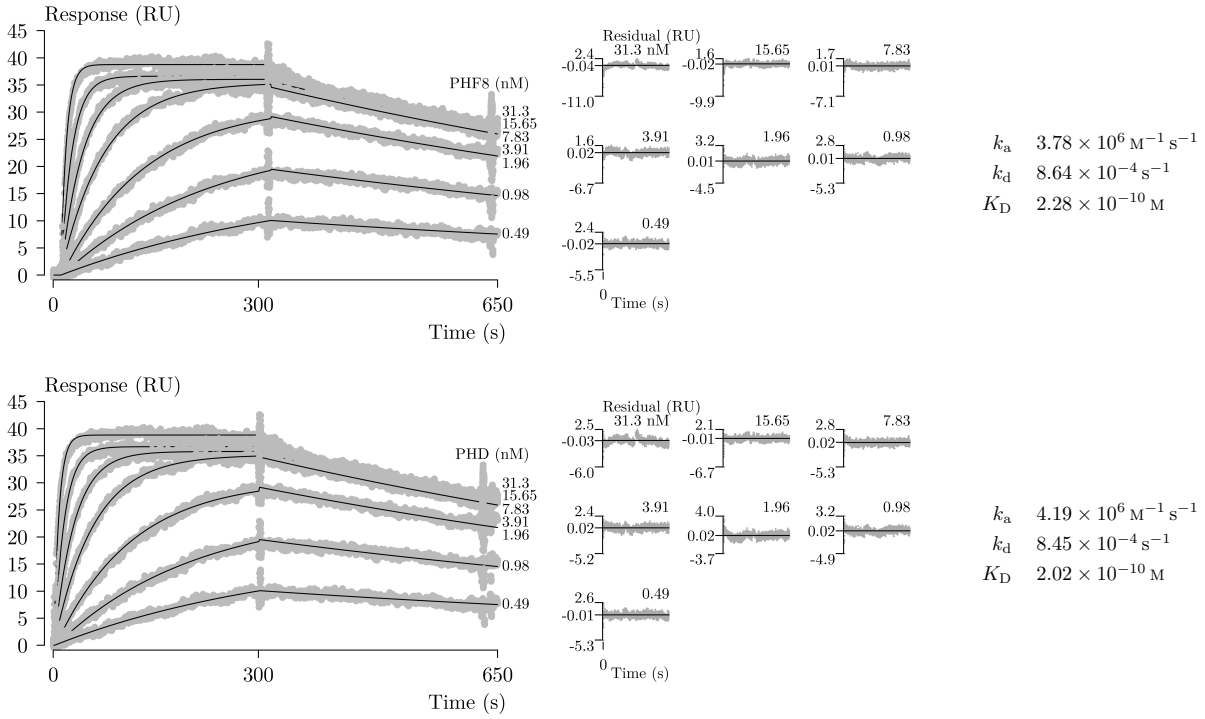


Figure 3.22: Size exclusion chromatograms showing that the antibody P-13 can form a complex with PHF8$_{1–447}$ in solution at a concentration of 1 µM.

The affinity of P-13 for PHF8$_{1–447}$, as measured by SPR, is subnanomolar. However, no effect on PHF8 activity was observed in initial testing, so the antibody was not investigated further.



k_a ($\mathrm{M^{-1}\,s^{-1}}$)	k_d ($\mathrm{s^{-1}}$)	K_D (M)
$3.98 \times 10^6 \divideontimes 1.05$	$8.54 \times 10^{-4} \divideontimes 1.01$	$2.15 \times 10^{-10} \divideontimes 1.06$

Figure 3.23: Duplicate SPR traces and kinetic values for the antibody P-13 binding to $PHF8_{1–447}$ in an SPR assay. Kinetic values are summarised by their geometic mean, and geometric standard deviation. P-13 was captured by surface-immobilised goat anti-rabbit Fab antibodies, and $PHF8_{1–447}$ injected at varying concentration.

Summary of Main Results

47 antibodies that recognised the KDM PHF8 by ELISA were isolated from the lymph nodes of two NZW rabbits. Four antibodies appeared to inhibit PHF8 enzymatic activity, but they were later shown to form little complex in a SEC assay. The seemingly contradictary observation is likely caused by antibodies that are for a subpopulation of the antigen. Though the nature of this subpopulation was not clear, binding in ELISA and not solution suggests partial denaturation of the antigen may be involved. The PHD of PHF8 was cloned and expressed, and four antibodies that recognise this domain also struggled to bind in solution despite binding in ELISA – implying the problem affected both the JmjC domain and PHD. Although most of the antibodies tested could not bind to PHF8 in SEC assays, P-13, bound to PHF8's JmjC domain with high affinity in an SPR assay, and efficiently in a SEC assay demonstrating that the problem of binding in ELISA but not in SEC does not effect all antibodies from the immunisation.

3.3 Conclusions

At the end of chapter 2, an observation was made that some antibodies that bound in an ELISA seemed not to bind in a SEC assay. The recovery of similar antibodies for PHF8 in this chapter suggests that there is a systematic error in antibody recovery for these two proteins – one that simply changing the antigen would not necessarily fix. What these experiments for PHF8 add is that these binding problems, for PHF8 at least, seem to come from an antigen that is heterogeneous. However, whether this heterogeneity is permanent or dynamic is unclear. What is likely is that complex formation probably requires an epitope be first exposed by partial denaturation of the antigen.

Two possible explanations of how antibodies binding to non-native epitopes could have been recovered here are that a bias in screening has some how favoured antibodies specific for denatured protein, or that during im-

munisation the antigen has denatured, and thus the immune response was direct primarily towards denatured protein. A practical resolution is to ask whether more antibodies that recognise PHF8's dominant solution state can be found using orthogonal techniques such as phage display from immune libraries. If these antibodies can be recovered, and in large numbers, it implies the problem was in screening. An orthogonal antibody recovery technique would also supply the starting point for inhibitor screening and structural studies, should those antibodies exist in the first place.

Even if the immunisation had been compromised from the beginning, and there were no antibodies useful for inhibition and structural studies, the nature of these antibodies, and the reasons for their generation are important. If it is a general property of intracellular proteins when they are immunised, then it has serious implications for this method of drug discovery. The affinity offered by antibodies recovered from an immunised host is important to antibody-inspired drug discovery. Antibody mimics of lower surface area, like small molecules and peptides, are almost guaranteed to lose affinity, and a high affinity starting point is insurance against this. But more generally, these binding problems potentially say something about PHF8 and KDM4D: that they have underlying properties, possible of biological relevance, that predispose them to this counter-intuitive antibody recognition.

References

[1] C. Loenarz et al. "PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an Nε-dimethyl lysine demethylase". In: *Human molecular genetics* 19.2 (2010), pp. 217–222.

[2] L. Yu et al. "Structural insights into a novel histone demethylase PHF8". In: *Cell research* 20.2 (2010), pp. 166–173.

[3] J. Qiu et al. "The X-linked mental retardation gene PHF8 is a histone demethylase involved in neuronal differentiation". In: *Cell research* 20.8 (2010), pp. 908–918.

[4] W. Feng et al. "PHF8 activates transcription of rRNA genes through H3K4me3 binding and H3K9me1/2 demethylation". In: *Nature structural & molecular biology* 17.4 (2010), pp. 445–450.

[5] J. R. Horton et al. "Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases". In: *Nature structural & molecular biology* 17.1 (2010), pp. 38–43.

[6] C. Huang et al. "Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4". In: *Cell research* 20.2 (2010), pp. 154–165.

[7] L. E. Siderius et al. "X-linked mental retardation associated with cleft lip/palate maps to Xp11.3-q21. 3". In: *American journal of medical genetics* 85.3 (1999), pp. 216–220.

[8] F. Laumonnier et al. "Mutations in PHF8 are associated with X linked mental retardation and cleft lip/cleft palate". In: *Journal of medical genetics* 42.10 (2005), pp. 780–786.

[9] F. Abidi et al. "A novel mutation in the PHF8 gene is associated with X-linked mental retardation with cleft lip/cleft palate". In: *Clinical genetics* 72.1 (2007), pp. 19–22.

[10] A. Koivisto et al. "Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate". In: *Clinical genetics* 72.2 (2007), pp. 145–149.

[11] Y. Qiao et al. "Autism-associated familial microdeletion of Xp11.22". In: *Clinical genetics* 74.2 (2008), pp. 134–144.

[12] C. Nava et al. "Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE". In: *Translational psychiatry* 2.10 (2012), e179.

[13] V. De Wolf et al. "A complex Xp11.22 deletion in a patient with syndromic autism: exploration of FAM120C as a positional candidate gene for autism". In: *American Journal of Medical Genetics Part A* 164.12 (2014), pp. 3035–3041.

[14] D. Kleine-Kohlbrecher et al. "A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation". In: *Molecular cell* 38.2 (2010), pp. 165–178.

[15] H. H. Qi et al. "Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development". In: *Nature* 466.7305 (2010), pp. 503–507.

[16] S. M. Smith, R. S. Kimyon, and J. J. Watters. "Cell-Type-Specific Jumonji Histone Demethylase Gene Expression in the Healthy Rat CNS: Detection by a Novel Flow Cytometry Method". In: *ASN Neuro* 6.3 (2014), AN20130050.

[17] N. E. Oey et al. "A neuronal activity-dependent dual function chromatin-modifying complex regulates arc expression". In: *eNeuro* 2.1 (2015), ENEURO–0020.

[18] W. Liu et al. "PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression". In: *Nature* 466.7305 (2010), pp. 508–512.

[19] T. Hori et al. "Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly". In: *Developmental cell* 29.6 (2014), pp. 740–749.

[20] H.-J. Lim et al. "The G2/M regulator histone demethylase PHF8 is targeted for degradation by the anaphase-promoting complex containing CDC20". In: *Molecular and cellular biology* 33.21 (2013), pp. 4166–4180.

[21] L. Sun et al. "Cyclin E-CDK2 protein phosphorylates plant homeodomain finger protein 8 (PHF8) and regulates its function in the cell cycle". In: *Journal of Biological Chemistry* 290.7 (2015), pp. 4075–4085.

[22] Z. Zhu et al. "PHF8 is a histone H3K9me2 demethylase regulating rRNA synthesis". In: *Cell research* 20.7 (2010), pp. 794–801.

[23] Ö. Erdoğan et al. "Proteomic dissection of LPS-inducible, PHF8-dependent secretome reveals novel roles of PHF8 in TLR4-induced acute inflammation and T cell proliferation". In: *Scientific reports* 6 (2016), p. 24833.

[24] Y. Tang et al. "Expression profiles of histone lysine demethylases during cardiomyocyte differentiation of mouse embryonic stem cells". In: *Acta Pharmacologica Sinica* 35.7 (2014), pp. 899–906.

[25] Y. Tang et al. "Plant homeo domain finger protein 8 regulates mesodermal and cardiac differentiation of embryonic stem cells through mediating the histone demethylation of pmaip1". In: *Stem Cells* 34.6 (2016), pp. 1527–1540.

[26] M. Björkman et al. "Systematic knockdown of epigenetic enzymes identifies a novel histone demethylase PHF8 overexpressed in prostate cancer with an impact on cell proliferation, migration and invasion". In: *Oncogene* 31.29 (2012), pp. 3444–3456.

[27] D. Tong et al. "The HIF/PHF8/AR axis promotes prostate cancer progression". In: *Oncogenesis* 5.12 (2016), e283.

[28] Y. Shen, X. Pan, and H. Zhao. "The histone demethylase PHF8 is an oncogenic protein in human non-small cell lung cancer". In: *Biochemical and biophysical research communications* 451.1 (2014), pp. 119–125.

[29] P. Shao et al. "Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis". In: *Nucleic acids research* 45.4 (2017), pp. 1687–1702.

[30] Q. Wang et al. "Stabilization of histone demethylase PHF8 by USP7 promotes breast carcinogenesis". In: *The Journal of clinical investigation* 126.6 (2016), p. 2205.

[31] X. Sun et al. "Oncogenic features of PHF8 histone demethylase in esophageal squamous cell carcinoma". In: *PLoS One* 8.10 (2013), e77353.

[32] M. F. Arteaga et al. "The histone demethylase PHF8 governs retinoic acid response in acute promyelocytic leukemia". In: *Cancer cell* 23.3 (2013), pp. 376–389.

[33] A. Yatim et al. "NOTCH1 nuclear interactome reveals key regulators of its transcriptional activity and oncogenic function". In: *Molecular cell* 48.3 (2012), pp. 445–458.

[34] X. Liu et al. "MiR-101 reverses the hypomethylation of the LMO3 promoter in glioma cells". In: *Oncotarget* 6.10 (2015), p. 7930.

[35] N. R. Rose et al. "Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases". In: *Journal of medicinal chemistry* 55.14 (2012), pp. 6639–6643.

[36] J. Dorosz et al. "Structure based design of a new scaffold for cell-penetrating peptidic inhibitors of the histone demethylase PHF8". In: *ChemBioChem* (2017).

[37] A. Kawamura. personal communication. 2015.

4 PHF8-PHD

4.1 Introduction

Many antibodies isolated in the previous chapter seem to not recognise PHF8 as it primarily exists in solution. What they do recognise is likely derived from PHF8 – for it not to be requires improbably many foreign contaminants to explain their specific cross-reactivity. But what directed these antibodies to this tiny fraction of PHF8-derived molecules was unclear. These antibodies could be symptoms of an antigen compromised by immunisation, or simply the outliers of an otherwise uneventful immunisation whose recovery was favoured by a bias in screening. PHF8 antibody P-13 bound well in solution, so it seemed reasonable that there may be others – be they exceptions to the norm or an overlooked majority. The problem was differentiating essentially full complex formation from partial complex formation.

The tiny antibody concentration in B-cell supernatants limited initial screening in the last chapter to assays, like ELISA, that amplify binding events. But without knowing the specific antibody concentration and their affinity for an antigen, weak affinity, strong affinity but low concentration, and recognition limited to a subset of antigen molecules cannot be discerned rigorously. In this chapter immune phage libraries are constructed, and PHF8's PHD panned in solution. By not being immobilised directly on plastic, the antigen's solution structure should be undisturbed. Any antibodies specific for only a subset of antigen molecules will behave as though they have weak affinity when forced to compete with those able to bind the antigen's primary solution form, and should be relegated by multiple rounds of panning.

I focus in this chapter on PHF8's PHD rather than the immunogen,

$PHF8_{1–447}$. Blocking recognition of H3K4me3 by this domain is enough to inhibit the turnover by the catalytic JmjC domain,[1] and the epitope of any antibody specific for the PHD is immediately placed within the PHD's 66 residues, rather than the immunogen's 447. The previous chapter identified antibodies that bound poorly in solution both for the PHD and JmjC domain, but the PHD offers the simpler model system to draw conclusions applicable to both. Structural studies on the PHD also need not rely on crystallography as they do for the JmjC whose molecular weight is over 40 kDa, as the 7 kDa PHD is readily accessible to NMR.

Phage Display

B-cells are custodians of monoclonality: once their single antibody clone is secreted it enters into a polyclonal solution, each antibody connected to its source B-cell only by sequence. This decoupling of antibody and B-cell is deliberate, evidenced by antibodies being auditioned while still tethered as a receptor to the cell membrane. Only after their source B-cell is selected is an antibody secreted. These secreted antibodies act at a distance, binding to their target antigen and recruiting elements of the immune system to any region they are permitted access. Antibody recovery begins here, with an antibody and an antigen. But, although antibodies may be identified, and even isolated as proteins, they must be recovered as DNA. Antibody recovery's task is, therefore, to find the genetic material from which an antibody is intentionally separated, curated by one B-cell, among a likeness of billions.

Unless selected, a B-cell's tendency *in vivo* is, by design, towards self-destruction. Perhaps by similar mechanism, this is also the impulse of B-cells cultured *in vitro*, their lifetime, not much longer than a week. This brevity is the constraint on which hybridoma and single-cell antibody recovery are designed, and from which they inherit their respective shortcomings. Their approach is to preserve, in B-cells, as much of an immune response as possible, so that it may be searched and desired antibody matched to

its source B-cell. Phage display's approach is instead to leave the corporeal limits of the B-cell for the bacteriophage. Resistant to temperature, pH, and time, the bacteriophage is a virus, optimised by evolution to proliferate its minimal, self-contained genome via a bacterial host. The format thus connects DNA and protein, but without the cellular machinery, superfluous to antibody recovery, that complicates B-cell-based techniques.

Engineering an antibody into the phage genome as a coat protein fusion, either directly or by separate, but similarly packaged "phagemid" DNA, imparts the antibody's specificity to its packaged DNA. Incubated with an immobilised antigen, phage particles are robust enough to tolerate washing away phage with little or non-specific affinity. Those whose displayed antibodies are specific for the antigen will remain bound, in numbers roughly proportional to their affinity for the antigen, and can be eluted with a low pH wash. If the solution of eluted phage is neutralised within minutes, it remains infectious, so incubated with the appropriate strain of bacteria the phage will replicate. Only centrifugation of the bacterial cells, and precipitation of phage are necessary to purify the amplified sub-library enriched for antigen-specificity. With each repetition of this "panning" cycle, those phage particles retained by chance or low affinity interaction, become fewer, outcompeted by genuine or stronger interactions.[a]

The phagemid system introduces an antibody-phage-coat-protein fusion as a separate plasmid which is integrated into the phage particle via a packaging signal.[2] Monoclonality of the polyclonal phage sub-library enriched by panning is restored by infecting bacteria subsequently cultured on solid media. Each plasmid shares the same origin of replication, and so only a single clone can be propagated to the infected bacteria's progeny. Thus each colony harbours a single antibody clone, that is, each colony is monoclonal. Culturing a single colony, and infecting it with helper phage, modified wild-type phage whose defective packaging signal causes the phagemid to be

[a]The relationship between affinity and amplification is, in reality, not quite so straightforward, as an antibody of poor affinity but exceedingly good expression will be amplified disproportionately. But, still panning remains an extremely efficient and versatile technique for the enrichment of large libraries of antibodies *in vitro*.

packaged with priority, will produce phage with a heterogeneous coat – some wild-type and some fused to the antibody body clone they package. There monoclonal phage particles can be screened by ELISA analogously to an antibody, and antibody-encoding DNA recovered from the original, infected colony by miniprep or similar.

Phage display's major weakness is in library construction. Antibody-encoding DNA must be introduced *en masse* into the phagemid, so B-cells are lysed, mRNA converted to cDNA, and antibody encoding DNA amplified selectively by PCR. But without a cell to contain them, heavy and light antibody chains, their mRNA transcribed separately, will mix. The libraries constructed here, scFv phagemid display libraries, use pull-through PCR to join heavy and light chain pairs as an scFvs and then ligate the fragment into the phagemid vector such that it becomes a coat protein fusion. The pairing of heavy and light chain in these scFvs is essentially random, likely to have little resemblance to the intended pairings. With sufficient amplification, all heavy chains could be paired with all light chains. The majority of pairs would be unnatural, but contained among them should be every natural ("cognate") pairing from the original immune response (though the cognate pair is note necessarily the best binding). But following ligation of scFv into phagemid, the DNA must be transformed into a bacterial host. The cells actually transformed are never more than a fraction of a percent, and transformation vessel size limited so that capturing all combinations of heavy and light chain becomes impractical. What saves phage display is firstly that many antibodies can remain effective in an unnatural pairing of heavy and light chain, and secondly that, in the case of immune phage display, antibodies have been pre-selected in an immune response for proliferation, their clones are therefore at an excess, and the recombination of cognate heavy and light chains favoured slightly.

4.2 Results

Immune Phage Display

Library Construction

Manually isolating B-cells and cloning the antibodies they express is time consuming enough that the previous chapters have been limited to a single niche: the lymph nodes. But in leaving antibody selection to the competition of panning, searching additional niches becomes practical. The natural separation of B-cells in lymph nodes, spleen, and bone marrow was used to construct six separate phage libraries, each one insulated from the clones of another and therefore, as much as possible, protected from certain antibodies completely dominating selection.

Library	Antigen	Rabbit	Source	Format	Size
PHF8-LN-1	PHF8	6142	Lymph Nodes	scFv	2.4×10^8
PHF8-LN-2	PHF8	6144	Lymph Nodes	scFv	8.0×10^7
PHF8-SP-1	PHF8	6142	Spleen	scFv	–
PHF8-SP-2	PHF8	6144	Spleen	scFv	1.4×10^8
PHF8-BM-1	PHF8	6142	Bone Marrow	scFv	–
PHF8-BM-2	PHF8	6144	Bone Marrow	scFv	2.6×10^7

Table 4.1: Phage libraries constructed for PHF8.

A small library may have many diverse binders and a large library few. Library size is really only a crude measure of potential for antibody discovery, but without sequencing the library in its entirety estimations of diversity are difficult. So, where it could be determined, approximate library size is indicated.[b] Assuming modest diversity, the total size of all libraries, which exceeds 4.9×10^8, should not itself impede antibody discovery.

[b] A library's size can only be measured once during construction, and a problem with culture plates meant for two libraries it could not be estimated.

Panning PHF8-PHD on Plastic

Each library was first panned against PHF8-PHD immobilised on plastic as it would be in an ELISA. ELISA was the assay in which all antibodies from the previous chapters appeared to bind regardless of their binding in solution, so PHF8-PHD displayed analogously should solicit at least some antibodies. Though the number of hits after two rounds of panning was low, among them are families of similar sequence – an indication that panning enriched the libraries, and that further panning risked losing diversity.

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-10.14	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.21	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.25	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.26	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.29	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.30	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.31	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.37	GFSLSSYAMS	VISRSGRTYYTR	G-----YPNRVDI	QSSQSVYNNNNLA	KASTLAS	AGGGSLPSDNT
P-10.4	GFSLSSYDMS	IISAGGNPYYAT	DSA--YWADAFDP	QSSQSVYNNNWLA	GASTLAS	QGGYPSGDARA
P-10.5	GFSLDSYDMS	IIAVSGNTYYAS	DSA--YWTDAFDP	QASQSVYNKNWLA	GASTLAS	QGGYGSGDARA
P-10.10	GFSLSSYDMN	IISSRGNPYYAS	EFYPDVGGGIEDL	QSSKSVYNKNWLA	GASTLAS	QGGYGSGDARA
P-10.8	GFSLSSYVMS	IIYASGSTSYAS	HPW--RRGWGDYL	QASQSIGSS-YLS	SASTLDS	AG--GYSDNRA
P-10.1	GIDLSRYAVI	IISRSATTYYPS	D----HPGYSFNI	QASQNIVG--RLA	SASNLAS	QGYYGGGIYG-
P-10.52	DIDLSSYHMS	ISSSSSNTYYAS	D-----DGSGSSI	QSSESVYSN-WLA	ATSTLDS	EGGYSSGDART
P-10.2	GFSLNNYAMT	TISDGGTAYYAS	---------SVGL	QSSQSVHNNNWLS	QASKLAS	QGGYSSGDAFA

Table 4.2: Sequences of antibodies isolated from panning on PHF8-PHD immobilised directly on plastic.

Among this modest, though diverse panel of antibodies are two whose sequences are familiar: P-10.1 resembles P-2 from the previous chapter:

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-2	GIDLSRYAVI	IISRSATTYYAS	DHPGYSFNI	QASQSVYNDNTLA	DSSNLAS	QG-YYSGAIYA
P-10.1	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQNIVG--RLA	SASNLAS	QG-YYGGGIYG

and P-10.52 is similar to P-23 and P-50:

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-50	GIDLSSYHMT	IISTGGTTYYAT	DDGSGSSI	QSSQSVYSNWLA	AASTLDS	EGGYSSGDGRA
P-23	GIDLSSYHMT	IISTGGTTYYAT	DDGSGSSI	QSSQSVYSNWLA	AASTLDS	EGGYSSGDGRA
P-10.52	DIDLSSYHMS	ISSSSSNTYYAS	DDGSGSSI	QSSESVYSNWLA	ATSTLDS	EGGYSSGDART

The heavy and light chains, which in constructing a phage library are separately amplified and recombined randomly, are in P-10.52 and P-10.1 found essentially in their original, cognate pairing which isolation from single B-cells preserves. Unless the heavy and light chain combination in these antibodies is particularly well expressed in E. coli, or the combination is the only one that has affinity for PHF8-PHD, then this particular heavy and light chain pairing, among 10^{14} alternatives[c], suggests the original antibody was in considerable excess. Which is to say, these antibodies were dominant clones, their host B-cells selected during the immune response for their ability to recognise the antigen and allowed to proliferate.

Panning PHF8-PHD in Solution (Rounds 1–2)

Phage display is used here to efficiently search an antibody response to PHF8 for antibodies against the dominant solution form of PHF8's PHD. Panning on plastic-immobilised PHF8-PHD enriched phage libraries for this subset of the original PHF8 immunogen, but returned antibodies, P-10.1 and P-10.52, whose sequences are almost identical to those that appeared not to recognise PHF8-PHD's dominant solution form in the previous chapter. If there are antibodies in these libraries that recognise PHF8-PHD's

[c]Libraries were constructed from approximately 10^7 cells.

dominant solution form, the appearence of P-10.1 and P-10.52 suggests the immobilisation of PHF8-PHD has disadvantaged their selection.

Presenting PHF8-PHD in its more natural solution form requires a chemical or biochemical handle. Chemical biotinylation of PHF8-PHD is facilitated by three lysines, two of which only two are at the C-terminus, part of a likely unstructured tail. Tethered to a streptavidin surface, the structure of biotinylated PHF8-PHD should be undisturbed. If biotinylated PHF8-PHD is incubated with a phage library before being captured (along with any phage bound to it) on streptavidin-coated beads, any surface binding effects can be minimised. Incubating biotinylated PHF8-PHD at a concentration of 1 µM with 500 µL of blocked phage libraries returned, after two-rounds, the following clones.

		CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
PHF8-SP-1	P-11.15	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQNIVGR--LA	SASNLAS	QGY-YGGGIY---G
	P-11.16	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQNIVGR--LA	SASNLAS	QGY-YGGGIY---G
	P-11.17	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQNIVGR--LA	SASNLAS	QGY-YGGGIY---G
	P-11.22	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.23	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.25	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.27	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.28	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.29	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.30	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.33	GIDLSRYAMI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.34	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIGST--LA	YASTLAS	QGY-YSGGIY---A
	P-11.13	GFSLSSSWMT	TIYAGTGRTWYAS	GYFSTNGTLN-----L	QASQNIHSN--LA	SAAKLAS	QSY-YGLTSSYN-G
	P-11.35	GFSLSAYWMS	IIFP-SGSTYYAT	-DPGSSRD-------I	QSSQSVYKNNYLA	QASKLAS	LAG-YDCSSADCTA
PHF8-LN-2	P-11.42	GIDLSSYHMT	IISS-AGNTYYAT	DDGSGSSI--------	QSSQSVYSN-WLA	ATSTLDS	EGG-YSSGDGR--A
	P-11.43	GIDLSSYHMR	IIIS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.45	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.46	GIDLSSYHMT	IISS-AGNTYYAT	DDGSGSSI--------	QSSQSVYSN-WLA	ATSTLDS	EGG-YSSGDGR--A
	P-11.51	GIDLSSYHMT	IISS-AGNTYYAT	DDGSGSSI--------	QSSQSVYSN-WLA	ATSTLDS	EGG-YSSGDGR--A
	P-11.60	GIDLSSYHMT	IISS-AGNTYYAT	DDGSGSSI--------	QSSQSVYSN-WLA	ATSTLDS	EGG-YSSGDGR--A
	P-11.61	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.64	GIDLSNYHMN	IISS-SDNTYYAS	DDGSGSSV--------	QSSQSVYSN-WLS	AASTLES	QGG-YGSGDAF--A
	P-11.65	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.66	GFSLSSNAIN	IISS-SDNTYYAS	DDGSGSSV--------	QSSQSVYSN-WLA	GTSTLDS	EGG-YPSGDAR--T
	P-11.67	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.68	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.47	GIDLSRYAVI	IISR-SATTYYPS	DHPGYSFN-------I	QASQSIYSG--LA	KASTLDS	QGY-VSGGIF---A
PHF8-LN-1	P-11.37	GIDLSSYHMT	IISS-AGNTYYAT	DDGSGSSI--------	QSSQSVYSN-WLA	ATSTLDS	EGG-YSSGDGR--A
	P-11.38	GIDLSSYHMR	IISS-AGNTYYAT	DDGSGSSI--------	QASQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.41	GFSLNNYAMT	IIAR-SGNTYYAT	ESYDDYGDYVGNYWNI	QSSQSVYNNNWLS	AASTLAS	QGG-YSSGDAN--A
	P-11.39	GFTISSYHMS	YISV-GGSAYYAS	GAPLGSAVYG---MDP	QSSESVYNNNYLS	DASKLAS	QGTYYSSGWWG--A
	P-11.36	GFSLNNYYMS	SITS-DGITYYAN	RNGGATRD-------I	QASESVYSNNRLA	LASTLTS	VG--YKDIATDGIA
	P-11.40	GIDLSTYHMG	IISY-GDNAYYAS	EAFSTGGD-------I	QSSQSVYNNNRLA	EASKLTS	QGG-YGSGDAR--A
PHF8-SP-2	P-11.3	GIDLSSYHMT	IISS-SGTTYYAS	DDGSGSSI--------	QSSQSVYSN-WFA	ATSTLDA	EGG-FSSGDAR--A
	P-11.9	GIDLSSNAMI	TIIP-TGGTAYAS	AGSGTGML--------	QASQSVYNNNNLA	GASPLAS	QGY-YSGGIF---V

Table 4.3: Antibody sequences isolated from panning PHF8-PHD in solution arranged by source library, and size of heavy chain CDR3 family.

These sequences (table 4.3) represent the antigen-recognising phage particles recovered from 94 colonies picked for each of the six libraries panned on PHF8-PHD. The recovered phage particles were screened by phage ELISA both on PHF8-PHD immobilised directly on plastic, as well as biotinylated PHF8-PHD, though there appeared to be no clones with a strong preference for a plastic- or streptavidin-immobilised antigen. Again there are sequences which have become familiar, but elaborated further than before, strengthening the argument that they are representatives of a common, dominant clone. There are five unique sequences related to P-23 and P-50 of the previous chapter:

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-23	GIDLSSYHMT	IISTGGTTYYAT	DDGSGSSI	QSSQSVYSNWLA	AASTLDS	EGGYSSGDGRA
P-50	GIDLSSYHMT	IISTGGTTYYAT	DDGSGSSI	QSSQSVYSNWLA	AASTLDS	EGGYSSGDGRA
P-11.3	GIDLSSYHMT	IISSSGTTYYAS	DDGSGSSI	QSSQSVYSNWFA	ATSTLDA	EGGFSSGDARA
P-11.37	GIDLSSYHMT	IISSAGNTYYAT	DDGSGSSI	QSSQSVYSNWLA	ATSTLDS	EGGYSSGDGRA
P-11.38	GIDLSSYHMR	IISSAGNTYYAT	DDGSGSSI	QASQSVYSNWFA	ATSTLDA	EGGFSSGDARA
P-11.64	GIDLSNYHMN	IISSSDNTYYAS	DDGSGSSV	QSSQSVYSNWLS	AASTLES	QGGYGSGDAFA
P-11.66	GFSLSSNAIN	IISSSDNTYYAS	DDGSGSSV	QSSQSVYSNWLA	GTSTLDS	EGGYPSGDART

and three sequences resembling P-2, though they differ slightly in CDR-L1:

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
P-2	GIDLSRYAVI	IISRSATTYYAS	DHPGYSFNI	QASQSVYNDNTLA	DSSNLAS	QGYYSGAIYA
P-11.15	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQNIVGR--LA	SASNLAS	QGYYGGGIYG
P-11.22	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIGST--LA	YASTLAS	QGYYSGGIYA
P-11.33	GIDLSRYAMI	IISRSATTYYPS	DHPGYSFNI	QASQSIGST--LA	YASTLAS	QGYYSGGIYA

Absent, after two rounds of solution-based panning, were antibodies similar to P-10.14 (CDR-H3: `GYPNRVDI`), the most prevalent clone recovered by panning on plastic-immobilised PHF8-PHD. Steptavidin- or plastic-immobilisation of PHF8-PHD changes the antibodies enriched by panning, and so partially validates the hypothesis that screening bias effected antibody recovery. But the prevalent clones from solution based panning, are clones already known to not recognise PHF8-PHD well in solution. There is a difference caused by antigen-immobilisation method, but not one that can alone explain why these clones, whose binding is apparently impaired, continue to be isolated.

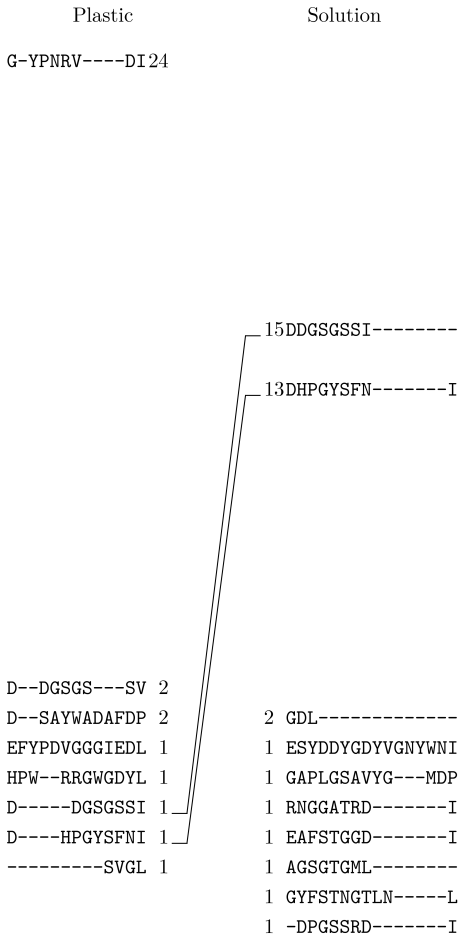


Figure 4.1: CDR-H3 sequences of antibodies isolated from panning on plastic-immobilised PHF8-PHD and biotinylated PHF8-PHD free in solution. The most dominant family of antibodies `GYPNRVDI` *disappeared from the output in solution-based panning, and less prevalent sequences,* `DDGSGSSI` *and* `DHPGYSFNI`*, dominate the output of solution-based panning.*

It was an unfortunate oversight that during library construction redundancy in the genetic code was not exploited to give each library a unique "barcode" in regions of shared sequence. These genetic barcodes would unambiguously identify each antibody's source even if they were contaminating other libraries. Instead, the most probable source library must be inferred

from each clones distribution among libraries. The most prevalent CDR-H3 sequence, `DDGSGSSI`, likely originates from the lymph nodes of rabbit 2:

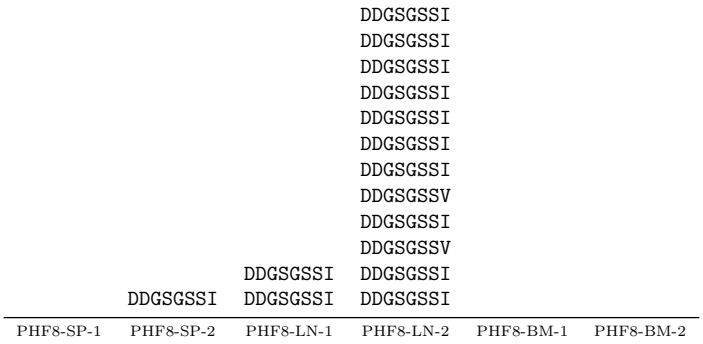


Figure 4.2: Distribution of the `DDGSGSSI` *CDR-H3 sequence was concentrated primarily in a library derived from the lymph nodes of rabbit 2.*

and the second most prevalent CDR-H3 sequence, `DHPGYSFNI`, likely comes from the spleen of rabbit 1:

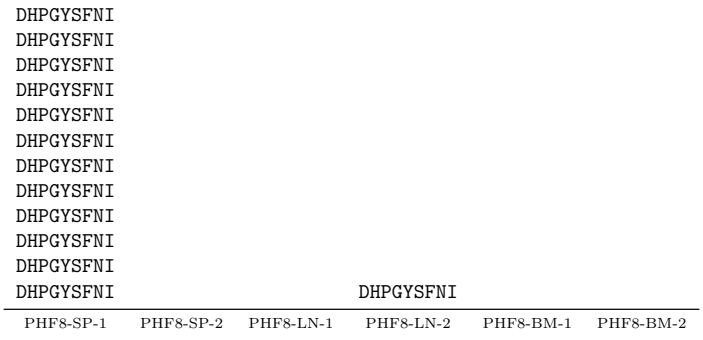


Figure 4.3: The distribution of the `DHPGYSFNI` *CDR-H3 was concentrated around a library derived from the spleen of a rabbit 1.*

Where these clones have spread to other libraries is probably due to contaminantion. When libraries are panned concurrently it is difficult not to spread the robust, easily aerosolised phage particle between independent

samples. It is very unlikely that these clones would be shared between two different animals, and though they may be found in different niches of the same animal, their low numbers, and no clearly distinguishing differences in sequence make contamination a more reasonable explanation.

Panning PHF8-PHD in Solution (Round 3)

After two rounds, the hit-rate for some libraries was still low, so there seemed to be value continuing to another round. From the 94 colonies picked for each of the six libraries panned, the following sequences were returned.

		CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
PHF8-SP-1	P-14.2	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.4	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.5	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.6	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.8	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.10	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.11	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.12	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.13	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.16	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.17	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.19	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.20	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.21	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.22	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.24	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.25	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.26	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.27	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYSGG---IYA
	P-14.1	GFSLSAYWMS	IIFPSGSTYYAT	D----PGSSRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-14.15	GFSLSAYWMS	IIFPSGSTYYAT	D----PGSSRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-14.29	GFSLSAYWMS	IIFPSGSTYYAT	D----PGSSRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-14.7	GFSLSTYALG	IISSSGSTAYAS	IG-NSDDERLDL	QASESIY--SGLA	AASTLAS	QQGYSSINV-DNA
PHF8-LN-1	P-14.30	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.31	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.32	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.34	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.35	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.36	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.38	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.39	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.40	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.41	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.42	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.43	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.44	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.45	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
	P-14.46	GFTISSYHMS	YISYGGSAYYAS	VG--YAGYGYDI	QASQSVYNNNWLA	QASKLAS	AGGYSGH---IYA
PHF8-SP-2	P-14.49	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.50	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.68	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.51	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.52	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.53	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.54	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.62	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.83	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.60	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.85	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.59	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.55	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.58	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.87	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.61	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---ILW
	P-14.67	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.72	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.77	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.69	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.70	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---ILW
	P-14.76	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
	P-14.56	GIDLSRYAVI	IISRSATTYYPS	D---HPGYSFNI	QASQNIV--GRLA	SASNLAS	QGYYGGG---IYG
PHF8-BM-1	P-14.88	GFSLSSFYMS	IIYPSGTTWYAS	GYSTSNTYIWGL			
	P-14.91	GFSLSSYYMS	IIYPSGTTWYAS	GYSTSNTYIWGL			
	P-14.92	GFSLSSYYMS	IIYPSGTTWYAS	GYSTSNTYIWGL			
	P-14.96	GFSLSSYYMS	IIYPSGTTWYAS	GYSTSNTYIWGL			

Table 4.4: (And overpage) Antibody sequences returned by panning PHF8-PHD in solution. Sequences arranged by source library and size of heavy chain CDR3 family. scFvs containing only a heavy chain were isolated from the library PHF8-BM-1.

		CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
PHF8-LN-2	P-18.82	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.84	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.36	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.68	DIDLSTYHMS	IISSSGSTYYAS	DDGS-GSSI	QSSQSVYS-NWLA	GASKLAS	QGGYG--SGDARA
	P-18.72	GIDLSNYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRI
	P-18.94	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.25	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.22	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.33	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.41	GFYLSSYDMK	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWLA	ATSTLDA	EGGYT--SGDARA
	P-18.56	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.64	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.91	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.93	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.48	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.53	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.57	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.27	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.31	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.44	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.23	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.32	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.37	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.38	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.39	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.43	GIDLSSYHMT	IISSAGNTYYAT	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.55	GIDLSTYHMT	IISTGGNTYYAN	DDGS-GSSI	QSSQSVYS-NWLA	ATSTLDS	EGGYS--SGDGRA
	P-18.61	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.71	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.87	GIDLSSYHMR	IISSAGNTYYAT	DDGS-GSSI	QASQSVYS-NWFA	ATSTLDA	EGGFS--SGDARA
	P-18.77	DIDLSTYHMS	IISSSGSTYYAS	DDGS-GSSI	QSSQSVYS-NWLA	GASKLAS	QGGYG--SGDARA
	P-18.78	DIDLSTYHMS	IISSSGSTYYAS	DDGS-GSSI	QSSQSVYS-NWLA	GASKLAS	QGGYG--SGDARA
	P-18.54	GIDLSNYHMN	IISSSDNTYYAS	DDGS-GSSV	QSSQNVYS-NWLA	GASTLAS	EGGYS--HGDARV
	P-18.81	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.52	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.88	GIDLSTFHMS	VISSSDNTYYTT	DDGS-GSSV	QSSQSVYS-NWLA	ATSTLDS	EGGYS--DGDARA
	P-18.30	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.66	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.80	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.83	GFSLSSNAIN	IISSSDNTYYAS	DDGS-GSSV	QSSQSVYS-NWLA	GTSTLDS	EGGYP--SGDART
	P-18.92	GIDLSTFHMS	VISSSDNTYYTT	DDGS-GSSV	PSSRSVYS-NWLA	ATSTLDS	EGCYS--DGDARA
	P-18.1	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.18	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.6	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.7	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.16	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.15	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.13	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.20	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.10	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.11	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.50	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIY--SGLA	KASTLDS	QGYVS--GGIF-A
	P-18.21	GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI	QASQSIG--STLA	YASTLAS	QGYYS--GGI-YA
	P-18.26	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.29	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.49	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.62	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.70	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.59	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.65	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.46	GIDLSTYHMG	IISYGDNAYYAS	EAFSTGGDI	QSSQSVYNNNRLA	EASKLTS	QGGYG--SGDARA
	P-18.3	GFSLSAYWMS	IIFPSGSTYYAT	DPGS-SRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-18.17	GFSLSAYWMS	IIFPSGSTYYAT	DPGS-SRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-18.19	GFSLSAYWMS	IIFPSGSTYYAT	DPGS-SRDI	QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA
	P-18.5	GFSLSAYWMS	IISPGGDTYYAS	TLRY-SYNI	QASESIG--SSLA	VASTLAS	AGGYS--RSSENA
	P-18.35	GFDLSSYAMG	FIDTGDSAYYAN	DPS--YFNI	QASQSVYNDNALS	KASTLAS	LGTYD--DDSSNL

A third round of panning introduced a new antibody from the PHF8-BM-1 library (CDR-H3: `GYSTSNTYIWGL`), but all clones were truncated after the heavy chain. Expressed without their phage coat protein fusion they would likely be unstable, and so were not recovered. This being the only antibody recovered from any bone marrow derived library may be due to an error in library construction. The family of antibodies from PHF8-LN-1 characterised by the heavy-chain CDR3 sequence `VGYAGYGYDI` was not seen in round two, but it was prevalent in round 3. The other clones are mostly amplified from round two, which is expected. The spread of some of these sequences to the outputs of different libraries is more likely caused by contamination than it is coincidence or a genuine relationship.

Solution Binding of Immune Phage Antibodies

Those antibodies that could be cloned from their phage coat protein fusions are given below.

	CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
2-13	GIDLSRYAVI	IISRSATTYYPS	DHPGY--SFNI	QASQSIGS--TLA	YASTLAS	-QGYY---SGGIYA
2-14	GIDLSRYAVI	IISRSATTYYPS	DHPGY--SFNI	QASQSIGS--TLA	YASTLAS	-QGYY---SGGIYA
2-1	GIDLSRYAVI	IISRSATTYYPS	DHPGY--SFNI	QASQNIVG--RLA	SASNLAS	-QGYY---GGGIYG
2-4	GFSLSSYDMS	IISAGGNPYYAT	DSAYWADAFDP	QSSQSVYNNNWLA	GASTLAS	-QGGYP--SGDARA
2-5	GFSLDSYDMS	IIAVSGNTYYAS	DSAYWTDAFDP	QASQSVYNKNWLA	GASTLAS	-QGGYG--SGDARA
2-12	GFSLSAYWMS	IIFPSGSTYYAT	DPGSS---REI	QSSQSVYKNNYLA	QASKLAS	-LAGYDCSSADCTA
2-15	GFSLSAYWMS	IIFPSGSTYYAT	DPGSS---RDI	QSSQSVYKNNYLA	QASKLAS	-LAGYDCSSADCTA
2-16	GFTISSYHMS	YISYGGSAYYAS	VGYAG-YGYDI	QASQSVYNNNWLA	QASKLAS	-AGGY---SGHIYA
2-10	GFSLSSYAMS	VISRSGRTYYTR	GWPNR---VDI	QASQSISSS-YLS	EASTLAS	QSAVYSTT--YLGT
2-11	GFSLSNYHMS	IISSSGSTYYAS	DDG---SGSSV	QSSQSVYSN-WLA	ATSTLDA	-EGGFS--SGDARA
2-17	GIDLSSNAMI	TIIPTGGTAYAS	AGSGT--GM-L	QASQSVYNNNNLA	GASPLAS	-QGYY---SGGIFN
2-3	GFSINNYHMS	IISDSGSTFYAS	DSY---ADFNL	QSSQNVYNNNYLS	GASTLAS	-QGGYP--SGDARA
2-2	GFSLNNYAMT	TISDGGTAYYAS	-------SVGL	QSSQSVHNNNWLS	QASKLAS	-QGGYS--SGDAFA

Figure 4.4: Selected antibodies identified from rounds 2 and 3 of phage panning and cloned as scFvs.

Not all of these antibodies could be expressed. Lacking constant regions, scFvs can often be unstable – especially after losing their phage coat protein fusion partner. Those scFvs that could be expressed, were purified by nickel affinity chromatography without SEC (unless these antibodies can recognise the dominant solution form of PHF8-PHD, they are not particularly useful, and complex formation should be obvious without the additional purifica-

tion step). The affinity-purified scFvs were first mixed with PHF8-PHD and incubated for at least half an hour in preparation for SEC analysis. However, when scFv and PHF8-PHD were mixed, a precipitate formed, slowly at first, but after several hours it was enough to prevent analysis by SEC.

It is very uncommon for two proteins that are stable alone, and that do not interact, to precipitate when mixed. Precipitation upon mixing is more often evidence a complex is unstable, that the complex's pI approaches the buffers pH and so becomes insoluble, or that complexation encourages aggregation. Changing the antigen to $PHF8_{1–447}$ shouldn't prohibit binding (after all, it is the original immunogen), but it will change the pI of an antibody complex, and may prevent aggregation. So $PHF8_{1\text{-}447}$ was combined at a concentration of 10 µM with each scFv at their concentration following purification. No precipitation was observed; though nor was strong evidence of binding.

Antibody 2-2

CDR-H1	CDR-H2	CDR-H3
FSLNNYAMT	TISDGGTAYYAS	SVGL
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQSVHNNNWLS	QASKLAS	QGGYSSGDAFA

Antibody 2-3

CDR-H1	CDR-H2	CDR-H3
GFSINNYHMS	IISDSGSTFYAS	DSYADFNL
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQNVYNNNYLS	GASTLAS	QGGYPSGDARA

Antibody 2-4

CDR-H1	CDR-H2	CDR-H3
GFSLSSYDMS	IISAGGNPYYAT	DSAYWADAFDP
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQSVYNNNWLA	GASTLAS	QGGYPSGDARA

Antibody 2-5

CDR-H1	CDR-H2	CDR-H3
GFSLDSYDMS	IIAVSGNTYYAS	DSAYWTDAFDP
CDR-L1	**CDR-L2**	**CDR-L3**
QASQSVYNKNWLA	GASTLAS	QGGYGSGDARA

Antibody 2-11

CDR-H1	CDR-H2	CDR-H3
GFSLSNYHMS	IISSSGSTYYAS	DDGSGSSV
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQSVYSNWLA	ATSTLDA	EGGFSSGDARA

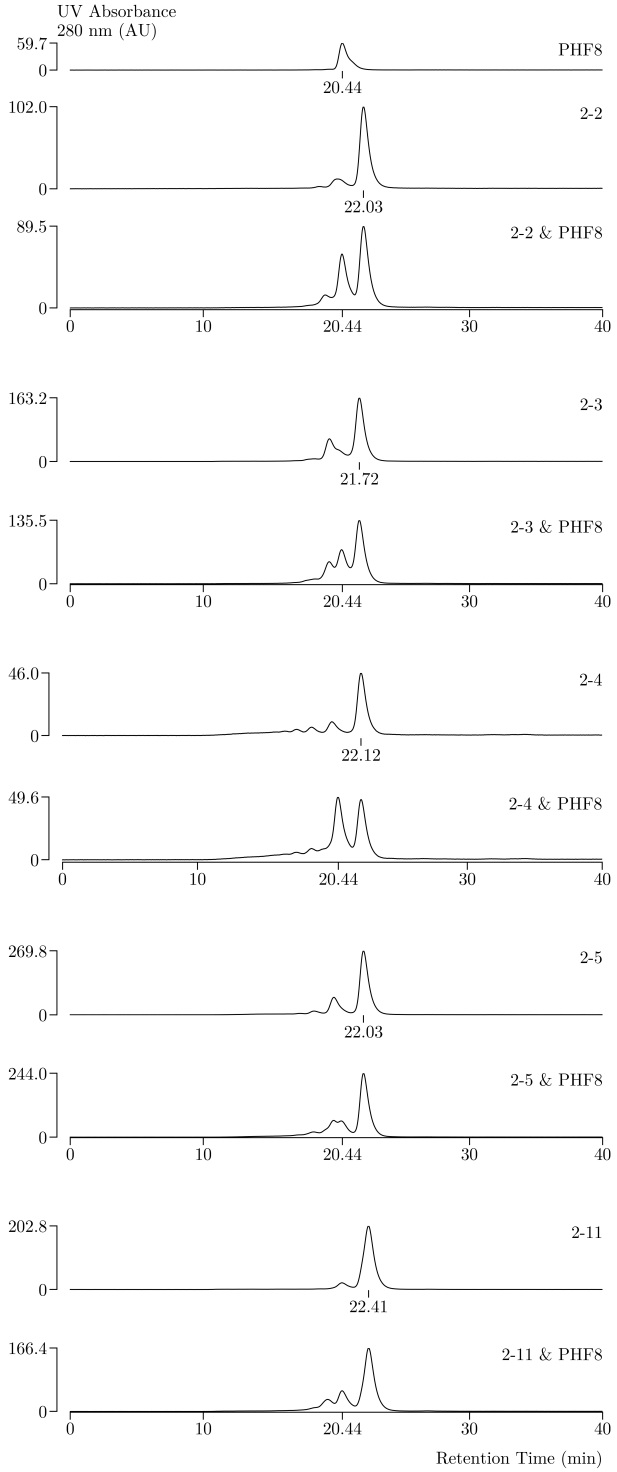


Figure 4.5: (And overpage) SEC chromatograms for scFvs and scFvs combined with PHF8$_{1–447}$. Evidence of binding (a peak of lower rentention time than that of the PHF8-only control) is rarely seen, and when it is, it is not consistent with full complexation, despite the PHF8 concentration being 10 µM.

Antibody 2-12

CDR-H1	CDR-H2	CDR-H3
GFSLSAYWMS	IIFPSGSTYYAT	DPGSSREI
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA

Antibody 2-13

CDR-H1	CDR-H2	CDR-H3
GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI
CDR-L1	**CDR-L2**	**CDR-L3**
QASQSIGSTLA	YASTLAS	QGYYSGGIYA

Antibody 2-14

CDR-H1	CDR-H2	CDR-H3
GIDLSRYAVI	IISRSATTYYPS	DHPGYSFNI
CDR-L1	**CDR-L2**	**CDR-L3**
QASQSIGSTLA	YASTLAS	QGYYSGGIYA

Antibody 2-15

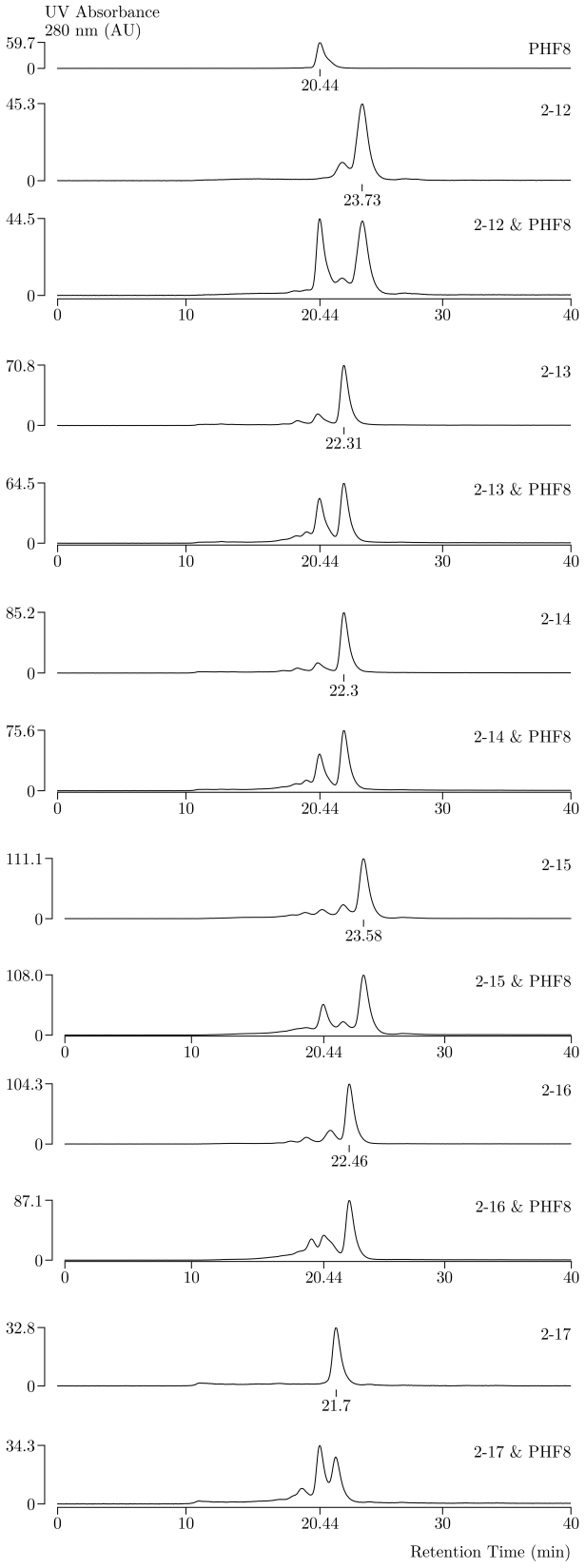
CDR-H1	CDR-H2	CDR-H3
GFSLSAYWMS	IIFPSGSTYYAT	DPGSSRDI
CDR-L1	**CDR-L2**	**CDR-L3**
QSSQSVYKNNYLA	QASKLAS	LAGYDCSSADCTA

Antibody 2-16

CDR-H1	CDR-H2	CDR-H3
GFTISSYHMS	YISYGGSAYYAS	VGYAGYGYDI
CDR-L1	**CDR-L2**	**CDR-L3**
QASQSVYNNNWLA	QASKLAS	AGGYSGHIYA

Antibody 2-17

CDR-H1	CDR-H2	CDR-H3
GIDLSSNAMI	TIIPTGGTAYAS	AGSGTGML
CDR-L1	**CDR-L2**	**CDR-L3**
QASQSVYNNNNLA	GASPLAS	QGYYSGGIFN



Among these chromatograms of various scFvs combined with $PHF8_{1–447}$ there is little evidence of a strong peak with retention time lower than PHF8's 20.44 min. Where there is a peak, for example antibodies 2-2, 2-16, and 2-17, it is small, reminiscent of the partial complex formation only seen in the previous chapters at extreme antigen excess. There remain some antibodies that did not express, but even if they were to be cloned as Fabs to improve stability, there are not enough to make the chances favourable that at least one antibody would recognise the dominant state of PHF8-PHD in solution, block enzymatic activity, and crystallise. It seems that at least for the PHD domain of PHF8, and probably the rest of the protein, the immunisation or immune response was some how compromised. If there are any antibodies that recognise PHF8's dominant solution state contained in these immune responses, they are likely few.

Preimmune Phage Display

Panning on PHF8-PHD

Even if it was understood what has caused so many antibodies to be specific for a seemingly insignificant form of PHF8-PHD, little could be changed. An immune response can be encouraged by immunisation with adjuvants, and a response directed to certain areas of a protein with an appropriately limited immunisation construct, but few other details of an immune response are within experimental control. So immunisation and the superior affinity it offers was set aside, and pre-immune llama phage libraries already available[3] panned on PHF8-PHD. A pre-immune library is created with B-cells derived from an animal before it has been exposed to anything but environmental immunogens, and so offers the enrichment of antibodies not biased by any effects of immunisation on an antigen. Presentation of the antigen is enabled by panning, so panned in solution, antibodies binding any other than PHF8-PHD's dominant solution form so disadvantaged that they are unlikely to appear.

A protein of PHF8-PHD's size offers relatively few epitopes to a preimmune phage library. Panned in solution, these epitopes must compete with the larger streptavidin to which they are tethered, as well as the plastic walls of the vessel that contains them. Despite, beginning from round 2, libraries were incubated with streptavidin coated beads before being exposed to PHF8-PHD, the majority of recovered antibodies were streptavidin specific. In fact, by round 3, only streptavidin-specific antibodies were recovered. So the output from round 2 was searched for PHF8-PHD specific antibodies. Of the 282 colonies rescued only three produced phage particles that bound to PHF8-PHD in a phage ELISA.

	CDR1	**CDR2**	**CDR3**
V_HH-24	RTILENYAIG	AIVPSEGT	----LPSS----RRTD
V_H-25	GFDFSTRAMS	SIYTDPNK	----------GGV-
V_H-29	GFNGSRHYMS	GISRGSGT	RRGGRVGSSYVDYDY-

Table 4.5: sequences of the V_HH and two heavy-chain's isolated from a pool of six naïve llama libraries after two rounds of panning in solution against PHF8-PHD.

Four mutations in framework two, F37V, K44G, R45L, and R47W[d], identify one of these antibodies as a V_HH, and two as V_Hs probably incorporated unintentionally during library construction.[4,5] The mutation of hydrophobic to charged residues, and an aromatic residues to a smaller hydrophic residue is thought afford V_HHs stability without a light chain,[4,5] though the V_Hs isolated here appear stable alone.

Solution Binding

This V_HH and two V_Hs, subcloned to remove their phage coat protein fusion partners, expressed in HEK293 cells, and purified, when combined with PHF8-PHD and analysed by SEC show strong evidence they form complexes. A peak appears, preceding antibody and antigen, and the antigen peak disappears, indicating that these antibodies from preimmune libraries, unlike those from the immunised rabbits, can recognise PHF8-PHD's dominant solution state.

[d] Residued numbered according to the *Chothia* definition

		CDR1		CDR2		CDR3	
V_HH-24	MAQVQLVESGGGLVQAGGSLRLTCTGS	RTILENYAIG	W f RQAPGK kr E r VT	AIVPSEGT	TWYAPSVKGRFTIFADNAKNTVYLQMNSLKPEDTAIYYCNI	----LPSS----RRTD	WGQGTQVTVSSAAAHHHHHH
V_H-25	MAEVQLVESGGGLVQPGGSLRLSCAAS	GFDFSTRAMS	W v RQAPGK gl E w VS	SIYTDPNK	TAYADSVKGRFTISRDNAEQTLYLQMNSLKPEDTAIYYCAL	----------GGV-	RGQGTQVTVSSAAAHHHHHH
V_H-29	MAQVQLVESGGGLVPPGGSLRLSCAAS	GFNGSRHYMS	W v RQAPGK gl E w VS	GISRGSGT	LFYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYCAA	RRGGRVGSSYVDYDY-	WGQGTLVTVSSAAAHHHHHH

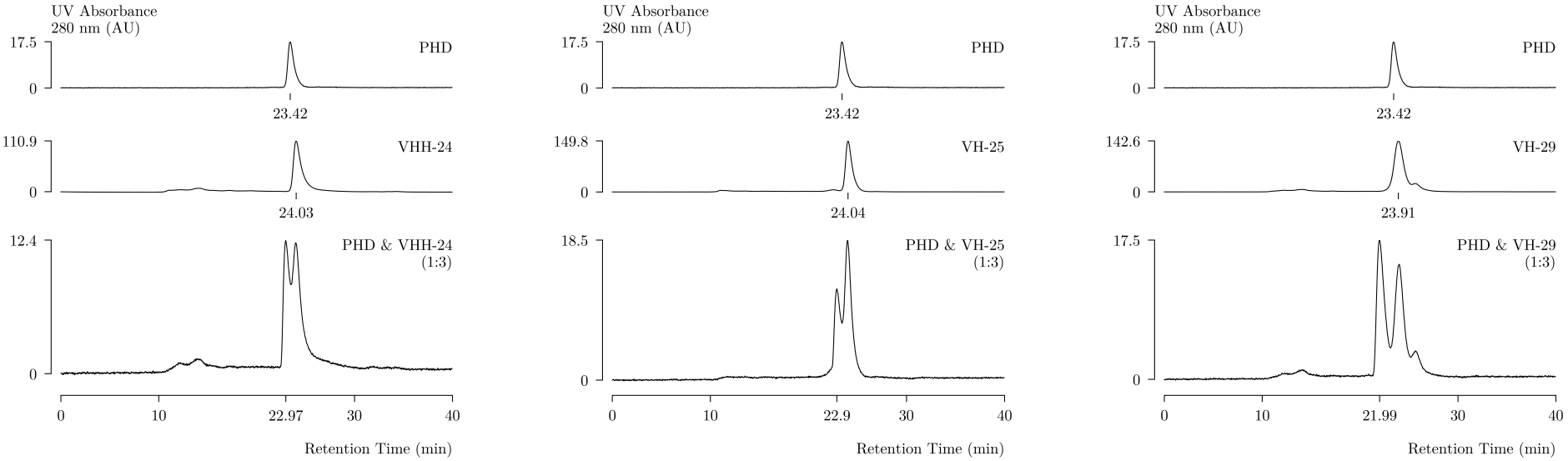


Figure 4.6: (Top) Sequences of V_HH-24, V_H-25, and V_H-29 showing residues that often differentiate V_HH and VH in lower case. (Bottom) size exclusion chromatograms for PHF8-PHD at a concentration of 10 µM *incubated with V_HH-24, V_H-25, and V_H-29 at an approximately three-fold molar excess. Each chromatogram displays the shift in mass, consistent with complex formation in solution and disappearance of the minor species (PHF8-PHD).*

Although V_HH-24, V_H-25, and V_H-29 all appear to form complexes with PHF8's PHD, they struggle to bind to the catalytically active construct, $PHF8_{1–447}$.

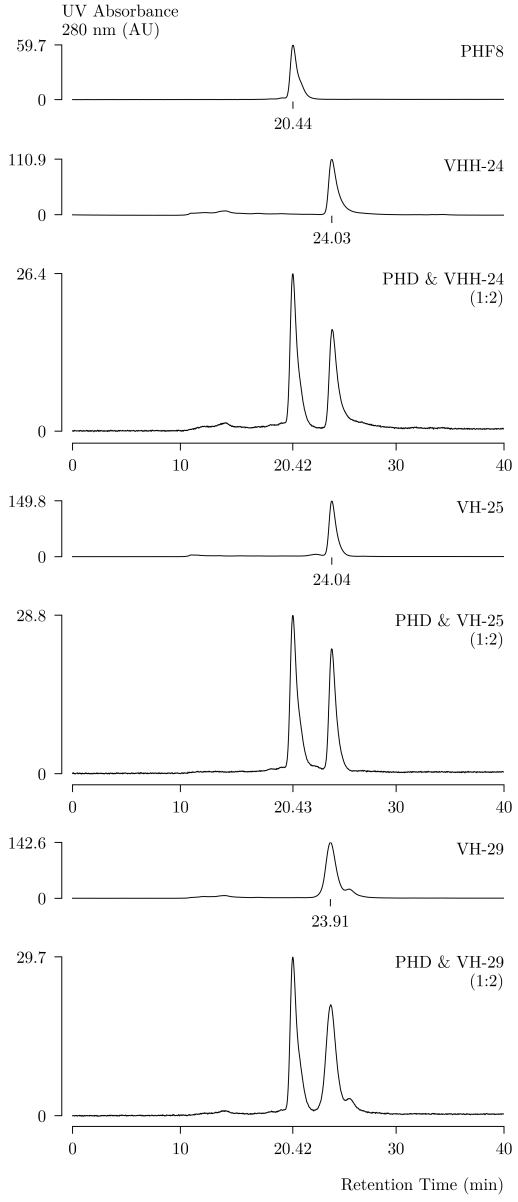


Figure 4.7: UV absorbance at 280 nm *from SEC of (from top) V_HH-24, V_H-27, V_H-29 with $PHF8_{1–447}$. Despite all successfully binding, even in solution to the PHD, they appear to not bind $PHF8_{1–447}$ – the catalytically active construct including both the PHD and JmjC domain.*

These antibodies are not intended to be inhibitors of $PHF8_{1–447}$ themselves, but rather guides in the design of smaller synthetic molecules. Their inability to bind PHF8-PHD when accompanied by the catalytic JmjC domain is not necessarily detrimental. The JmjC domain more likely prevents anti-

body recognition by steric hinderance than it does by disallowing certain conformations accessible in its absence. The two domains are separated by a linker region that is likely unstructured, and although the domains probably move relative to each other, there is no evidence to suggest coordination of conformational changes between the two. Though parts of PHF8's PHD may be inaccessible to antibodies whose molecular weights are over twice its own, the substrate peptide occupies an interface between the JmjC domain and PHD that is not solvent exposed, implying a similar accessiblity to antibody mimetics. In fact, separating JmjC domain and PHD may have exposed epitopes difficult to explore otherwise.

Mixing V_HH-24, V_H-25, and V_H-28 in various combinations with PHF8-PHD showed no strong evidence of ternary or quaternary complex formation in a SEC assay. There is a small shift in retention time when V_HH-24 and V_H-25 are combined, together, with PHF8-PHD, but it could also be caused by variation between SEC runs. The retention time of a putative V_H-29 complex, 21.99 min, differentiates it from V_HH-24 and V_H-25 whose retention times are, respectively, 22.97 min and 22.99 min. In combination with either V_HH-24 or V_H-25, the retention time is that of the V_H-29-PHF8-PHD complex, implying V_H-29 appears to dominate complex formation, though whether this is due to affinity or conformational change is uncertain.

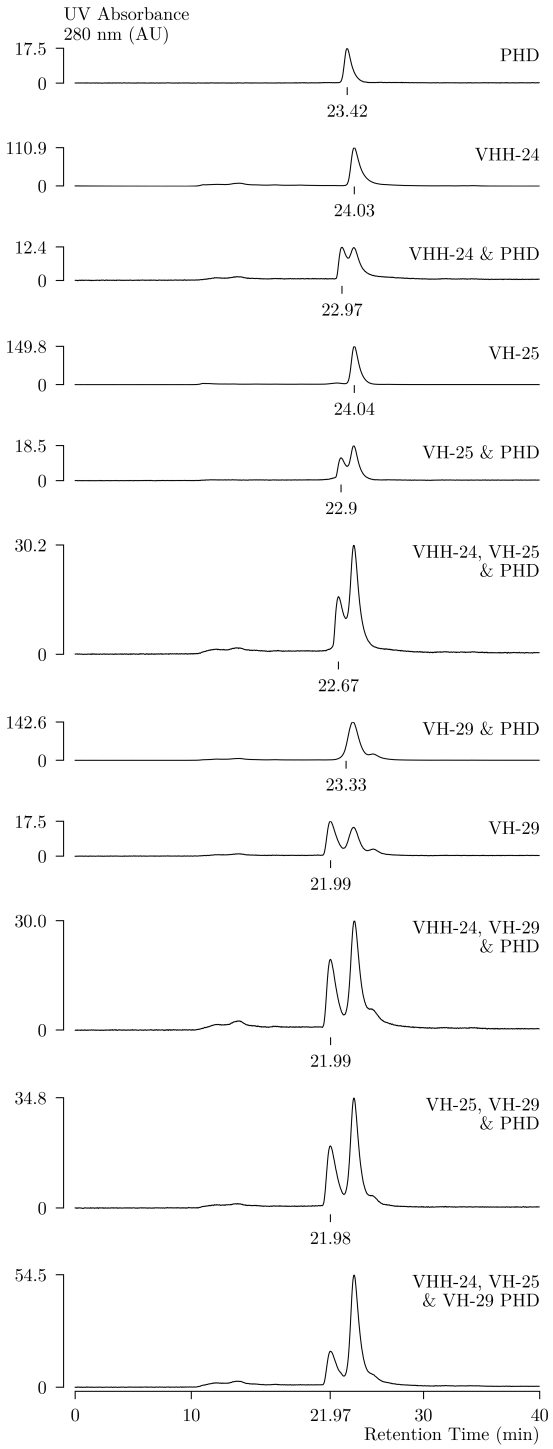


Figure 4.8: SEC chromatograms of V_HH-24, V_H-25, and V_H-29 in various combinations with PHF8-PHD obtained using a S200 column. There is no strong evidence of complexes involving multiple antibodies forming, though there is a slight shift in retention time when V_HH-24 and V_H-25 are simultaneously combined with PHF8-PHD.

Inhibition of the PHD-H3K4me3 Interaction

The PHD of PHF8 binds to H3K4me3 so that H3K9me2/1 can be positioned towards, and demethylated by, PHF8's JmjC domain. Although the JmjC domain alone is unable to demethylate H3K9me2/1, the PHD alone can bind to H3K4me3. For a C-terminal biotinylated H3K9me3$_{1–14}$ peptide captured on a streptavidin-coated SPR sensor surface, the PHF8-PHD-H3K4me3 interaction's K_{D} is 4.12 µM ⁜ 1.03.

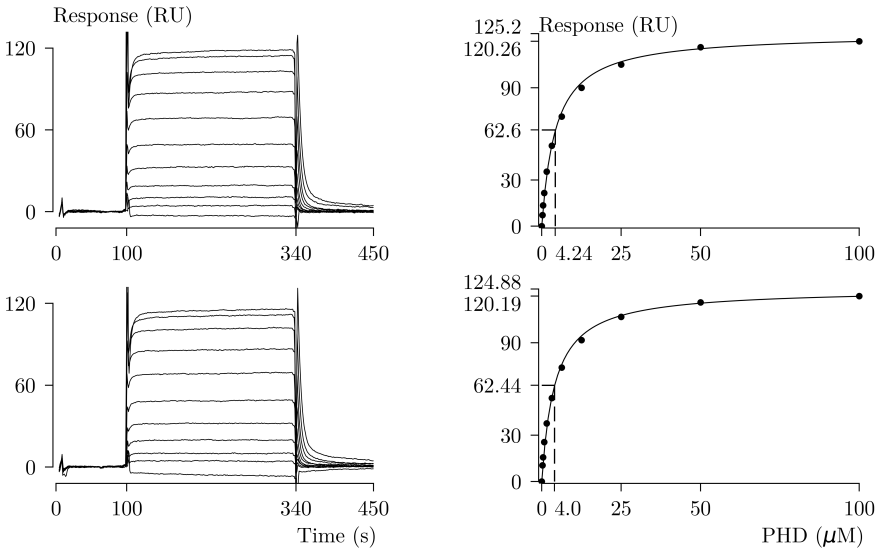


Figure 4.9: The K_D of PHF8-PHD for H3K4me3$_{1–14}$ captured on a streptavin coated SPR sensor surface was 4.12 µM ⁜ 1.03.

The on- and off-rates of PHF8-PHD binding to this H3K4me3 peptide substrate are too fast to measure accurately by SPR, but the signal level, by being roughly proportional to molecular weight, gives a good indication of the extent of binding. Incubating an antibody with PHF8, and injecting the mixture over the immobilised H3K4me3 peptide reflects interference with substrate recognition in a decreased binding signal. For instance, V_{H}-29's prevention of PHF8-PHD-H3K4me3 is evidenced by an almost total loss of response signal.

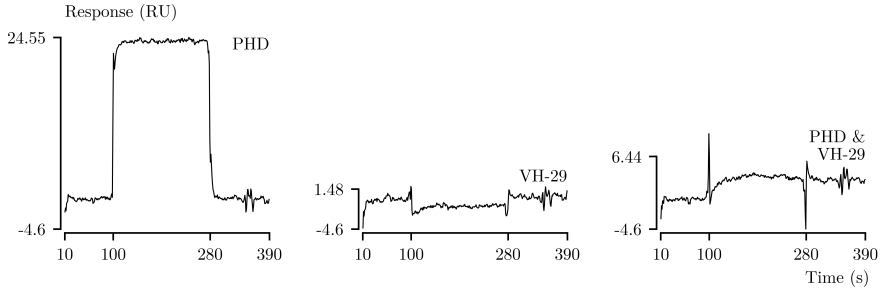


Figure 4.10: (From left) SPR traces of PHF8-PHD, V_H-29, and V_H-29 and PHF8-PHD all at a concentration of 1 µM*. Incubating PHF8-PHD with V_H-29 essentially completely reduces the signal associated with PHF8-PHD binding to the surface-immobilised H3K4me3 peptide.*

In contrast to V_H-29, incubating V_H-25 with PHF8-PHD equimolar at a concentration of 1 µM didn't appear to prevent PHF8-PHD binding to H3K4me3; the response actually increases.

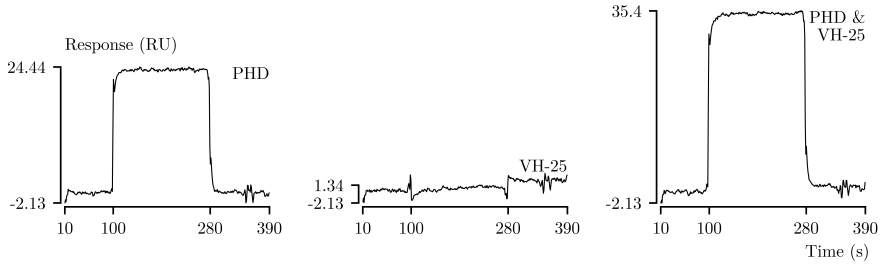


Figure 4.11: (From left) SPR traces of PHF8-PHD, V_H-25, and V_H-25 and PHF8-PHD all at a concentration of 1 µM *injected across a surface coated in streptavidin bound to biotinylated H3K4me3.*

Since there doesn't appear to be significant "background binding" of V_H-25 to the sensor surface, the increased signal of the V_H-25-PHF8-PHD mixture, relative to the PHF8-PHD-only control, is evidence the V_H-25-PHF8-PHD complex is binding to the H3K4me3-coated sensor surface. However, the signal has only increased 1.4 times relative to the PHF8-PHD-only control. If every molecule of PHF8-PHD was bound by one of V_H-25, the signal should increase roughly three times, since the molecular weight of V_H-25 is approximately twice that of PHF8-PHD. Signals any lower than this theoretical maximum indicate that at a concentration of 1 µM complex formation between V_H-25 and PHF8-PHD is still low. Increasing the concentration of V_H-25 10 µM, a ten-fold excess over that of PHF8-PHD, brought the signal

closer to its theoretical maximum and demonstrates V_H-25 binding is most probably compatible with the PHF8-PHD-H3K4me3.

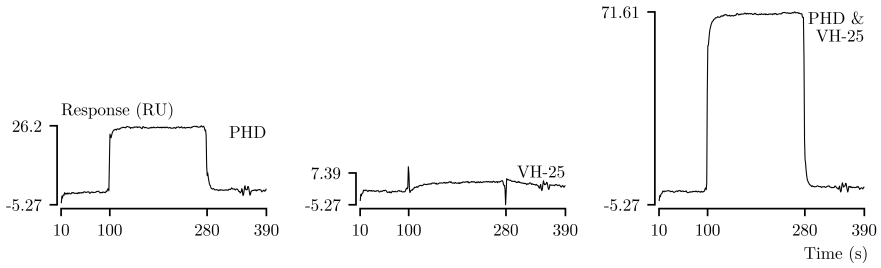


Figure 4.12: (From left) SPR traces of PHF8-PHD, V_H-25, and V_H-25 and PHF8-PHD injected across a surface coated in streptavidin bound to biotinylated H3-K4me3. The concentration of PHF8-PHD was 1 µM *and the concentration of V_H-25 was* 10 µM*.*

The effect V_HH-24 has on PHF8-PHD-H3K4me3 recognition is complicated by V_HH-24 binding to the streptavidin-H3K4me3 sensor surface. At a concentration of 1 µM, this binding is not prohibitive, but the effect on PHF8-PHD's recognition of H3K4me3 is also not conclusive.

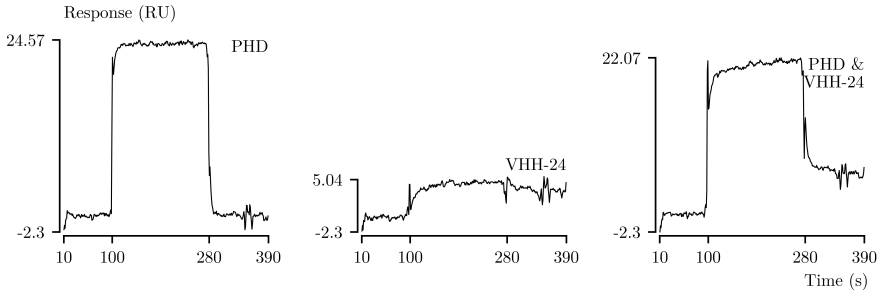


Figure 4.13: (From left) SPR traces of PHF8-PHD, V_HH-24, and V_HH-24 mixed with PHF8-PHD injected across a surface coated in streptavidin bound to biotinylated H3K4me3. The concentration of V_HH-24 and PHF8-PHD was 1 µM*.*

Despite background binding, there is a small decrease in the maximum response relative to the PHF8-PHD-only control when V_HH-24 mixed with PHF8-PHD is injected. This signal decrease is evidence V_HH-24 at least partially disrupts the interaction between PHF8-PHD and H3K4me3. It is possible that the putative disruption in PHF8-PHD binding to H3K4me3 is caused by competition with the non-specific interaction between V_HH-24 and the sensor surface. However, since the molecular weight of V_HH-24 is roughly twice that of PHF8-PHD, even if only one molecule of V_HH-24 could bind for every two molecules of PHF8-PHD, the signal level would not

change. The signal decreases, and since the sensor surface is not saturated (the response reaches 120 RU in the affinity measurement of figure 4.9) it is strong evidence that there is PHF8-PHD in the mobile phase unable to bind to H3K4me3 by already being in complex with V_HH-24. Increasing the concentration of V_HH-24 should increase this inhibition, but it is harder to see as it also increases non-specific binding between V_HH-24 and the sensor surface.[e]

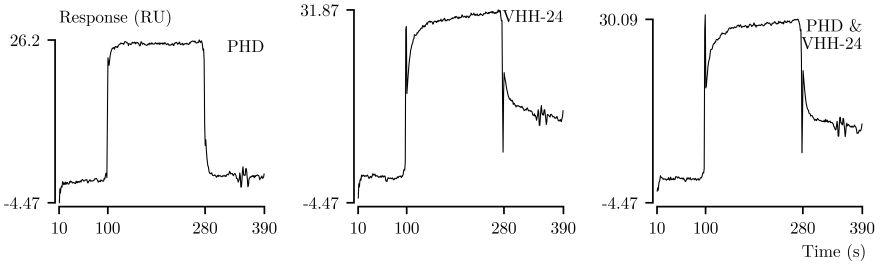


Figure 4.14: (From left) SPR traces of PHF8-PHD, V_HH-24, and V_HH-24 and PHF8-PHD injected across a surface coated in streptavidin bound to biotinylated H3K4me3. The concentration of PHF8-PHD was 1 µM *and the concentration of V_HH-24 was* 10 µM.

Since the sensor surface is not saturated, response from V_HH-24 alone can be subtracted from the response to the PHF8-PHD and V_HH-24 mixture to show, if any, interaction between PHF8-PHD and H3K4me3.

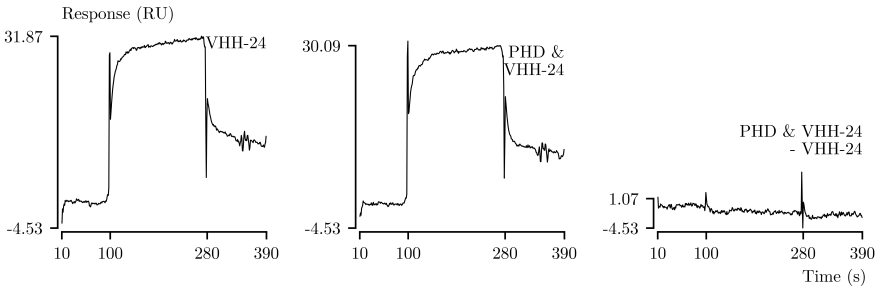


Figure 4.15: (From left) SPR traces of V_HH-24, and V_HH-24 and PHF8-PHD, and V_HH-24 and PHF8-PHD with the V_HH-24 only control subtracted. Samples were injected across a surface coated in streptavidin bound to biotinylated H3K4me3. The concentration of PHF8-PHD was 1 µM *and the concentration of V_HH-24 was* 10 µM.

Subtracting the V_HH-24 only control from the response to injecting a 10 µM to 1 µM mixture of V_HH-24 and PHF8-PHD indicates no binding. V_HH-24, therefore, almost certainly inhibits the PHF8-PHD-H3K4me3 interaction.

[e]Figure 4.15 and 4.12 were recorded during the same instrument run, and therefore share the same PHF8-PHD-only control.

Assuming this inhibition is competitive, the pronounced increase in inhibition when the antibody concentration was increased from 1 µM to 10 µM implies a K_D in the high nanomolar or low micromolar range. This is also true of V_H-25, which almost reaches the theoretical signal maximum at an antibody concentration of 10 µM. V_H-29, whose inhibition of PHF8-PHD-H3K3me3 binding is clear at an antibody concentration of 1 µM, assuming competitive inhibition, likely has a K_D in the nanomolar range. Summarised, the relationship of each antibody with the PHF8-PHD-H3K4me3 is:

V_HH-24	Binding prevents H3K4me3 recognition
V_H-25	Binding is compatible with H3K4me3 binding
V_H-29	Binding prevents by H3K4me3 recognition

Epitope Mapping by NMR

Chemical shift may be influenced both by effects that propagate through covalent bonds and effects that propagate through space. A ligand that interacts non-covalently with a protein with sufficient affinity will, therefore, effect the chemical shifts of nuclei close to the interaction site. Determining the proximity of interacting ligand atoms to interacting protein atoms using perturbed chemical shifts is non-trivial, as the total change in chemical shift is a linear combination of changes induced by multiple effects, but a threshold can, with reasonable accuracy, separate residues involved in binding from residues not involved. The molecular weight of PHF8's PHD, even in complex with a V_HH or V_H, is low enough to be accessible to NMR studies, and PHF8-PHD can be made visible to NMR by isotopic labelling. Chemical shifts perturbed by the addition of an unlabelled binding partner, when mapped on a structure of PHF8-PHD which is already available, should reveal an epitope.

Purification of ^{13}C- and ^{15}N-labelled PHF8-PHD

Since the natural abundance of NMR-active nuclei in a protein is low, PHF8-PHD (residues 1–66) was doubly labelled with ^{13}C and ^{15}N by expressing the construct in E. coli grown in labelled, minimal media.

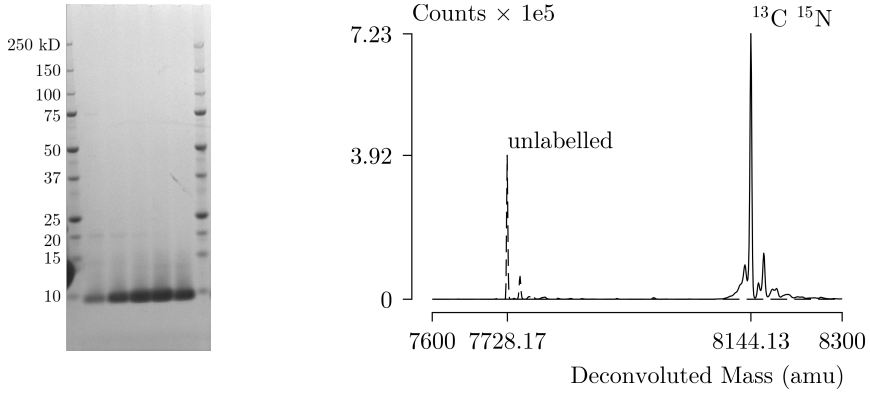


Figure 4.16: (Left) SDS-PAGE of SEC-purified ^{15}N- and ^{13}C- labelled PHF8-PHD (residues 1–66). (Right) an overlay of mass spectra for PHF8-PHD without isotopic labelling, and with double labelling.

Backbone Assignment of PHF8$_{PHD}$

PHF8-PHD's backbone resonances were assigned by triple resonance experiments. The first, CBCANH,[6] transfers magnetisation from Hα to Cα, and from Hβ to Cβ. The magnetisation evolves simultaneously on both Cα and Cβ before being transferred through Cα to the backbone amide nitrogen. Transfer to this nitrogen is best from Cα_i, that is, Cα of the same residue. But also transferred, albeit with weaker intensity, is magnetisation from Cα_{i-1}, Cα from the residue preceding the amide nitrogen. The signal now encodes, in one dimension, four resonances: Cα_i, Cβ_i, Cα_{i-1} and Cβ_{i-1}. This magnetisation evolves on the amide nitrogen before being transferred to, and detected on, the amide hydrogen. The final signal encodes in the first dimension Cα and Cβ of residue i, and with weaker intensity $i-1$, and in the remaining two dimensions backbone amide nitrogen and hydrogen resonances. Through the weaker Cα_{i-1} and Cβ_{i-1} resonance, each NH is connected to its predecessor (where the same resonances are more intense),

and also its successor (where the same carbon resonances are less intense).

CBCAHN gives a sequence of NH resonances broken only by proline. Theoretically it is enough for assignment, but the weak $C\alpha_{i-1}$ and $C\beta_{i-1}$ resonances are sometimes lost, so they are supplemented here with the CBCA(CO)NH[7] experiment. The experiment is similar to HNCACB, but magnetisation simultaneously evolved on Cα and Cβ is transferred to the backbone nitrogen through the backbone carbonyl carbon, allowing only magnetisation transferred from $C\alpha_{i-1}$ to reach the amide nitrogen, evolve, and be detected on the amide hydrogen. The resulting spectrum resembles that of the HNCACB, but each NH resonance reports only the $C\alpha_{i-1}$ and $C\beta_{i-1}$ resonances which can then be matched with the CBCANH spectrum.

Assignment usually starts from the Cβ resonances of serine and theronine which, close as they are to oxygen, are characteristically low. Continuous segments broken by proline or ambiguous resonances, are assembled, guided by the average carbon chemical shifts of each residue and the protein's amino acid sequence. The result is each NH resonance in the heteronuclear single quantum coherence (HSQC) spectrum is placed in sequence and assigned to a residue. These NH resonances, sensitive to changes is chemical environment, can thus report on binding events between an assigned isotopically-labelled protein and an unlabelled binding partner. The assignment of PHF8-PHD captured resonances from alanine 3 to arginine 66, 60/63 (95 %) of non-proline residues.

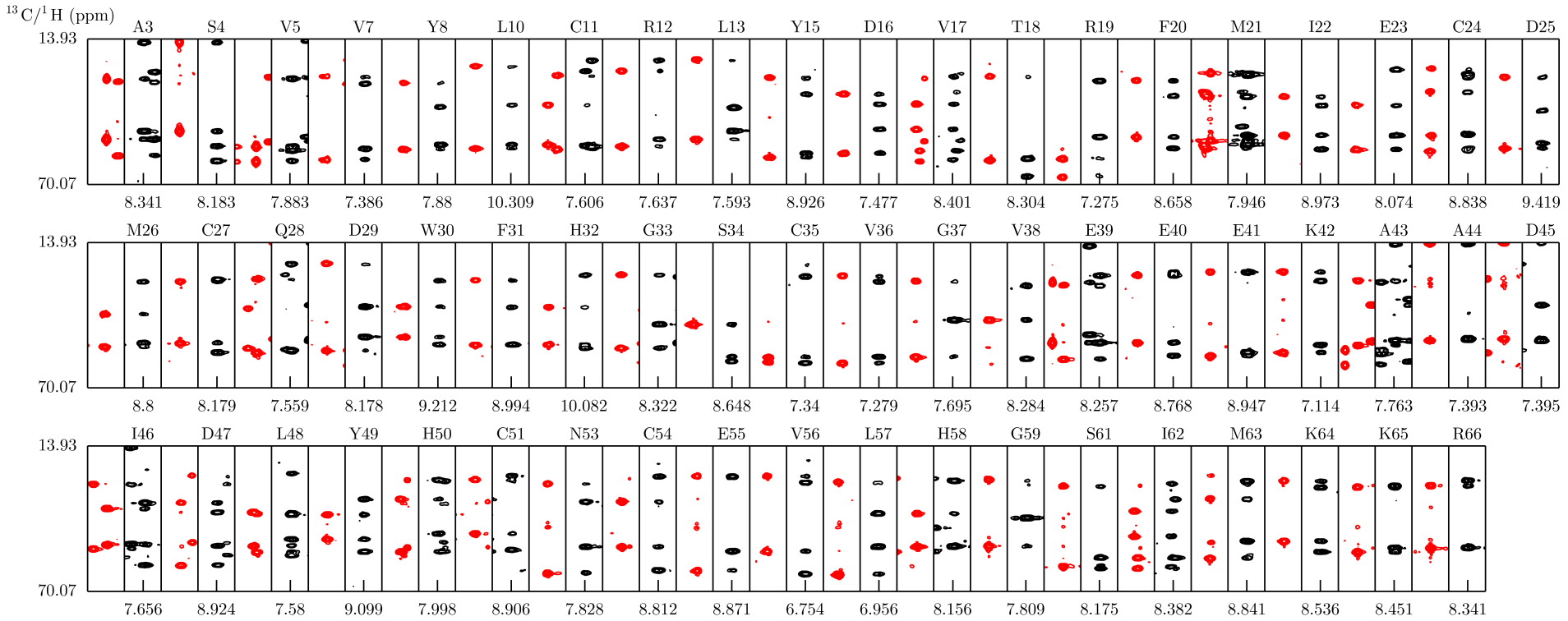


Figure 4.17: ^{13}C-^{1}H strips from the backbone assignment of ^{13}C/^{15}N-PHF8-PHD. Peaks in red were obtained from a CBCA(CO)NH spectrum and the corresponding resonance in the CBCANH spectrum is represented in black with the assigned residue indicated above. The spectra were recorded using ^{13}C/^{15}N-PHF8-PHD, at a concentration of 450 µM*, and a temperature of* 298.15 K *in sodium phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

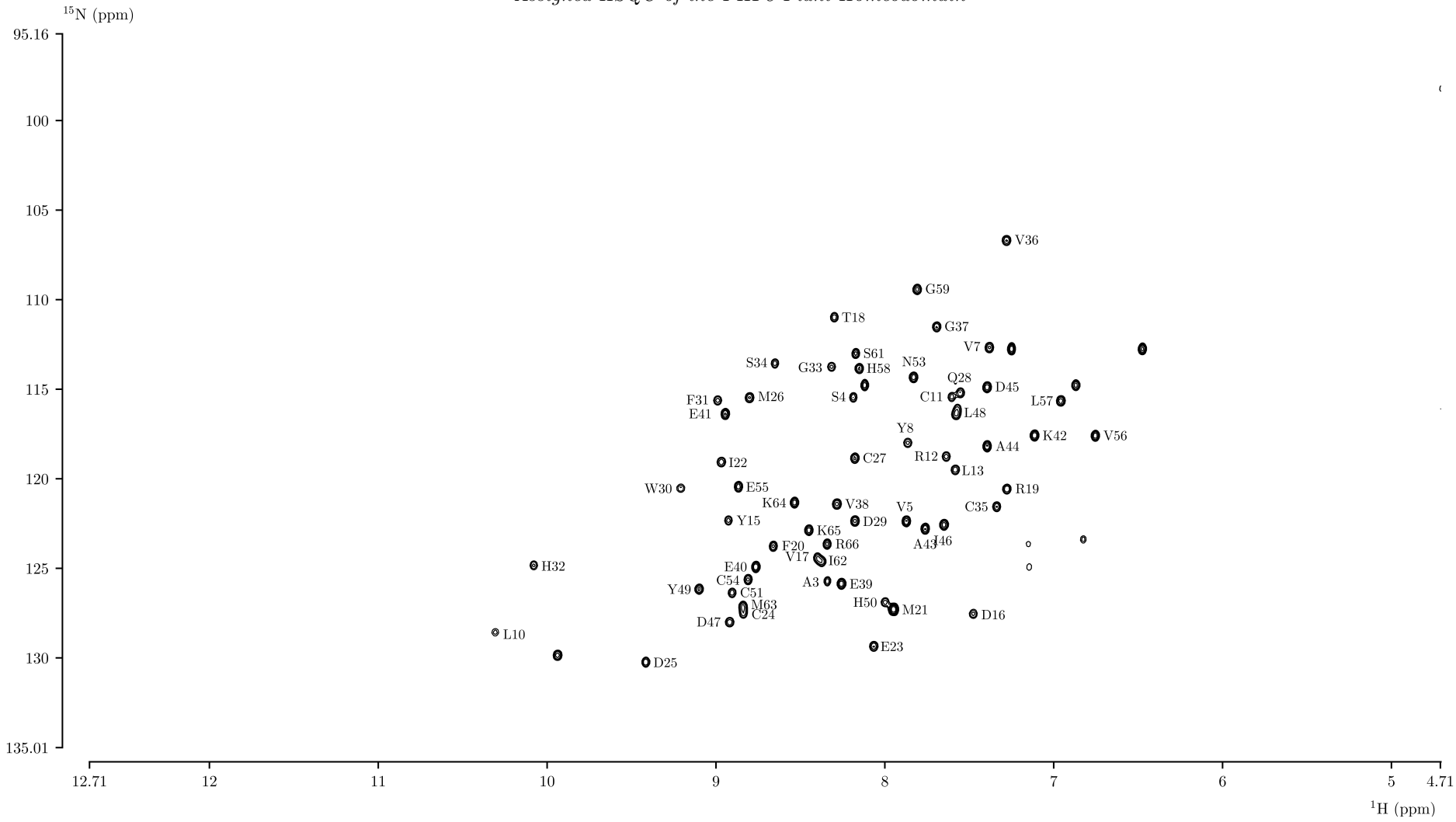


Figure 4.18: The assigned HSQC spectrum of $^{13}C/^{15}N$-PHF8-PHD, recorded at a concentration of 450 µM*, and a temperature of* 298.15 K *in sodium phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

Minimal Shift Assignment Algorithm

For many proteins whose HSQC spectra have been assigned, the perturbation to chemical shift induced by a binding partner can be estimated without reassignment. The assumption that connects assigned resonances to perturbed resonances is that the majority will have shifted the minimum amount, that is, the source of a chemical shift in a perturbed spectrum, is likely the closest chemical shift in the unperturbed spectrum.

The so-called minimal chemical shift perturbation can be calculated using a matrix that records the distance from every assigned chemical shift to every perturbed chemical shift. If A is a set of assigned resonances from a HSQC sepctum, and a_i^N and a_i^H are, respectively, the nitrogen and hydrogen chemical shifts of the i^{th} resonance, then

$$A = \left\{(a_1^N,\ a_1^H),\ (a_2^N,\ a_2^H),\ (a_3^N,\ a_3^H),\ \ldots,\ (a_n^N,\ a_3^H)\right\},$$

and if U is a list of unassigned resonances (i.e. those recorded in the presence of a ligand):

$$U = \left\{(u_1^N,\ u_1^H),\ (u_2^N,\ u_2^H),\ (u_3^N,\ u_3^H),\ \ldots,\ (u_m^N,\ u_m^H)\right\},$$

a distance matrix, D, can be constructed

$$D = \begin{array}{c} \\ (a_1^N,\ a_1^H) \\ (a_2^N,\ a_2^H) \\ (a_3^N,\ a_3^H) \\ \vdots \\ (a_m^N,\ a_m^H) \end{array} \begin{array}{c} \begin{array}{ccccc} (u_1^N,\ u_1^H) & (u_2^N,\ u_2^H) & (u_3^N,\ u_3^H) & \ldots & (u_n^N,\ u_3^H) \end{array} \\ \begin{bmatrix} D_{11} & D_{12} & D_{13} & \ldots & D_{1n} \\ D_{21} & D_{22} & D_{23} & \ldots & D_{2n} \\ D_{31} & D_{32} & D_{33} & \ldots & D_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{m1} & D_{m2} & D_{m3} & \ldots & D_{mn} \end{bmatrix} \end{array}$$

where D_{ij} is the (scaled) euclidean distance between points a_i and u_j given by:

$$D_{ij} = \sqrt{\frac{1}{2}\left((a_i^H - u_j^H)^2 + \alpha(a_i^N - u_j^N)^2\right)}. \tag{4.1}$$

The scaling factor, α, can be calculated by dividing the range of hydrogen resonances by the range of nitrogen resonances, but is usually set to 0.14 for simplicity. One approach to calculating the minimal perturbed chemical shift, δ_i, of the i^{th} assigned residue is to assign it the closest perturbed chemical shift:

$$\delta_i = \min_{j \in \{1,2,3,\ldots n\}} D_{ij} \quad \text{for } i = 1, 2, 3, \ldots, m$$

Although this approach can usually give a reasonable estimation of a residue's chemical shift perturbation, it can easily assign more than one resonance to the same perturbed resonance.

The strategy I have used here is to assign resonances in order from smallest shift to largest shift, and once a peak is assigned, to remove it from the remaining pool of possible assignments. The algorithm is closer to the way peaks would be matched by a spectroscopist (starting from the smallest, most obvious shifts), and although this approach, as most others, struggles to assign peaks that have undergone large shifts, by eliminating obvious shifts first they are less likely to be erroneously assigned. The goal in these experiments is to determine significant shifts, and although the chemical shift perturbation of a given residue may be incorrect, if the shift remains significant (above a threshold) then the conclusion is the same: the residue is near a binding interface.

The assignment begins with a distance matrix, D:

$$\begin{bmatrix} D_{11} & D_{12} & D_{13} & \ldots & D_{1n} \\ D_{21} & D_{22} & D_{23} & \ldots & D_{2n} \\ D_{31} & D_{32} & D_{33} & \ldots & D_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{m1} & D_{m2} & D_{m3} & \ldots & D_{mn} \end{bmatrix}$$

The first assignment is made by finding the minimum element in D, that is, the smallest distance between any assigned and unassigned peak:

$$\delta_1 = \min_{i \in I,\ j \in J} D_{ij}$$

$$\text{where } I = \{1, 2, 3 \ldots, m\},\ J = \{1, 2, 3, \ldots n\}$$

The next assignment is made by finding the next smallest value in the distance matrix, excluding the row and column of the assignment previously made. Assuming, for example's sake, that the minimum element, δ_1 is D_{23}, then row 2 and column 3 will be excluded.

$$\delta_2 = \min_{i \in I,\ j \in J} D_{ij}$$

$$\text{where } I = \{1, 2, 3, \ldots, m\} \setminus \{2\},\ J = \{1, 2, 3, \ldots n\} \setminus \{3\}$$

The operation is equivalent to finding the minimum value of the distance matrix without row 2, and column 3:

$$\begin{bmatrix} D_{11} & D_{12} & \mathbf{D_{13}} & \ldots & D_{1n} \\ \mathbf{D_{21}} & \mathbf{D_{22}} & \mathbf{D_{23}} & \ldots & \mathbf{D_{2n}} \\ D_{31} & D_{32} & \mathbf{D_{33}} & \ldots & D_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ D_{m1} & D_{m2} & \mathbf{D_{m3}} & \ldots & D_{mn} \end{bmatrix} \longrightarrow \begin{bmatrix} D_{11} & D_{12} & \ldots & D_{1n} \\ D_{31} & D_{32} & \ldots & D_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ D_{m1} & D_{m2} & \ldots & D_{mn} \end{bmatrix}$$

Assuming again, for example, that the result of this operation, δ_2, is D_{12}, then row 1 and column 2 are also excluded when making the next assignment:

$$\delta_3 = \min_{i \in I\, j \in J} D_{ij}$$

$$\text{where } I = \{1, 2, 3, \ldots, m\} \setminus \{2, 1\},\ J = \{1, 2, 3, \ldots n\} \setminus \{3, 2\}$$

Which is, again, equivalent to finding the minimum element in the distance matrix with row 2 and 1, and columns 3 and 2 removed:

$$\begin{bmatrix} \mathbf{D_{11}} & \mathbf{D_{12}} & \dots & \mathbf{D_{1n}} \\ D_{31} & \mathbf{D_{32}} & \dots & D_{3n} \\ \vdots & \vdots & \vdots & \vdots \\ D_{m1} & \mathbf{D_{m2}} & \dots & D_{mn} \end{bmatrix} \longrightarrow \begin{bmatrix} D_{31} & \dots & D_{3n} \\ \vdots & \vdots & \vdots \\ D_{m1} & \dots & D_{mn} \end{bmatrix}$$

These steps are continued until there are no more rows, or no more columns in the matrix D to exclude, that is, there are no more assigned resonances left, or no more unassigned resonances left.

H3K4me3 Epitope

Since there was already a crystal structure of PHF8$_{1–447}$ in complex with a H3K4me3K9me2 peptide,[1] comparing this structure to the epitope mapped by incubating PHF8-PHD and H3K4me3$_{1–14}$ is a natural way to validate epitope mapping on PHF8-PHD by NMR. Setting the threshold that separates residues involved in binding, from those that are not, is necessarily arbitrary when only chemical shift changes are available, so there are several approaches for analysis.[8] Here, the standard deviation of all chemical shift perturbations was calculated, and residues whose shifts are above different multiples of this standard deviation indicated. I feel this provides the most unbiased impression of these semi-quantitative epitope maps since, as discussed later, the assumption of a one-to-one correspondence assigned and perturbed resonances may not hold for PHF8-PHD with its various binding partners, and decision of a hard threshold may obscure the underlying epitope. Incubating PHF8-PHD and an H3K4me3$_{1–14}$ at concentration of 50 µM and 60 µM, respectively, produced the following spectrum.

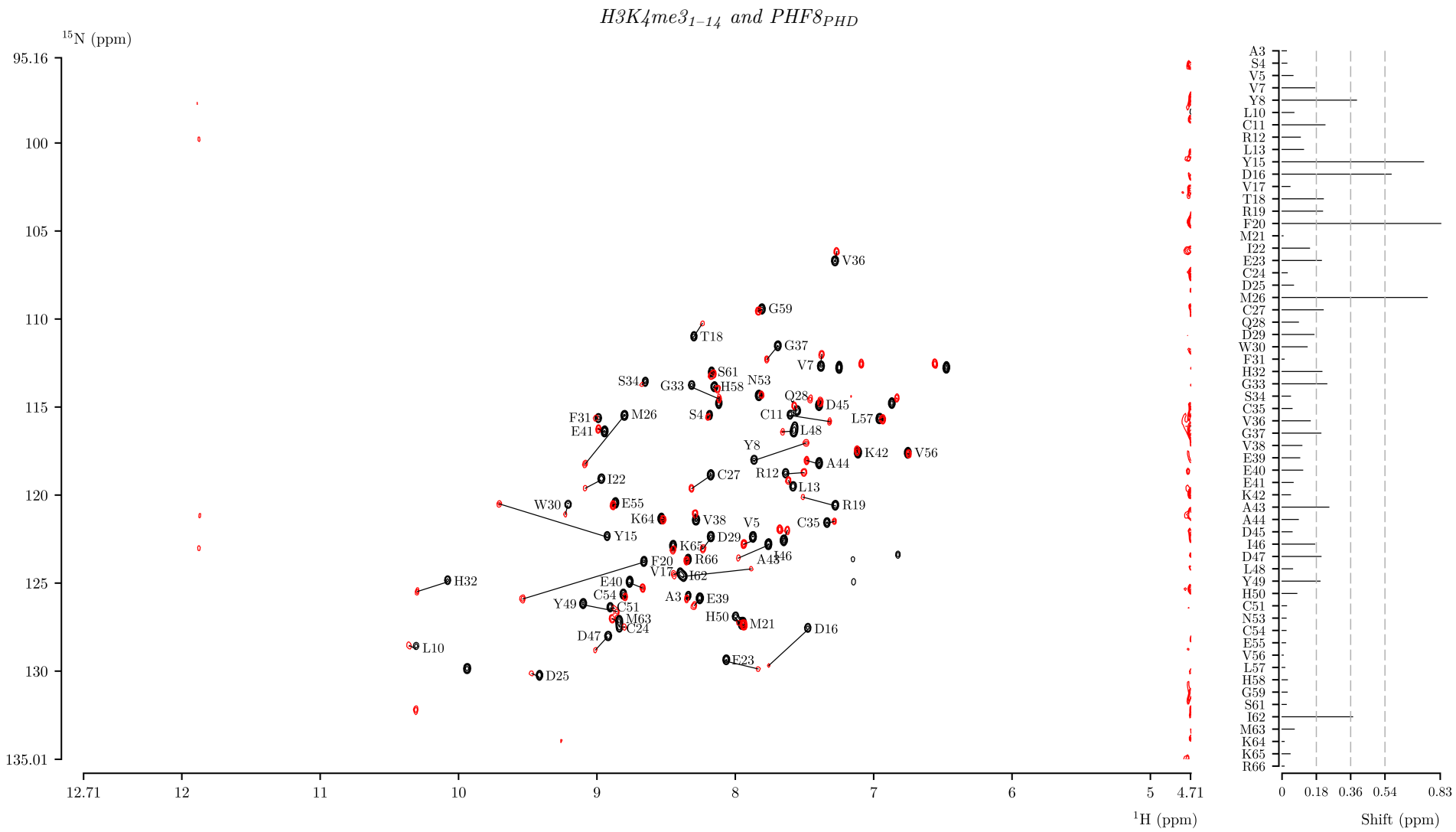


Figure 4.19: An overlay of HSQC spectra for PHF8-PHD at a concentration of 400 µM *and PHF8-PHD (*50 µM*) incubated with a H3K4me3 14mer (*60 µM*). Peaks in the assigned spectrum (black) are connected to their assigned peaks in the perturbed spectrum (red) by line segments. Chemical shift perturbation calculated according to equation 4.1. Both spectra were recorded at* 298.15 K *in a phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

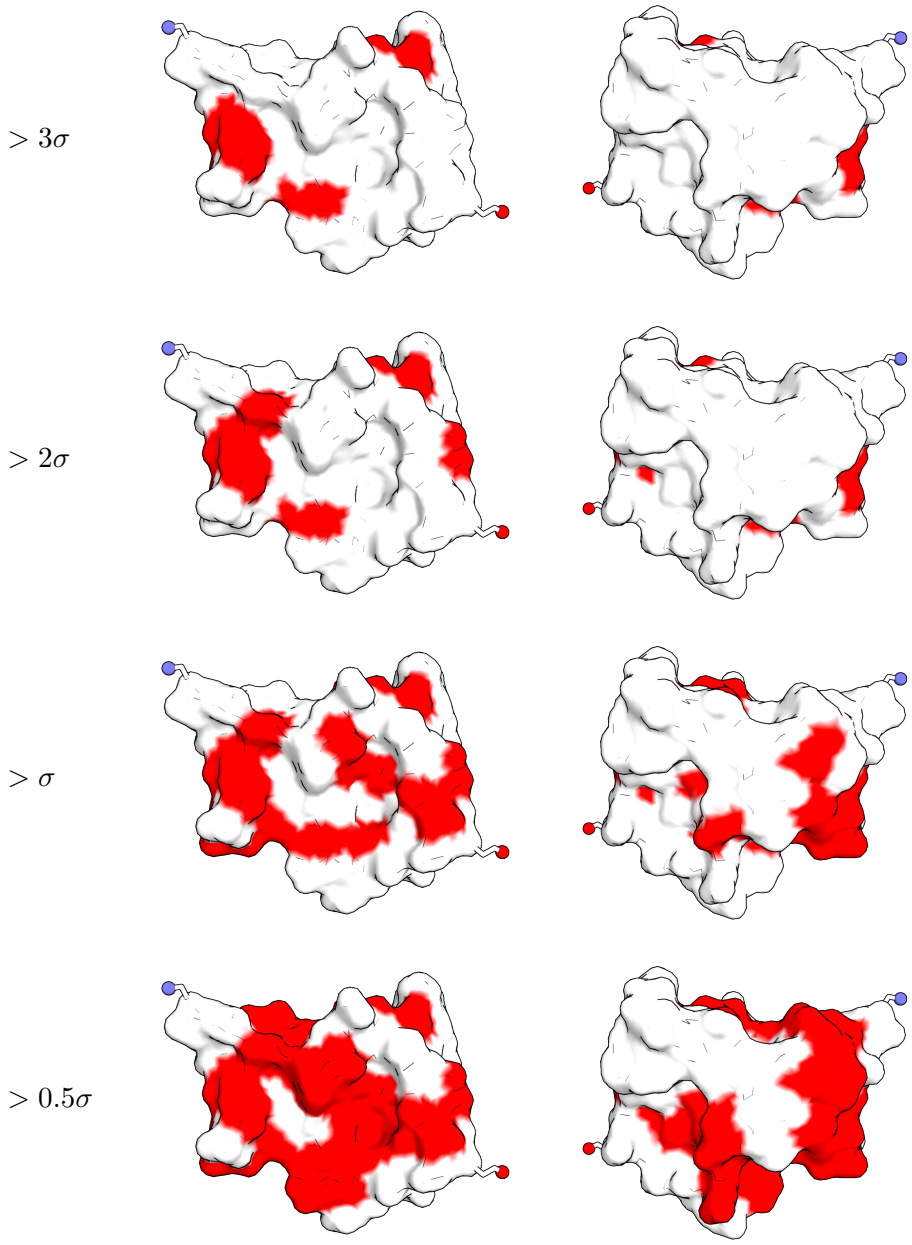


Figure 4.20: Residues whose chemical shifts are perturbed by more than (from top) 3σ, 2σ, σ, and 0.5σ by the addition of $H3K4me3_{1-14}$, where σ is the standard deviation calculated from all perturbed chemical shifts for all residues. The N-terminal amine nitrogen and C-terminal carbonyl oxygen are shown as spheres in blue and red, respectively.

As the threshold separating interacting from non-interacting residues approaches σ, an epitope appears in the shadow of the H3K4me3 substrate peptide. The putative epitope captures, on the left, the K4me3 binding pocket that gives the PHD its specificity, and on the right the arginine binding pocket – two features common to PHDs specific for H3K4.

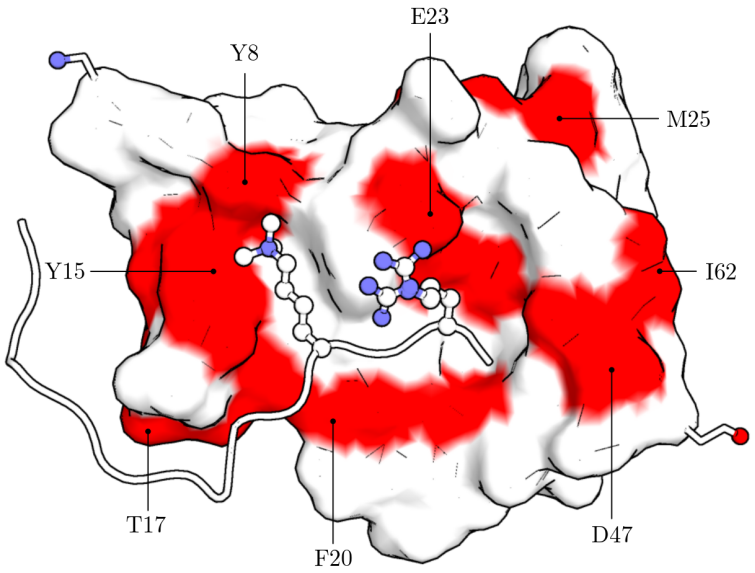


Figure 4.21: The PHD, residues 1–70, and the substrate peptide, residues 1–14, rendered from a crystal structure of PHD and JmjC domain bound to a H3K4me3-K9me2 peptide (PDB: 3KV4).[1] Residues coloured red have chemical shift changes greater than the standard deviation of all chemical shift changes of all residues. The side chain of arginine 2 is shown in two conformations to reflect occupancy of multiple states in the electron density. Selected ersidues on the likely interaction surface are labelled.

Epitopes of V_HH-24, V_H-25, and V_H-29

Since V_HH-24, V_H-25, or V_H-29 with PHF8-PHD all appeared to form complexes in solution each antibody was combined with ^{15}N-labelled PHF8-PHD and NMR spectra recorded. From NMR-derived epitope maps the putative epitopes for V_HH-24 and V_H-29 cross the H3K4me3 binding site, and so support binding preventing substrate recognition, whereas the putative epitope for V_H-25 is adjacent to the substrate binding site, and thus it is compatible with substrate recognition – as observed in the previous section.

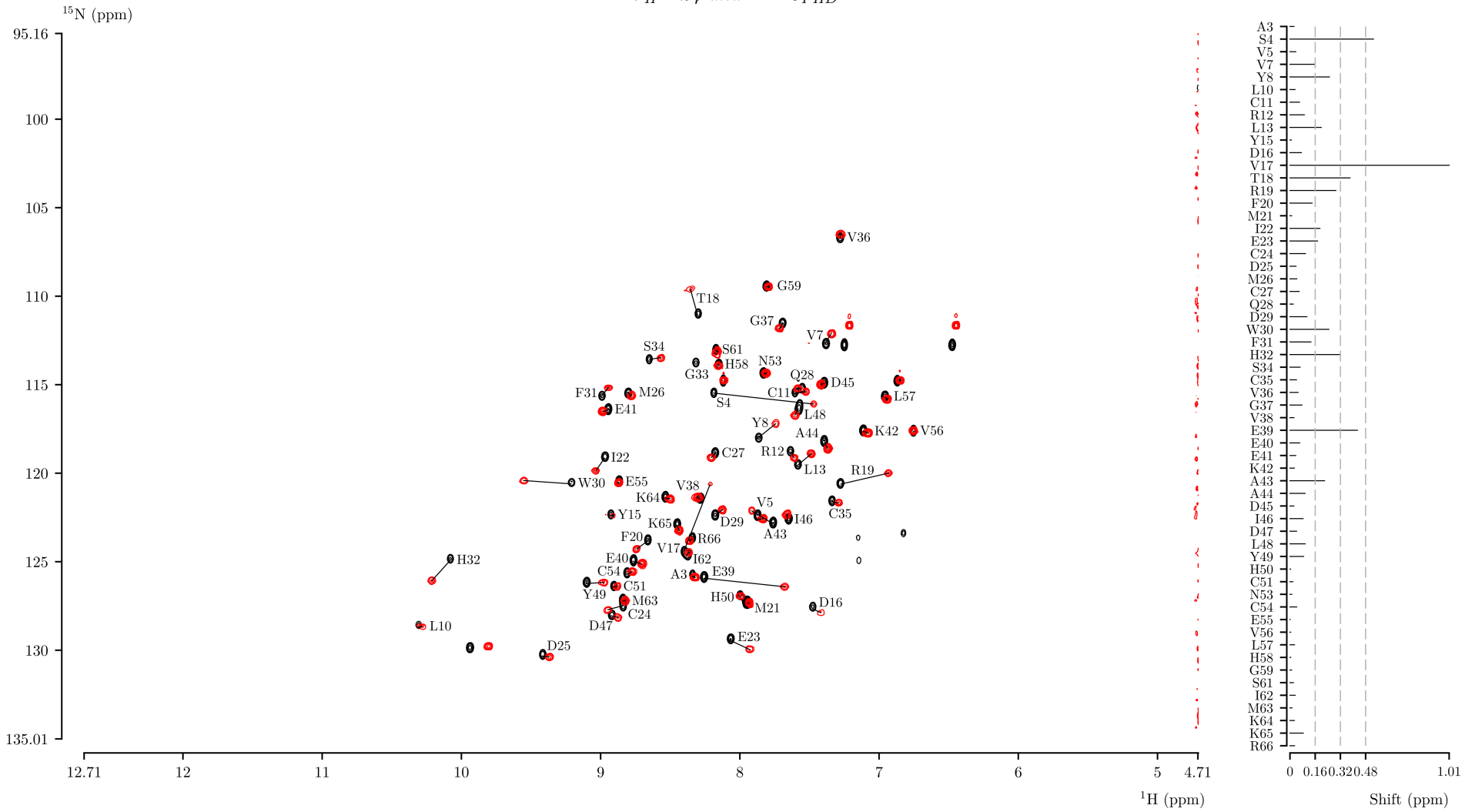


Figure 4.22: An overlay of HSQC spectra for PHF8-PHD at a concentration of 400 µM *and PHF8-PHD* (50 µM) *incubated with V_HH-29* (60 µM)*. Peaks in the assigned, apo, spectrum (black) are connected to their assigned peaks in the perturbed spectrum (red) by line segments. Chemical shift perturbation was calculated according to equation 4.1. Both spectra were recorded at* 298.15 K *in a phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

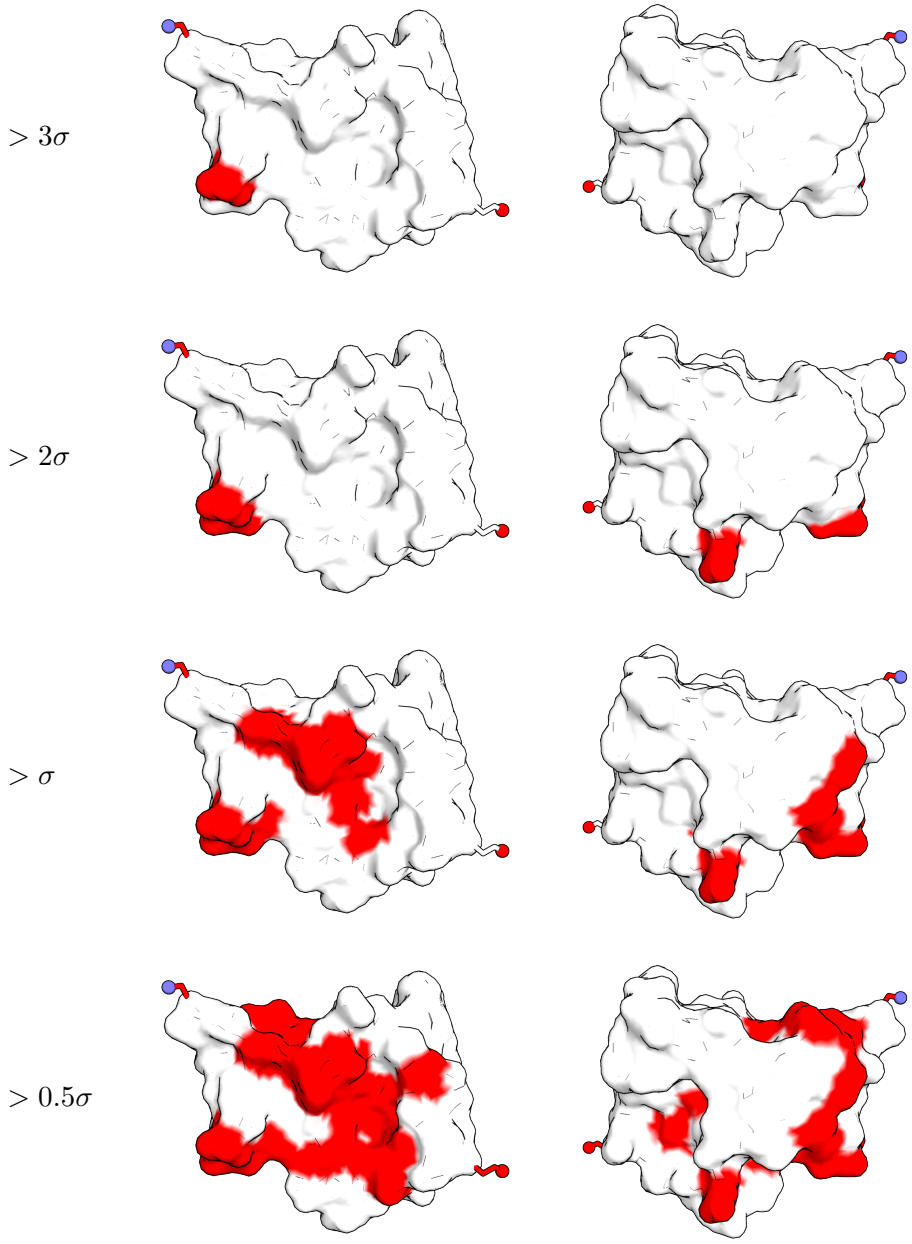


Figure 4.23: Residues whose chemical shifts are perturbed by more than (from top) 3σ, 2σ, σ, and 0.5σ by the addition of V_HH-24, where σ is the standard deviation calculated from all chemical shifts perturbations.

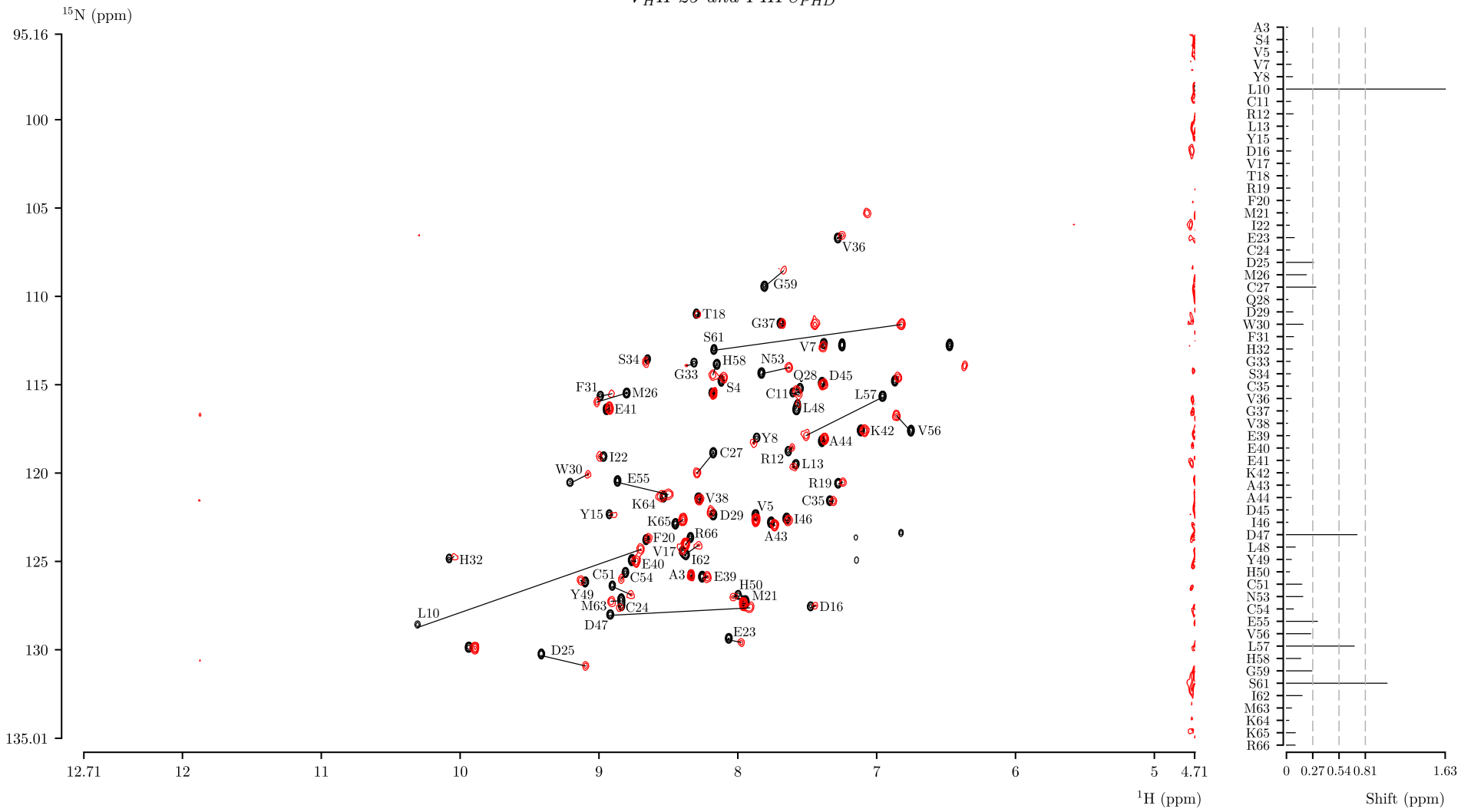


Figure 4.24: An overlay of HSQC spectra for PHF8-PHD at a concentration of 400 µM *and PHF8-PHD (*50 µM*) incubated with V_HH-25 (*60 µM*). Peaks in the assigned spectrum (black) are connected to their assigned peaks in the perturbed spectrum (red) by line segments. Chemical shift perturbation was calculated according to equation 4.1. Both spectra were recorded at* 298.15 K *in a phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

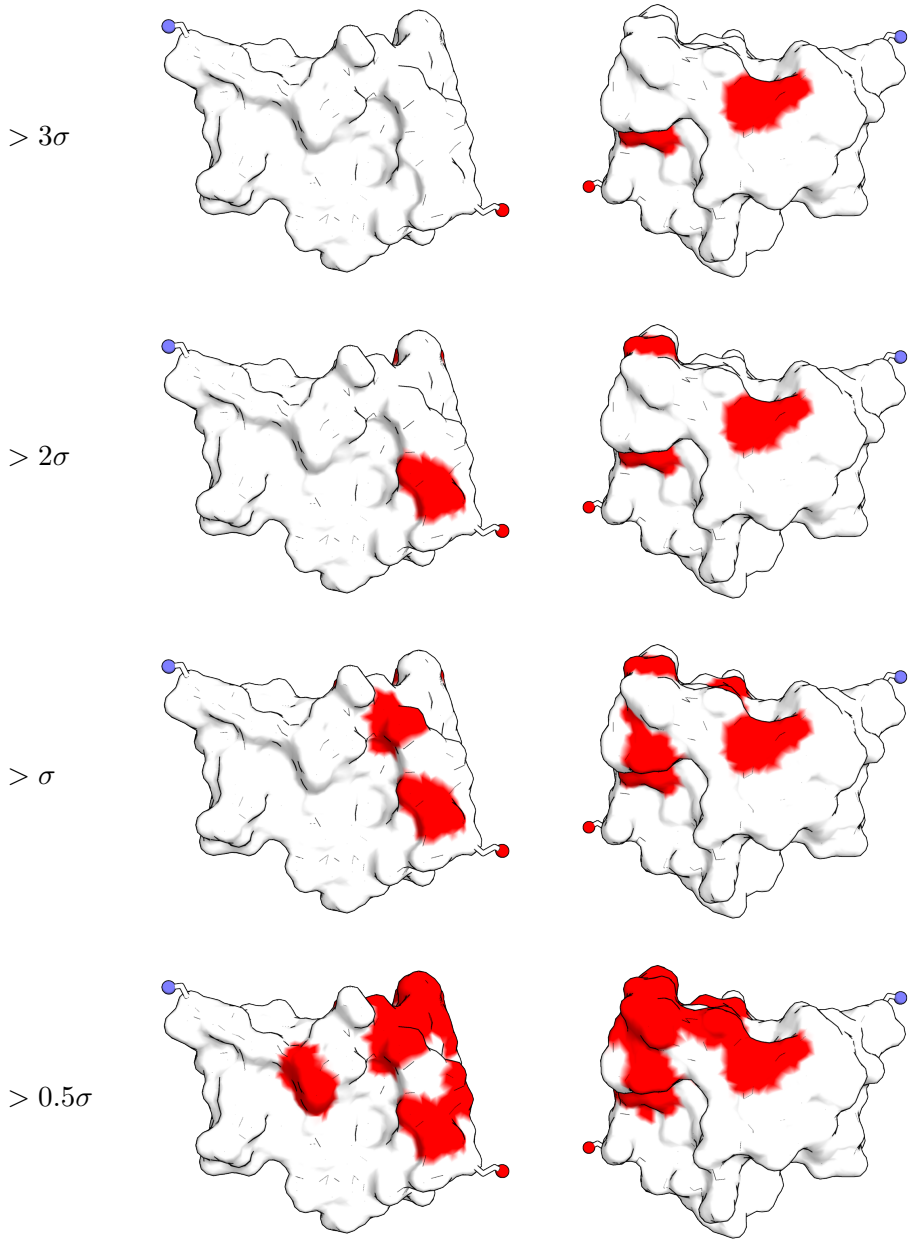


Figure 4.25: Residues whose chemical shifts are perturbed by more than (from top) 3σ, 2σ, σ, and 0.5σ by the addition of V_H-25, where σ is the standard deviation calculated from all perturbed chemical shifts.

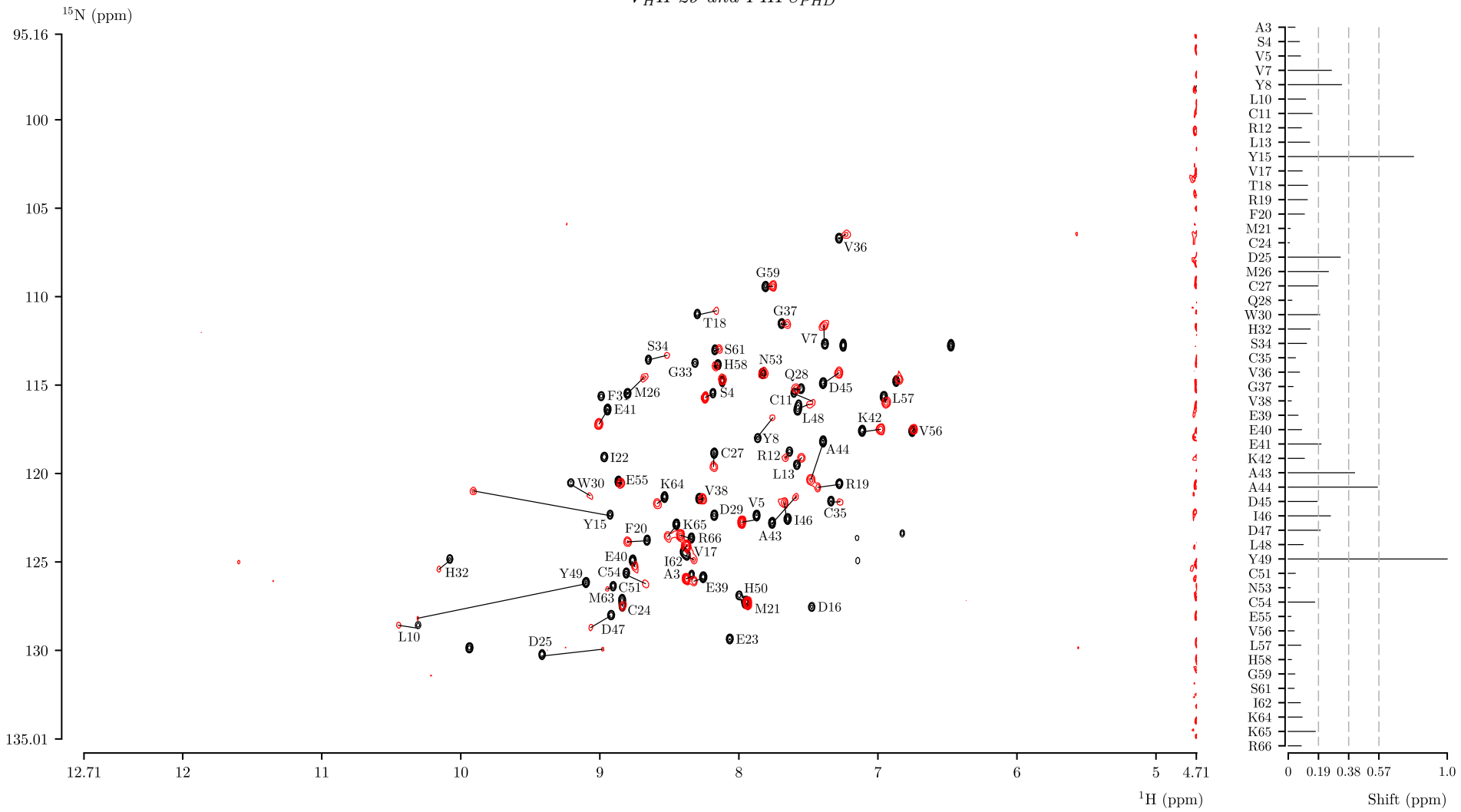


Figure 4.26: An overlay of HSQC spectra for PHF8-PHD at a concentration of 400 µM *and PHF8-PHD (*50 µM*) incubated with V_H-29 (*60 µM*). Peaks in the assigned spectrum (black) are connected to their assigned peaks in the perturbed spectrum (red) by line segments. Chemical shift perturbation was calculated according to equation 4.1. Both spectra were recorded at* 298.15 K *in a phosphate buffer, pH* 6.3*, with* 50 mM *NaCl.*

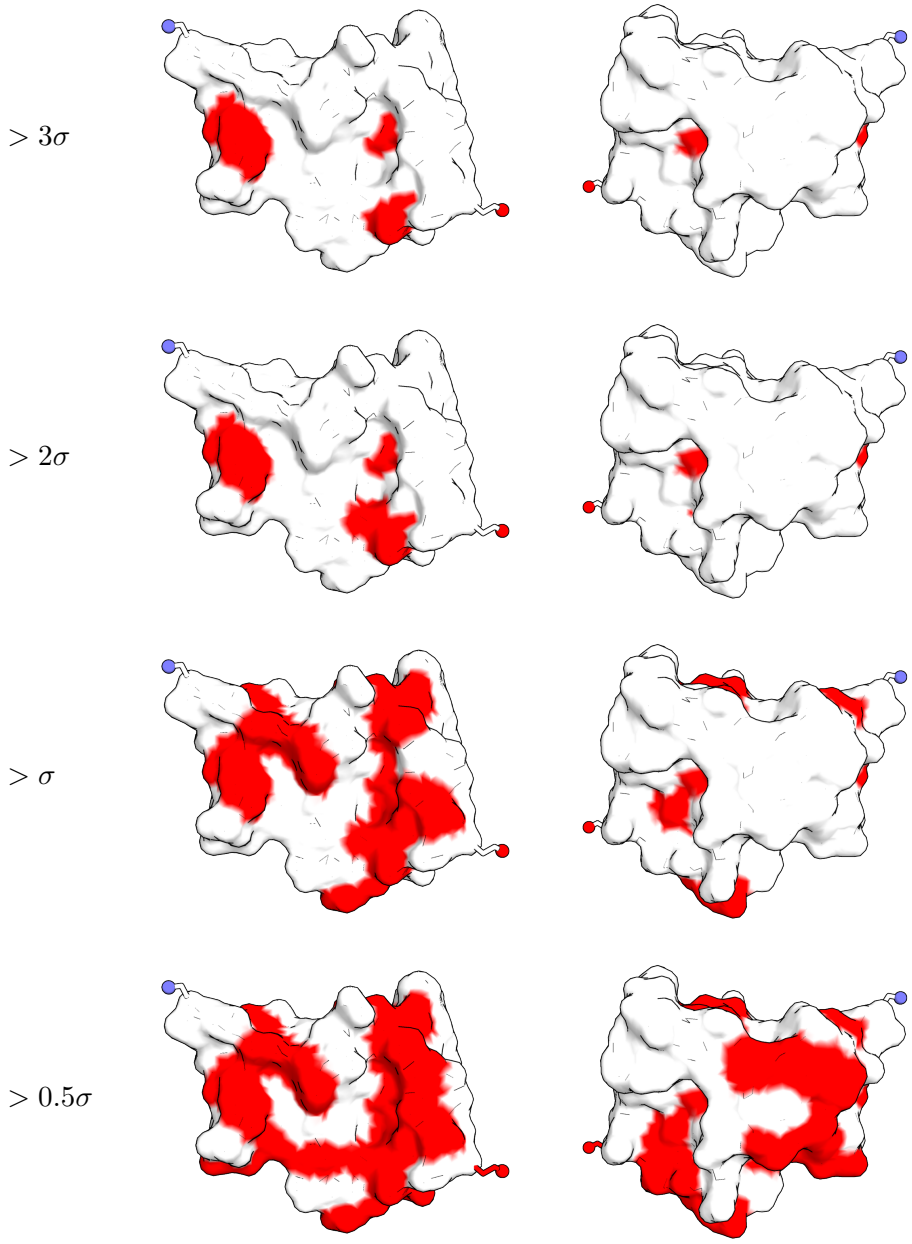
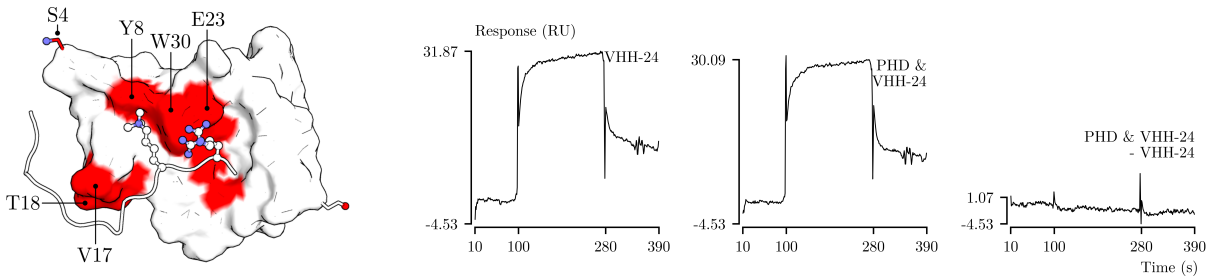
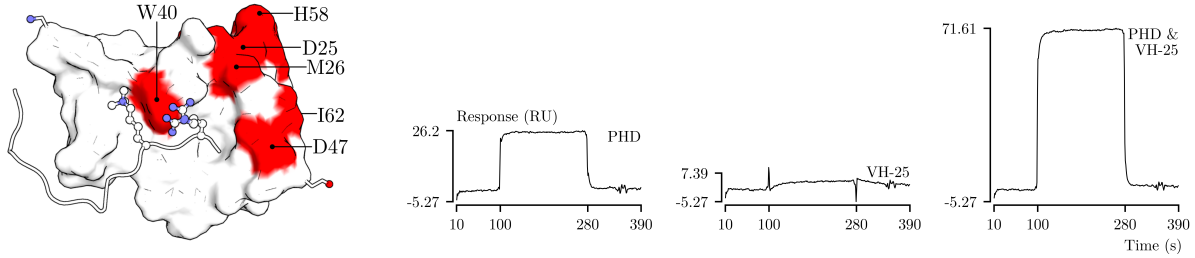


Figure 4.27: Residues whose chemical shifts are perturbed by more than (from top) 3σ, 2σ, σ, and 0.5σ by the addition of V_H-29, where σ is the standard deviation of chemical shift perturbations.

V_HH-24



V_H-25



V_H-29

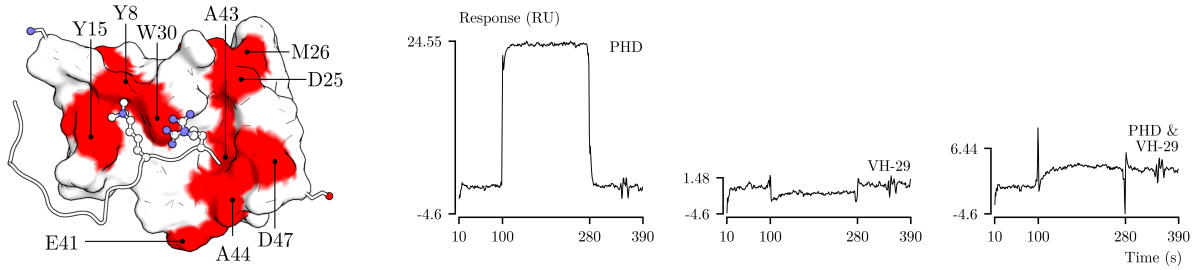


Figure 4.28: Putative epitopes for V_HH-24, V_H-25, and V_H-29 (threshold level σ) highlighted on structures of PHF8-PHD, with the substrate peptide shown for reference (structures derived from PDB: 3kv4*). Selected residues at predicted binding site are indicated. These putative epitopes help explain the SPR traces already presented which show V_HH-24 and V_H-29 prevents recognition of H3K4me3 by PHF8, and V_H-25 which is compatible with PHF8 binding to H3K4me3.*

Although these chemical shift perturbation for VHH-24 and VH-29 appear to cover similar areas of the proteins, the individual residues identified differ between two. Although the area covered by antibodies in complex is relatively large, specific interactions only appear to made by a small fraction of residues available at the surface and for two antibodies of differing sequences these are not necessary the same interactions. In addition, chemical shift perturbation, as it is used here, is not a quantitative method. The magnitude of shifts, and by extention the shifts themselves, need not represent engagement in binding, but simply a change in the chemical environment brought about by proximity to changed residues, conformational change or similar.

Besides an epitope, these perturbed spectra suggest recognition of H3-K4me3 by PHF8-PHD, and PHF8-PHD by V_HH-24, V_H-25, and V_H-29 is more than just the interaction of two static biomolecules. The spectrum of PHD with H3K4me3 contains more resonances than the spectrum of the PHD alone: an extra peak appears close to isoleucine 46 (though the true origin is uncertain without re-assignment). And in the spectrum of V_H-25 with ^{15}N-PHF8-PHD, there is similarly an additional unassigned resonance near valine 36.

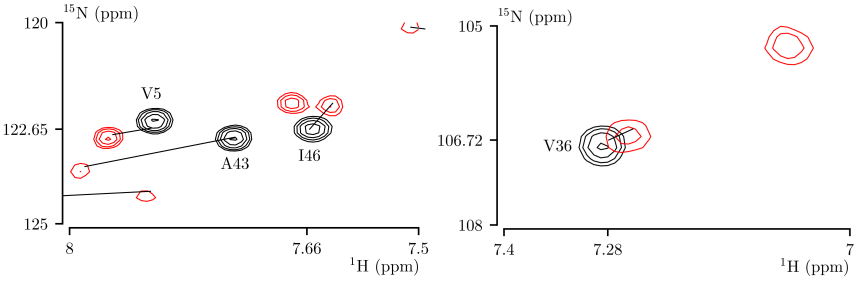


Figure 4.29: In the spectrum of ^{15}N-PHF8-PHD with H3K4me3 (left) there is an extra resonance near isoleucine 46, and in the spectrum of ^{15}N-PHF8-PHD with V_H-25 (right) there is an extra resonance near valine 36.

These spectra only represent ^{15}N-PHF8-PHD since H3K4me3 and V_H-25 are unlabelled, so more resonances than there are spin systems implies some residues experience two environments. There are many potential explanations: PHF8-PHD may adopt more than one conformation when stabilised

by a binding partner, or the partner itself could bind in multiples poses similar in energetic favourability. PHF8-PHD or its binding partner could also be moving, occupying different states often enough to appear, on average, as multiple populations. Alternatively, the few residues at the N- and C-terminus of PHF8-PHD that could not be assigned, most probably due to their being unstructured, may be stabilised enough by complexation to introduce a resonance previously unobservable. Each of these explanation is united in an implication of protein dynamics, and the theme continues elsewhere in these spectra.

In the spectrum of V_H-25 with ^{15}N-PHF8-PHD, the resonance assigned to glycine 33 almost disappears, and for V_H-29 with ^{15}N-PHF8-PHD, the intensity of the aspartate 25 resonance is also reduced. A general reduction in resonance intensity due to line broadening can be expected for an antibody binding to PHF8-PHD: a V_H, or V_HH adds twice the molecular weight of the PHD itself, slowing its tumbling and accelerating the relaxation of transverse magnitisation. But these decreases in glycine 33 and aspartate 25 are decreases relative to their companion resonances. Such a decrease can be caused by a spin system spread among different environments, whether by dynamic motion or discrete conformations, a single resolved resonance is lost to a spread in resonances.

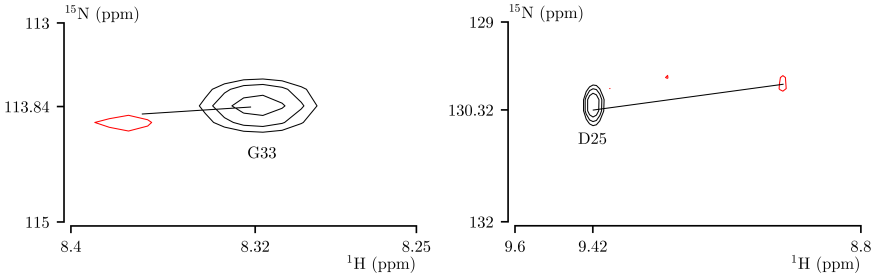


Figure 4.30: For (left) V_H-25 and (right) V_H-29 with ^{15}N-PHF8-PHD the intensity of some resonances was significantly reduced.

A complement to the reduction in intensity of aspartate 25's resonance, which is, in general, made sensitive to environmental changes by an acidic proton, are the absence of any resonances that can be assigned to isoleucine

22 and phenylalanine 31.

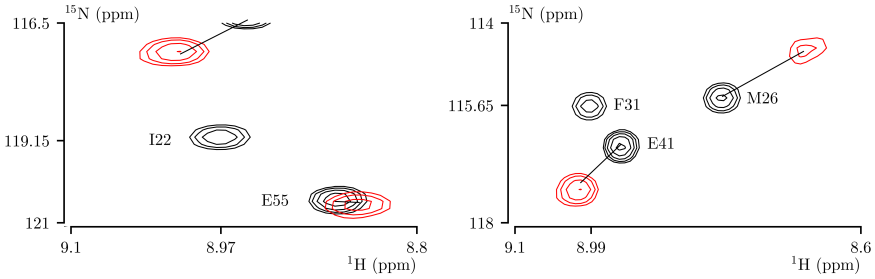


Figure 4.31: In the spectrum of V_H-29 and I22 and F31 seem to disappear – possibly a symptom of protein motion.

With protein motion, some dramatic changes in chemical shift can be reinterpreted. For PHF8-PHD combined with V_H-25, the assumed one-to-one correspondence between native and perturbed resonances has caused leucine 10 and aspartate 47 have been assigned perturbed chemical shifts of potentially unphysical magnitude.

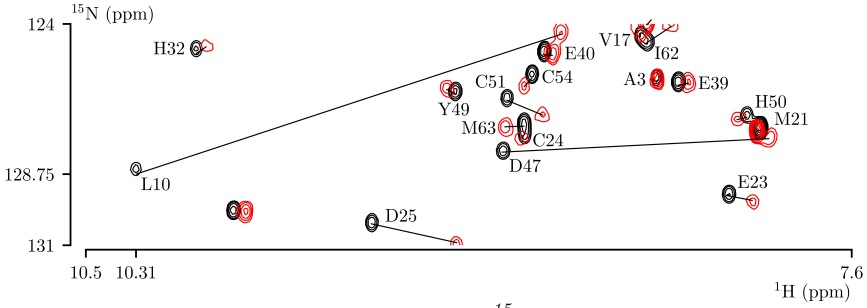


Figure 4.32: In the spectrum of V_H-29 and ^{15}N-PHF8-PHD L10 and D47 may have been erroneously assigned due to line broadening and resonance splitting.

However, leucine 10's resonance could have broadened beyond detection and been assigned, erroneously, to a different resonance: one that has spilt. For instance, leucine 10 could actually have been assigned to a split resonance from glutamine 40; or what is assigned to aspartate 47 could actually belong to histidine 50 or methionine 27. The spectra of V_HH-24 contains an assignment, valine 17, that could be interpreted similarly.

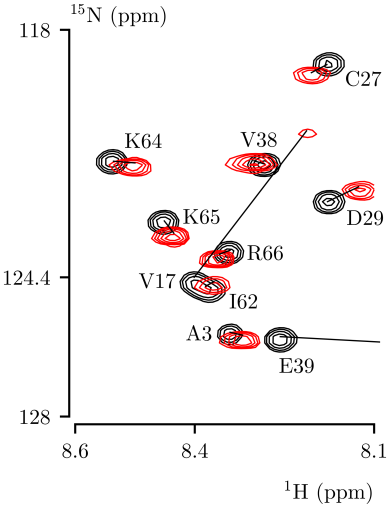


Figure 4.33: In the spectrum of V_HH-24 and ^{15}N-PHF8-PHD, V17 is potentially assigned to a peak originating from another residue.

In the spectrum of V_H-29 with ^{15}N-PHF8-PHD, tyrosine 49 is assigned a resonance near leucine 10. The distance between these peaks is substantial, perhaps even unrealistic. Another interpretation is that tyrosine 49's resonance has not been detected, but leucine 10's resonance has split in two and replaced the lost tyrosine resonance.

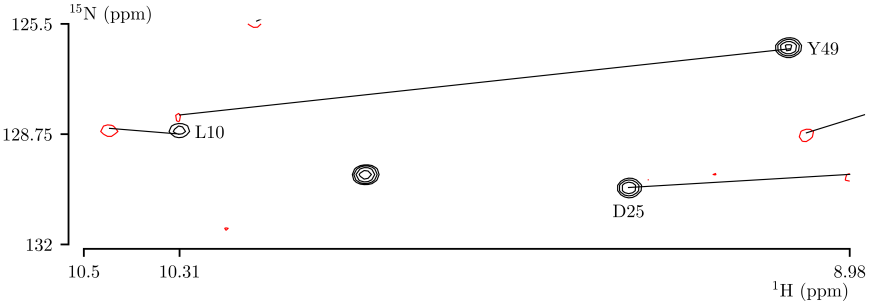


Figure 4.34: Tyrosine 49, in the spectrum of V_H-29 and ^{15}N-PHF8-PHD could have been incorrectly assigned to multiple resonances actually originating from leucine 10.

The source of this probable motion in either antibody or antigen, evidenced by the appearance and disappearance of resonances upon complex formation, is unclear without at least a backbone assignment of each antibody, and a reassignment of both antibody and PHF8-PHD in complex. But its existence, not only for the antibodies V_HH-24, V_H-25, and V_H-29, but also PHF8-PHD's substrate peptide, is enough to suggest that PHF8-PHD is

more flexible than assumed based on the structures currently available, and that this property may have consequences for inhibitor design.

Crystallography

A structure of V_HH-24 or V_H-29 in complex with PHF8-PHD is desirable because both antibodies inhibit recognition of H3K4me3 by binding at the substrate recognition site. But PHF8-PHD is also small enough that V_H-25, which binds slightly outside the substrate recognition site, could still be useful for inhibitor design. All three of the antibodies were put into crystal trials at 18 °C at a concentration of 10 mg mL^{-1}, alone and in complex with PHF8-PHD. From these six crystallisation trials the only hits were for V_HH-24 and PHF8-PHD, in conditions with TRIS between pH 7 and 9, with 10 % (w/v) to 30 % (w/v) polyethylene glycol (PEG) of average molecular weight 6000 Da to 8000 Da, and sometimes with NaCl or NaOAc at a concentration of 100 mM. The hits were mostly needles, crystalline but not large enough to collect an adequate dataset. Grid screens on the initial hit conditions improved their form somewhat, but even at 4 °C small crystal would finish growing in only 8 h to 10 h. Crushing these small crystals to seed hanging drops produced a few slightly larger crystals, but they did not diffract. The addition of NaOAc seemed to slow crystal growth a little, perhaps by increasing protein solubility slightly, though the crystals still remained small. In the end, the few larger crystals that sporadically appeared in replicate grid screens were mounted. Several diffracted to 1.39 Å.

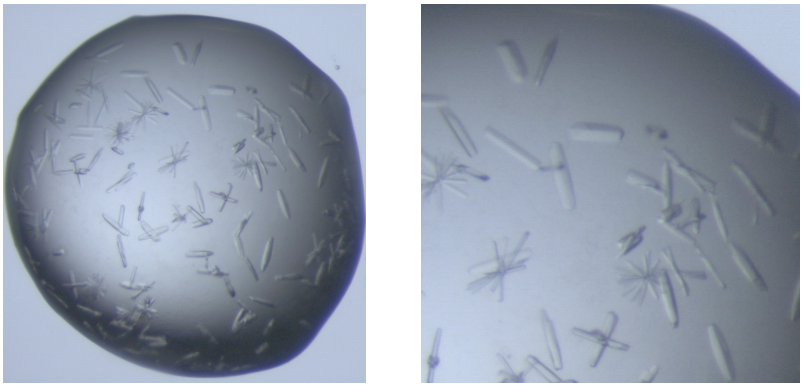


Figure 4.35: Small crystals from V_HH-24 in complex with PHF8-PHD.

The water content of these crystals was the first indication that V_HH-24 and PHF8-PHD could not both occupy a unit cell. On solving the structure it became clear V_HH-24 had crystallised without PHF8-PHD. A structure of V_HH-24 alone is still useful, but accurate information on the interactions made between antibody and antigen will probably require at least a backbone assignment of V_HH-24 and paratope mapping.

V_HH-24 adopts the immunoglobulin fold expected of both conventional antibodies and V_HHs. Perhaps the feature that differentiates V_HH-24 most is CDR1, which is an α-helix. The helix positions tyrosine 34 to point outwards, and since the structure is not conserved among V_HH CDR1s it may identify the PHF8-PHD paratope. CDR1 also includes the positively charged residues arginine 28 and asparagine 33 and negatively charged glutamine 32, while CDR3 positions lysines 105 and 104 towards solvent on a loop. These charged residues would, perhaps, not be significant but for PHF8-PHD's specificity for permanently positively charged trimethyl-lysine and an adjacent arginine on its H3K4me3 substrate – a predisposition possibly exploited by V_HH-24.

Wavelength	0.928 19 Å
Space Group	P $2_12_12_1$
Unit Cell (a, b, c)	32.28 33.64 94.93
	90 90 90

	Overall/Inner/Outer
Low Resolution Limit (Å)	27.44/27.44/1.41
High Resolution Limit (Å)	1.39/7.62/1.39
R_{merge} (within I/I−)	0.031/0.016/0.51
R_{merge} (all I+ and I−)	0.034/0.016/0.534
Reflections	92869/844/803
(Unique)	(20241)/(172)/(594)
Completeness	94.1/98.8/56.6
Multiplicity	4.6/4.9/1.4
$I/\sigma(I)$	22.1/72.8/1.2

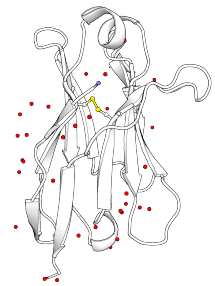
R/R_{free}	0.189/0.2172
RMS Bond Length (Å)	0.0270
RMS Bond Angle (°)	2.2083

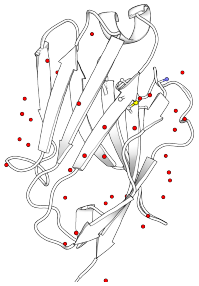
Overall B-factor (Å^2)	20.25
Protein	20.1
Water	24.67

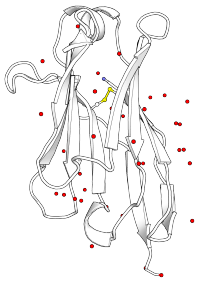
Ramachandran Preferred	110
Allowed	2
Outliers	0

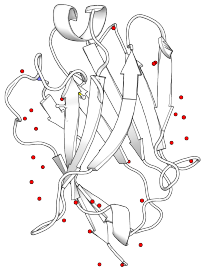
Figure 4.36: Crystallographic and refinement details for V_HH-24.

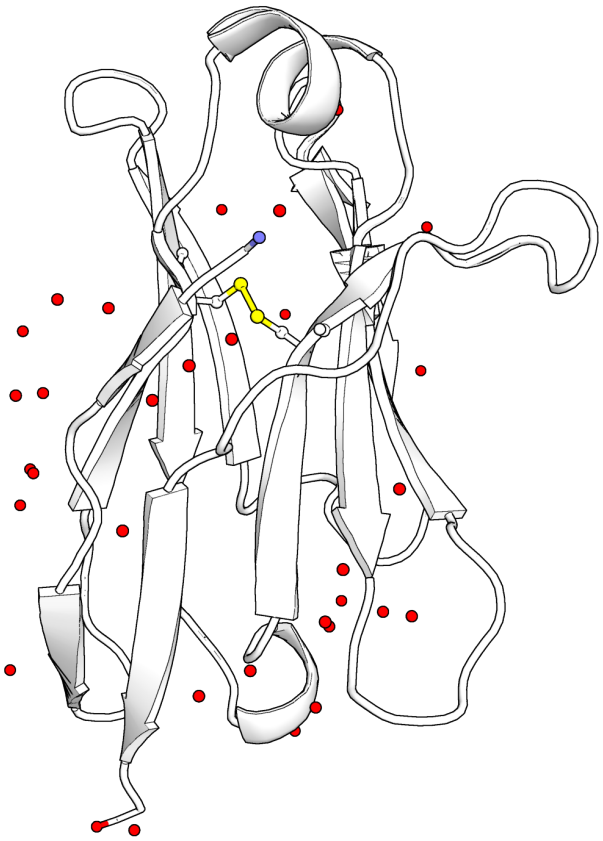
Figure 4.37: (Overpage) (Right) A "cartoon" representation of V_HH-24 showing crystallographic water molecules as red spheres, a single intramolecular disulfide bond in yellow, and N-terminal amine nitrogen and C-terminal carbonyl oxygen as blue and red spheres, respectively. (Left) The same structure rotated through 360° in increments of 90°.

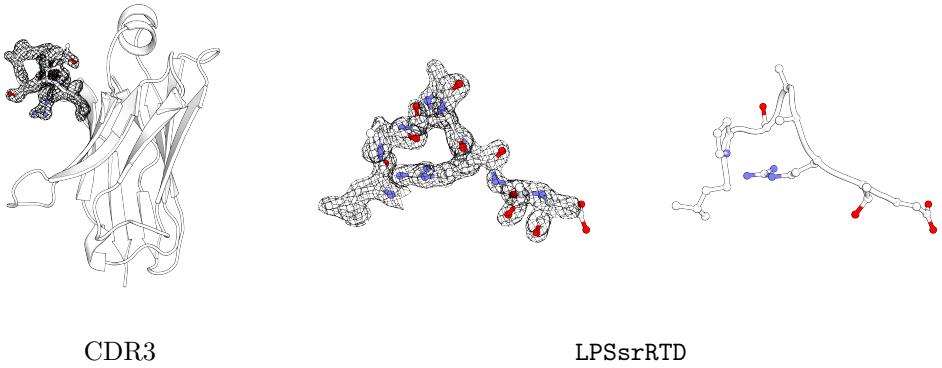




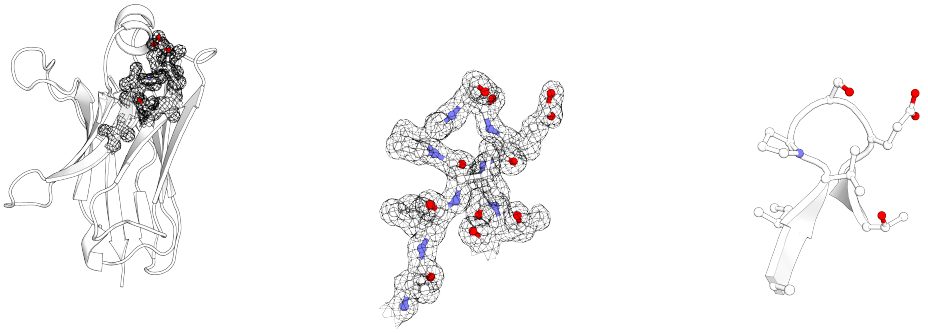




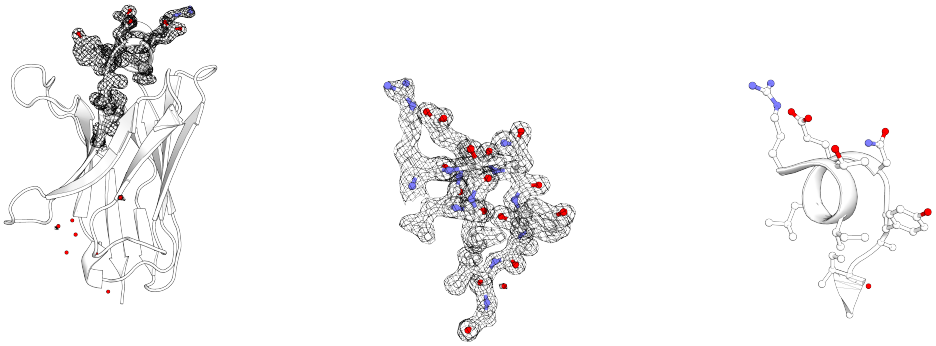




CDR3 LPSsrRTD



CDR2 AIVPSEGT



CDR1 RTILENYAIG

Figure 4.38: CDRs of V_HH-24 and their corresponding electron density maps at 1σ. Sequences printed left to right from the N-terminus as they are shown in each structure. The side chains of residues represented by lowercase letters have been "stubbed" due to missing electron density.

Summary of Main Results

Six immune phage libraries were constructed from lymph node, bone marrow, and spleen-derived B-cells of the two rabbits immunised with $PHF8_{1-447}$ in the previous chapter. These libraries were panned against PHF8-PHD in solution, and against PHF8-PHD immobilised on plastic, which changed the clones that were recovered. Some clones matched both the heavy and light chain sequence, as well as the heavy and light chain pairing, of antibodies previously isolated directly from B-cells – an indication they are dominant clones. None of the antibodies recovered from this panning, and that could be expressed as scFvs, formed more than a small amount of complex with PHF8-PHD in an SEC assay.

One V_HH, V_HH-24, and two V_Hs, V_H-25 and V_H-29, all capable of forming a complex with PHF8-PHD in solution, were identified by solution-based panning of six pre-immune llama phage libraries. In an SPR assay, V_HH-24 and V_H-29 blocked recognition of H3K4me3 by PHF8-PHD, but V_H-25 appeared compatible with H3K4me3 recognition. Epitopes for V_HH-24 and V_H-29 mapped by NMR chemical shift perturbation overlapped with PHF8-PHD's substrate binding site, whereas V_H-25's epitope bordered the site. Although purified as a complex with PHF8-PHD, upon solving the structure, V_HH-24 was found to have crystallised alone.

4.3 Conclusions

V_HH-24 and V_H-29 engage PHF8-PHD at its H3K4me3 binding site, and V_H-25's epitope is not far removed. Each could potentially inspire a small molecule inhibitor, but thorough structural characterisation is essential. Though V_HH-24 only crystallised alone, and V_H-25 and V_H-29 didn't crystallise either alone or in complex with PHF8-PHD, NMR remains an option to complete the characterisation. A structure of a V_HH or V_H and PHF8-PHD complex could be solved by NMR on its own. This entirely NMR-based approach will probably be needed for V_H-25 and V_H-29, for which there is

no auxiliary structural information. The considerable time such studies take is, however, made worthwhile by the absence of suitable PHD inhibitors; a domain whose homologues number some 150.

A potential complement to this structural information would be computational prediction. From the epitope of V_HH-24 inferred by chemical-shift perturbation, and the antibody's apo structure, macromolecular docking could be constrained. Backbone assignment of V_HH-24, and paratope mapping by chemical shift perturbation could provide further refinement as in the model of IL-1β bound to an scFv,[9] and other investigations of protein-protein interactions.[10,11,12] However, for V_H-25 and V_H-29, which could not be crystallised, unless their structures are determined by NMR, homology modelling must precede docking.

In general, prediction of antibody structure from sequence is helped by structural conservation of the immunoglobulin fold, and CDRs, besides heavy chain CDR3, adopting a small set of canonical structures[13,14*f*]. However, molecular recognition can be sensitive even to small changes in structure, and what is an acceptable deviation between a homology model and experimentally-determined structure, may make docking misleading. A straightforward approach to validating these models experimentally would be to dock PHF8-PHD and modelled antibody structures based on NMR-determined epitope constraints, cluster the structures according to binding pose using established methods,[17] and test predicted interactions by mutating the residues involved to alanine. Decreases in binding affinity would thus imply genuine interactions. This mutagenesis information, potentially in combination with hydrogen deuterium exchange mass spectrometry or residual dipolar couplings from NMR could also further refine any models.[18]

V_HH-24, V_H-25, and V_H-29, seemingly the only antibodies discovered here able to form a complex with PHF8-PHD in solution, originate not

[*f*]The extent to which camelid antibodies respect these canonical structures is still unclear. V_HHs, and potentially V_Hs, seem to extend the canonical structural repertoire.[15,16]

from an immunisation, but from a pre-immune llama phage display library. Neither the direct isolation of antibodies from the B-cells of rabbits immunised with $PHF8_{1–447}$, nor immune phage display libraries constructed from the same and panned in solution on PHF8-PHD returned antibodies that formed more than a tiny amount of complex in a SEC assay. Among the antibodies recovered from phage display were sequences already familiar from directly isolated, cognate pairs, preserved through the random recombination of heavy and light chains that accompanies library construction. Such a reunion of heavy and light chain is almost impossible without being somehow favoured. The recovery of multiple distinct cognate pairs implies they were dominant in the original immune response, their recombination favoured by the excess of B-cells expressing the clones.

If, as the recovery of cognate pairs from phage display suggests, these antibodies that seem to barely recognise PHF8-PHD in solution were dominant in the original immune response, it implies a compromised immunisation: what these antibodies have been presented *in vivo* does not resemble what they are presented *in vitro*. Since the sample of $PHF8_{1–447}$ used for immunisation was also that used in screening, such a compromise must either directly precede or follow immunisation. The antigen has almost certainly changed in some way as the solution binding problem effected antibodies from both rabbits immunised with $PHF8_{1\text{-}447}$. That immunisation is implicated, heavily favours either denaturation or proteolysis of the immunogen as the probable cause.

The discovery of V_HH-24, V_H-25, and V_H-29 concluded my investigations into the inhibition of PHF8 and KDM4D. Though what they have to tell small-molecule discovery is not realised fully in these pages, this initial characterisation demonstrate considerable potential. Potential that is within the reach of NMR studies and need not hindered by crystallography.

References

[1] J. R. Horton et al. "Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases". In: *Nature structural & molecular biology* 17.1 (2010), pp. 38–43.

[2] H. Qi et al. "Phagemid vectors for phage display: properties, characteristics and construction". In: *Journal of molecular biology* 417.3 (2012), pp. 129–143.

[3] A. Scott-Tucker. unpublished. 2016.

[4] S. Muyldermans. "Single domain camel antibodies: current status". In: *Reviews in Molecular Biotechnology* 74.4 (2001), pp. 277–302.

[5] A. Sircar et al. "Analysis and modeling of the variable region of camelid single-domain antibodies". In: *The Journal of Immunology* 186.11 (2011), pp. 6357–6367.

[6] S. Grzesiek and A. Bax. "An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins". In: *Journal of Magnetic Resonance (1969)* 99.1 (1992), pp. 201–207.

[7] S. Grzesiek and A. Bax. "Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR". In: *Journal of the American Chemical Society* 114.16 (1992), pp. 6291–6293.

[8] M. P. Williamson. "Using chemical shift perturbation to characterise ligand binding". In: *Progress in nuclear magnetic resonance spectroscopy* 73 (2013), pp. 1–16.

[9] I. C. Wilkinson et al. "High resolution NMR-based model for the structure of a scFv-IL-1β complex". In: *Journal of Biological Chemistry* 284.46 (2009), pp. 31928–31935.

[10] C. Williams et al. "Structural insights into the interaction of insulin-like growth factor 2 with IGF2R domain 11". In: *Structure* 15.9 (2007), pp. 1065–1078.

[11] L. Muresanu et al. "The electron transfer complex between cytochrome c552 and the CuA domain of the thermus thermophilus ba3 xxidase: a combined NMR and computational approach". In: *Journal of Biological Chemistry* 281.20 (2006), pp. 14503–14513.

[12] D. Owen et al. "Molecular dissection of the interaction between the small G proteins Rac1 and RhoA and protein kinase C-related kinase 1 (PRK1)". In: *Journal of Biological Chemistry* 278.50 (2003), pp. 50578–50587.

[13] B. Al-Lazikani, A. M. Lesk, and C. Chothia. "Standard conformations for the canonical structures of immunoglobulins1". In: *Journal of molecular biology* 273.4 (1997), pp. 927–948.

[14] A. Sircar. "Methods for the homology modeling of antibody variable regions". In: *Homology Modeling*. Springer, 2011, pp. 301–311.

[15] K. Decanniere, S. Muyldermans, and L. Wyns. "Canonical antigen-binding loop structures in immunoglobulins: more structures, more canonical classes?" In: *Journal of molecular biology* 300.1 (2000), pp. 83–91.

[16] V. K. Nguyen et al. "Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire". In: *The EMBO journal* 19.5 (2000), pp. 921–930.

[17] D. Kozakov et al. "Optimal clustering for detecting near-native conformations in protein docking". In: *Biophysical journal* 89.2 (2005), pp. 867–875.

[18] A. D. Van Dijk, R. Boelens, and A. M. Bonvin. "Data-driven docking for the study of biomolecular complexes". In: *The FEBS journal* 272.2 (2005), pp. 293–312.

5 Problems in Solutions

5.1 Introduction

Immobilising an antigen on plastic can expose epitopes otherwise hidden, and so, elicit antibodies specific for both the antigen and its mode of presentation. This appeared to be the condition of P2, P3, P23, and P36, four antibodies specific for the PHD of PHF8. By ELISA, the antibodies recognised plastic-immobilised PHF8-PHD, but even at concentrations exceeding 20 µM, complex formation in solution, measured by SEC, was very low. These antibodies represent the majority raised against $PHF8_{1–447}$, which in SEC assays showed only low complex formation. Here, P2, P3, P23, P36, their epitopes limited to the PHD's 66 residues, are used as a model to understand the details of antigen recognition by these antibodies and the causes underlying their poor solution binding.

5.2 Results

SPR Affinities

Even though the fraction of PHF8-PHD that can form complexes with P2, P3, P23, or P36 is likely very small, and binding only obvious with PHF8-PHD in excess, the interaction between antibody and antigen can be detected by SPR. Capturing these Fab fragments with a surface-immobilised anti-species antibodies, and presenting PHF8-PHD in the mobile phase, kinetic parameters were determined from SPR sensorgrams (figures 5.1, 5.2, and 5.3).

Ab	Ag	k_a ($M^{-1}\,s^{-1}$)	k_d (s^{-1})	K_D (M)
P36	PHD	$2.11 \times 10^4 \divideontimes 1.02$	$0.20 \times 10^{-3} \divideontimes 1.68$	$0.96 \times 10^{-8} \divideontimes 1.64$
P36	PHF8	$1.17 \times 10^4 \divideontimes 1.07$	$0.87 \times 10^{-3} \divideontimes 1.13$	$7.41 \times 10^{-8} \divideontimes 1.20$
P23	PHD	$1.49 \times 10^4 \divideontimes 1.23$	$2.42 \times 10^{-3} \divideontimes 1.08$	$1.62 \times 10^{-7} \divideontimes 1.14$
P23	PHF8	$1.43 \times 10^4 \divideontimes 1.39$	$2.52 \times 10^{-3} \divideontimes 1.12$	$1.76 \times 10^{-7} \divideontimes 1.24$
P2	PHD	2.27×10^4	5.38×10^{-4}	2.37×10^{-8}
P2	PHF8	2.41×10^4	1.50×10^{-3}	6.23×10^{-8}
P3	PHD	2.25×10^4	1.01×10^{-3}	4.51×10^{-8}
P3	PHF8	3.54×10^4	2.53×10^{-3}	7.15×10^{-8}

Table 5.1: On-rate, off-rate and dissociation constants, determined by SPR, of the antibodies P2, P3, P23, and P36 for antigens PHF8 and PHF8-PHD. Where duplicate affinity measurements were made the geometric mean and standard deviation are given.

For PHF8 to demethylate H3K4me3-K9me2 its JmjC domain and PHD need to close on each other, burying their substrate at the interface.[1] It is presumably relative motion between these two domains that makes them accessible in the first place, and that explains why, apart from P23, antibody affinity for the PHD is greater in the absence of the JmjC domain. But given that complex formation is so low in solution, the kinetic parameters for each domain are contradictory. If every molecule of PHF8-PHD could bind to one of P36 or P23, at the 20 µM concentration used for SEC in chapter 3, the law of mass action predicts 98 % of P36 to be in complex:

$$\begin{aligned}\frac{[A_bA_g]}{[A_b]} &= \frac{([A_b^T] + [A_g^T] + K_D) - \sqrt{([A_b^T] + [A_g^T] + K_D)^2 - 4[A_b^T][A_g^T]}}{2[A_b]} \\ &= \frac{(20 + 20 + 0.0096) - \sqrt{(20 + 20 + 0.0096)^2 - 4 \times 20 \times 20}}{2 \times 20} \\ &= 0.98 \text{ (2 sf.)}\end{aligned}$$

and 91 % of P23:

$$\begin{aligned}\frac{[A_bA_g]}{[A_b]} &= \frac{([A_b^T]+[A_g^T]+K_\mathrm{D})-\sqrt{([A_b^T]+[A_g^T]+K_\mathrm{D})^2-4[A_b^T][A_g^T]}}{2[A_b]}\\ &= \frac{(20+20+0.162)-\sqrt{(10+10+0.162)^2-4\times20\times20}}{2\times10}\\ &= 0.91 \text{ (2 sf.)}\end{aligned}$$

In reality, less than 10 % of molecules were in complex when analysed by SEC (chapter 3, figures 3.18 and 3.19). Since excess PHF8-PHD, but not excess antibody lead to more complex formation, it was assumed not all molecules of PHF8-PHD were capable of binding at any given time. In these SPR experiments, the quantity of antibody captured in the sensor surface is known, so their capacity to bind can be estimated from the quantity of antigen they capture. The predicted equilibrium SPR response can be calculated from the concentration, C, association constant, K_A, and maximum response at saturation, R_max, using:[2]

$$R_\mathrm{eq} = \frac{K_\mathrm{A}CR_\mathrm{max}}{K_\mathrm{A}C+1}.$$

Since the calculated response is comparable to what is recorded at saturation (figure 5.4), essentially all antibody molecules must bind, so to explain poor solution complex formation, many PHF8-PHD molecules must not. The small of fraction of PHF8-PHD that can bind is quickly exhausted by high affinity antibodies in SEC. But when replenished, as it is in the mobile phase of an SPR experiment, eventually enough will pass to register a signal and even saturate the surface. The affinities calculated from these SPR sensorgrams however, will be at best upper bounds, underestimated because the state of PHF8-PHD actually binding is diluted an unknown amount by states that cannot bind.

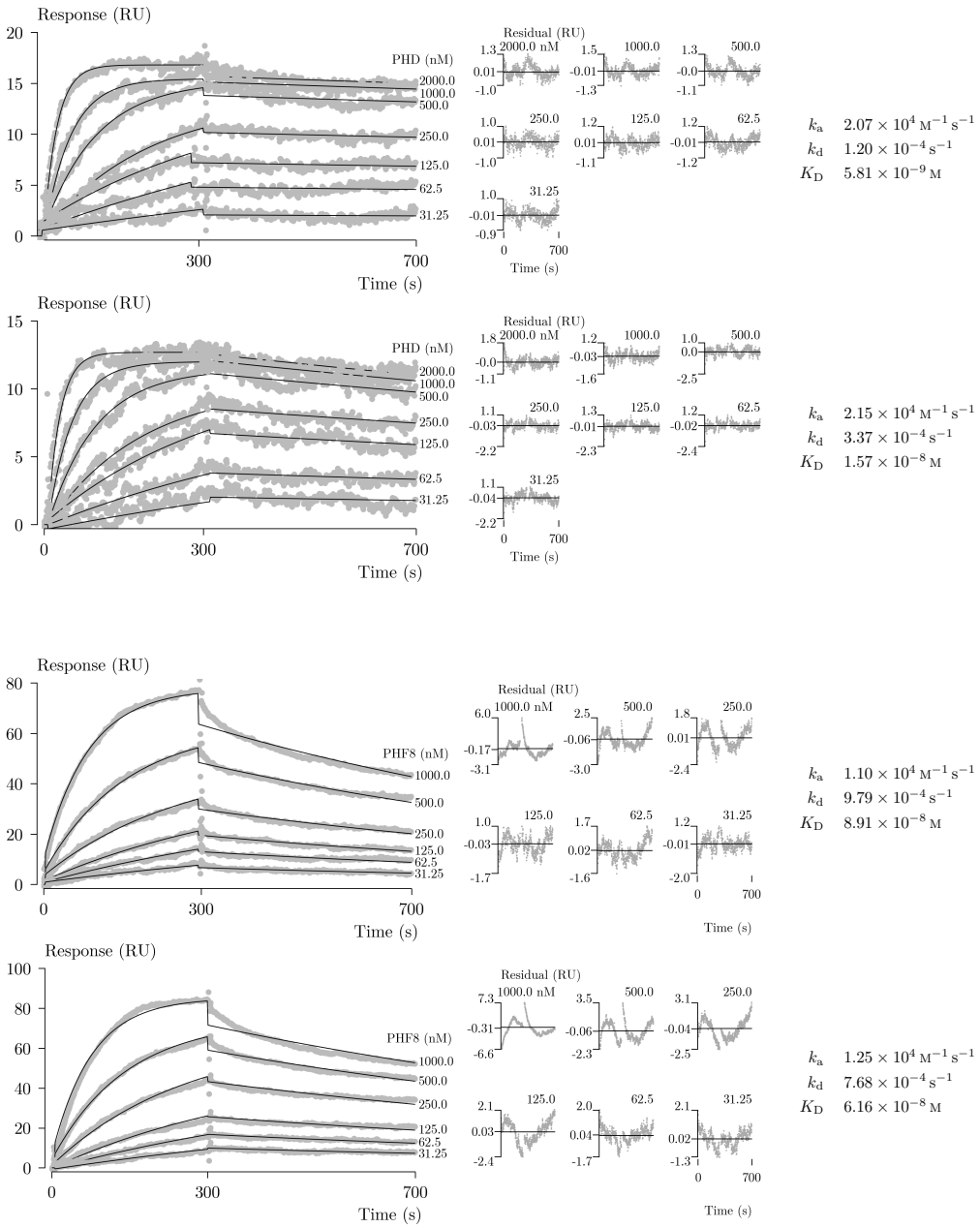


Figure 5.1: Duplicate SPR sensorgrams recorded for P36 captured by goat anti-rabbit F(ab')$_2$ antibody and presented (upper two traces) PHF8-PHD and (lower two traces) PHF8$_{1-447}$ in the mobile phase.

P23 and PHF8

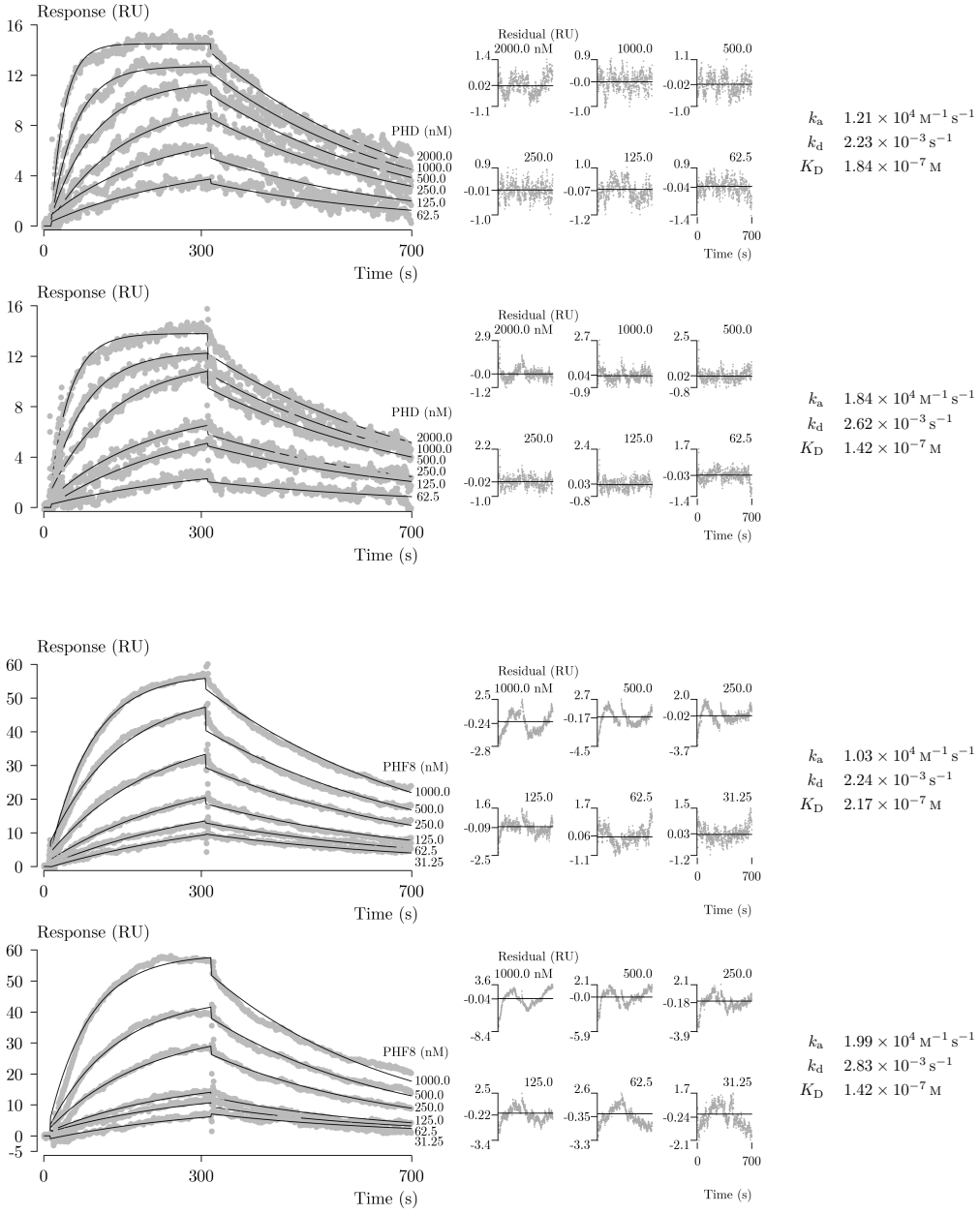


Figure 5.2: Duplicate SPR sensorgrams recorded for P23 captured by goat anti-rabbit F(ab')$_2$ antibody and presented (upper two traces) PHF8-PHD and (lower two traces) PHF8$_{1-447}$ in the mobile phase.

P2 and P3 with PHF8

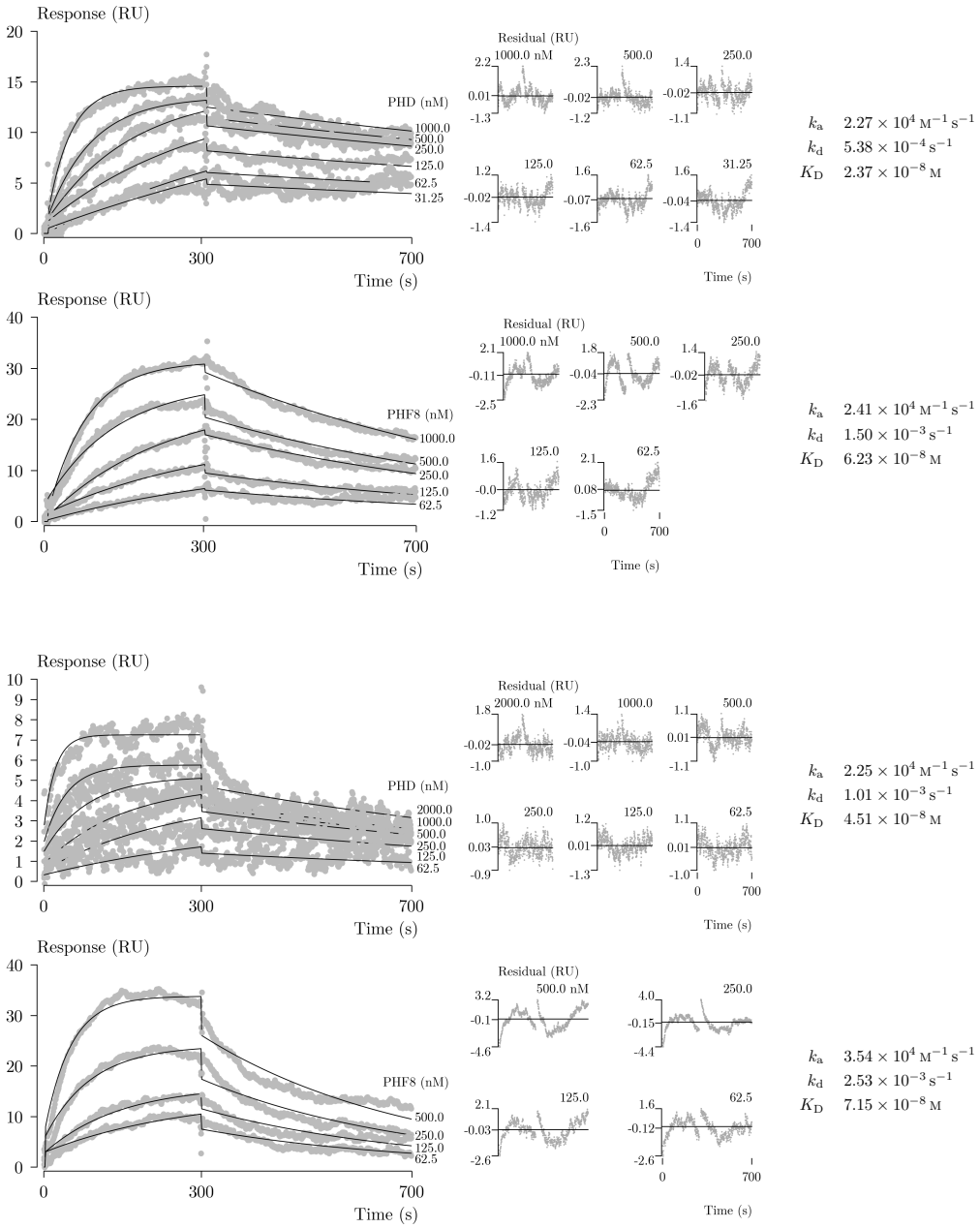
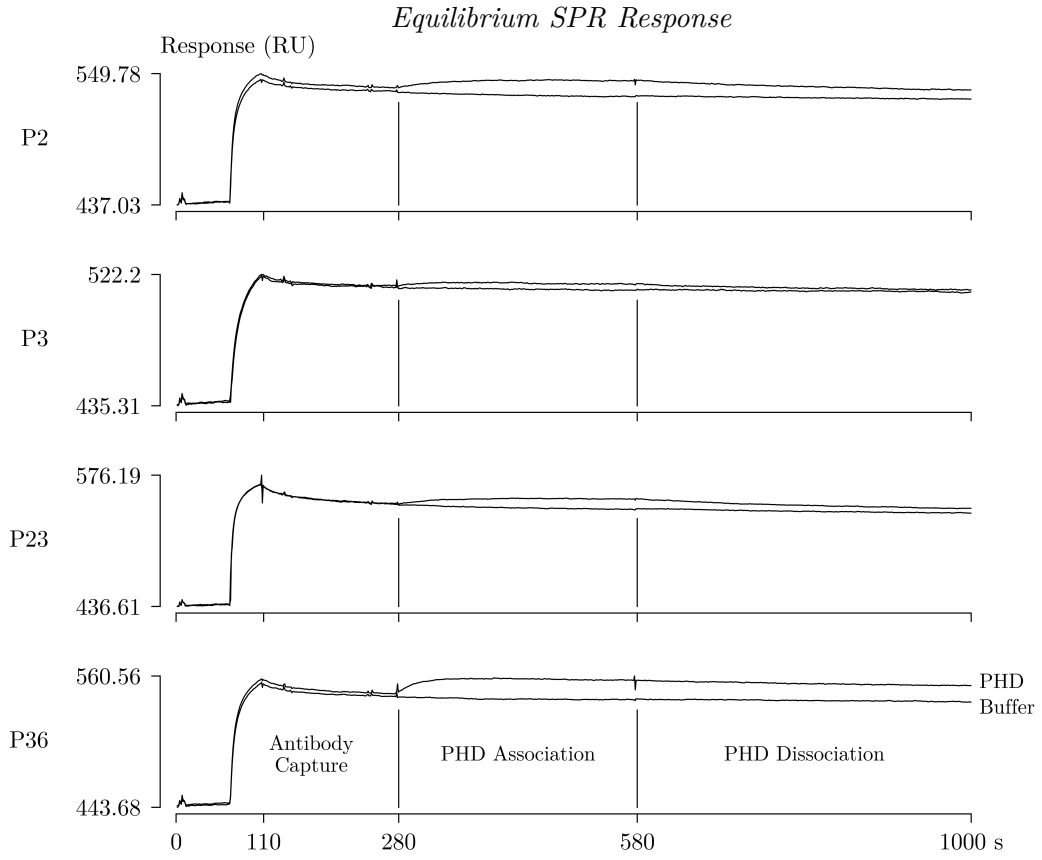


Figure 5.3: SPR sensorgrams recorded for P2 (upper two traces) and P3 (lower two trace) captured by goat anti-rabbit F(ab')$_2$ antibody and presented (upper trace) PHF8 and (lower trace) PHF8$_{1-447}$ in the mobile phase.



Ab	K_D (nM)	Conc. (μM)	R_{max} (RU)	R_{eq} (RU)	Measured R_{eq} (RU)
P2	2.37×10^{-8}	2	9.7	9.7	13.3 ± 0.2
P3	4.51×10^{-8}	1	4.8	4.8	4.1 ± 0.2
P23	1.84×10^{-7}	2	12.8	12.8	10.9 ± 0.1
P36	5.81×10^{-9}	2	15.8	15.8	17.4 ± 0.3

Figure 5.4: (Top) SPR sensorgrams for P2, P3, P23, and P36 with PHF8-PHD at the concentrations indicated in the table below. The buffer-only control, used for double referencing, is included as a reference for the level of immobilised antibody. (Bottom) The measured R_{eq}, an average of the response between 560 s *and* 570 s *shown with one standard deviation, is comparable to that calculated, indicating the majority of antibodies are bound at the concentration indicated.*

Binding Cooperativity

Since P2, P3, P23, and P36 share specificity for a domain of only 66 residues, and are all similarly affected by poor complex formation in solution, whether their's are properties of a single epitope, or ones spread among many is particularly significant. An antibody immobilised directly to the SPR sensor surface, after capturing its antigen, presents a complex to the mobile phase. Should a different antibody that would ordinarily recognise the antigen, be unable to recognise this complex, the two antibodies likely share an epitope. This is the relationship between P2 and P3, and also between P23 and P36, which cannot bind simultaneously.

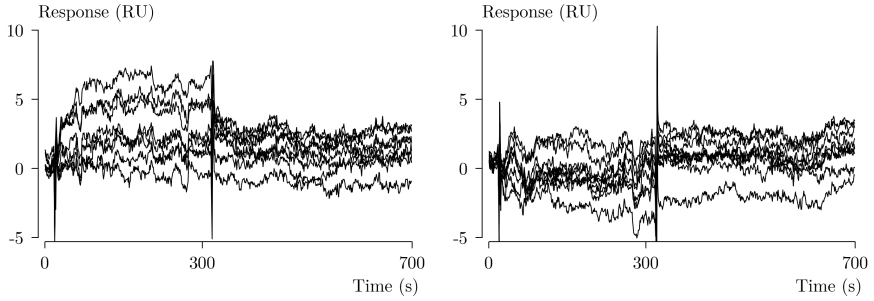


Figure 5.5: (Left) SPR response of P3, at concentrations ranging from 1000 nM *to* 16 nM *PHF8-PHD, captured by surface-immobilised-P2 and (Right) the same traces but for P36 binding to PHF8-PHD captured by surface-immobilised-P23.*

However, when PHF8-PHD is captured by P2 immobilised on the SPR sensor surface, the affinity of P23 for PHF8-PHD appears to increase by almost an order of magnitude, and for $PHF8_{1–447}$, it increases roughly two-fold. Reversing the system so that P23 or P36 capture PHF8-PHD or $PHF8_{1–447}$, the affinities of P2 seem similarly increased in the following kinetic parameters, determined from the sensorgrams of figures 5.6, 5.7, and 5.8.

Ab_1	Ag	Ab_2	k_a ($M^{-1}\,s^{-1}$)	k_d (s^{-1})	K_D (M)
	PHD	P2	2.27×10^4	5.38×10^{-4}	2.37×10^{-8}
P36	PHD	P2	$2.46 \times 10^4 \divideontimes 1.06$	$2.03 \times 10^{-4} \divideontimes 1.13$	$8.25 \times 10^{-9} \divideontimes 1.07$
P23	PHD	P2	$3.39 \times 10^4 \divideontimes 1.29$	$1.12 \times 10^{-3} \divideontimes 1.06$	$3.30 \times 10^{-8} \divideontimes 1.36$
	PHF8	P2	2.41×10^4	1.50×10^{-3}	6.23×10^{-8}
P36	PHF8	P2	$2.21 \times 10^4 \divideontimes 1.06$	$1.20 \times 10^{-4} \divideontimes 1.04$	$5.43 \times 10^{-9} \divideontimes 1.11$
P23	PHF8	P2	$1.79 \times 10^4 \divideontimes 1.07$	$1.10 \times 10^{-3} \divideontimes 1.01$	$6.18 \times 10^{-8} \divideontimes 1.09$
	PHD	P23	$1.49 \times 10^4 \divideontimes 1.23$	$2.42 \times 10^{-3} \divideontimes 1.08$	$1.62 \times 10^{-7} \divideontimes 1.14$
P2	PHD	P23	$4.35 \times 10^4 \divideontimes 1.07$	$1.43 \times 10^{-3} \divideontimes 1.13$	$3.28 \times 10^{-8} \divideontimes 1.05$
	PHF8	P23	$1.43 \times 10^4 \divideontimes 1.39$	$2.52 \times 10^{-3} \divideontimes 1.12$	$1.76 \times 10^{-7} \divideontimes 1.24$
P2	PHF8	P23	$1.71 \times 10^4 \divideontimes 1.14$	$1.32 \times 10^{-3} \divideontimes 1.04$	$7.75 \times 10^{-8} \divideontimes 1.19$

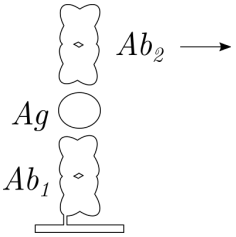


Table 5.2: Kinetic parameters determined by SPR of an antibody, Ab_2, binding to an antigen, Ag, captured by a different antibody, Ab_1, immobilised to the SPR sensor surface by amine coupling. Above each group of binding parameters, in smaller type, is the affinity originally determined for the antibody in the previous section.

An increase in apparent on-rate is expected: previously, not all molecules of PHF8-PHD could bind, so the effective concentration of the mobile phase was overestimated. With PHF8-PHD captured to the surface, precluding mass transport effects, affinity determination depends only on concentration of antibody in the mobile phase. Since all antibody molecules appear to bind, the antibody concentration reflects the population able to bind. But a change in off-rate, which is independent of concentration, suggests PHF8-PHD captured by an antibody is different from PHF8-PHD alone – most likely because it has undergone conformational change. For instance, the dissociation of P23 from the complex of PHF8-PHD and P2, is slower than dissociation of PHF8-PHD from P23. Likewise P2 dissociates more slowly from the complex of P23 and PHF8-PHD than PHF8-PHD dissociates from P2. In combination with P2, the off-rate of P36 improved so much it couldn't be accurately determined (figure 5.9). P2 and P23, and P2 and P36 are thus not only connected by recognising a similar, rare state of PHF8-PHD, but also in inducing conformational changes favourable to each other's binding.

P36-PHF8 with P2

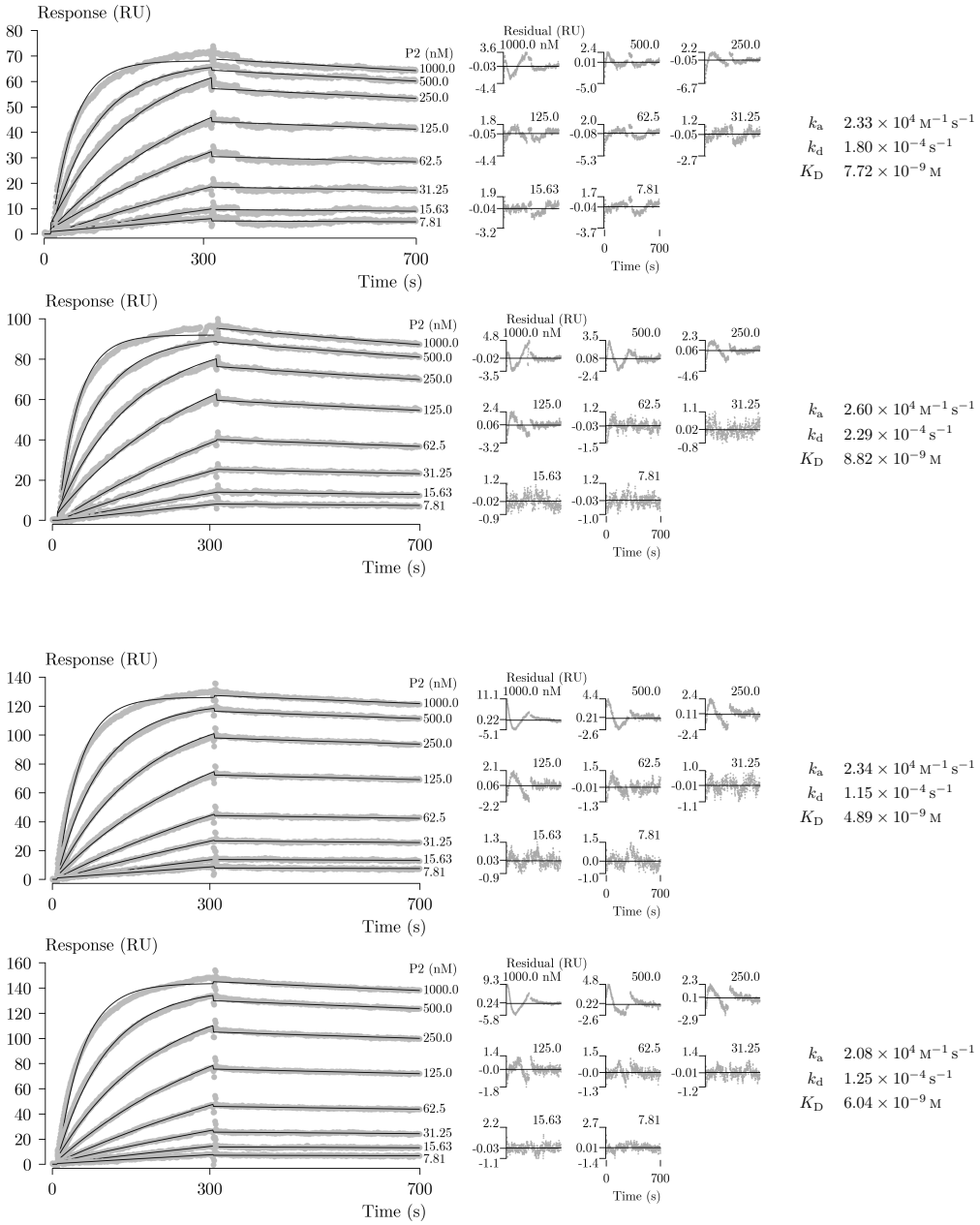


Figure 5.6: Duplicate SPR sensorgrams for P2 binding to (upper two traces) PHF8-PHD and (lower two traces) PHF8$_{1–447}$ captured by P36 immobilised to the sensor surface through amide coupling.

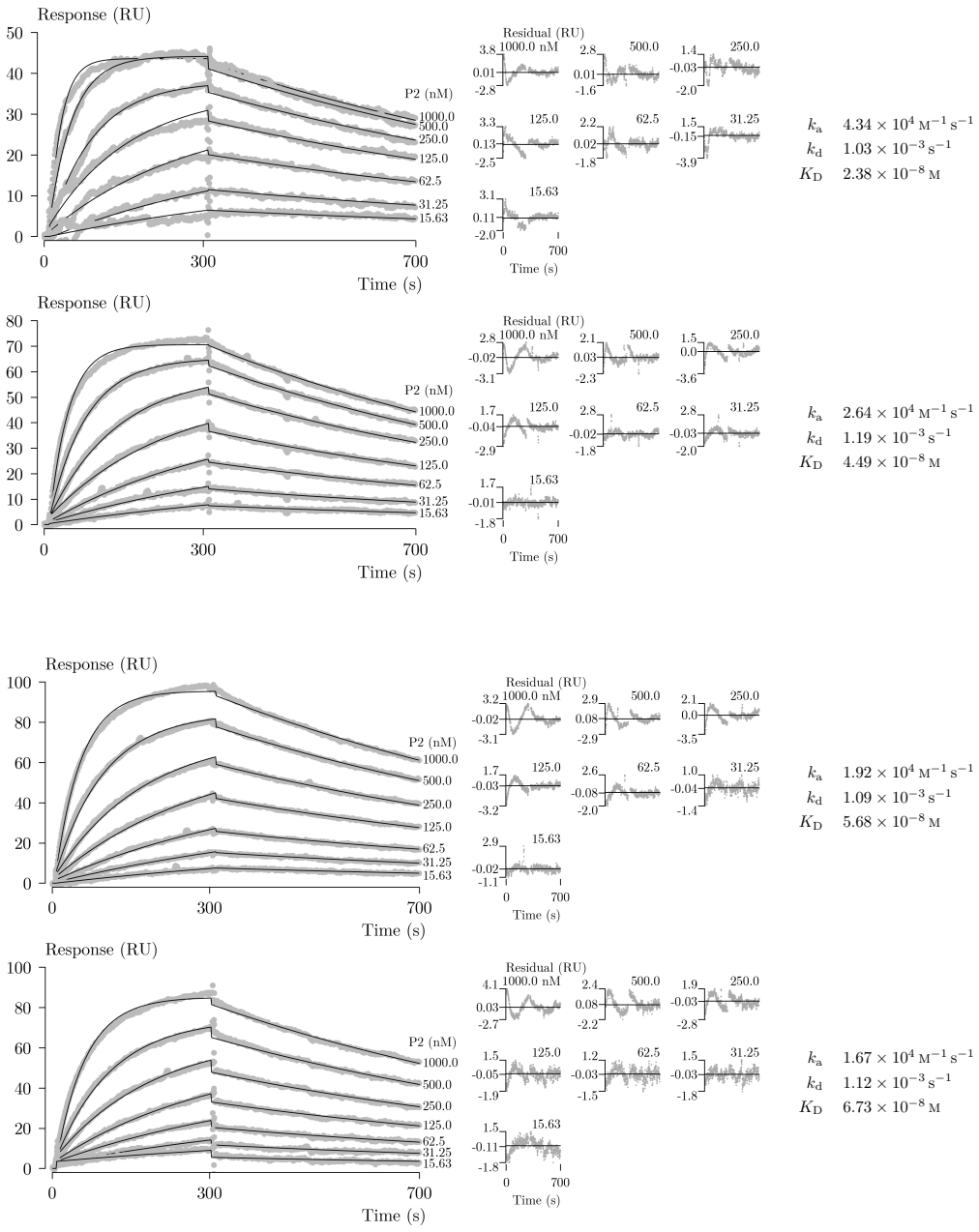


Figure 5.7: Duplicate SPR sensorgrams for P2 binding to (upper two traces) PHF8-PHD and (lower two traces) PHF8$_{1–447}$ captured by P23 which was immobilised to the sensor surface through amide coupling.

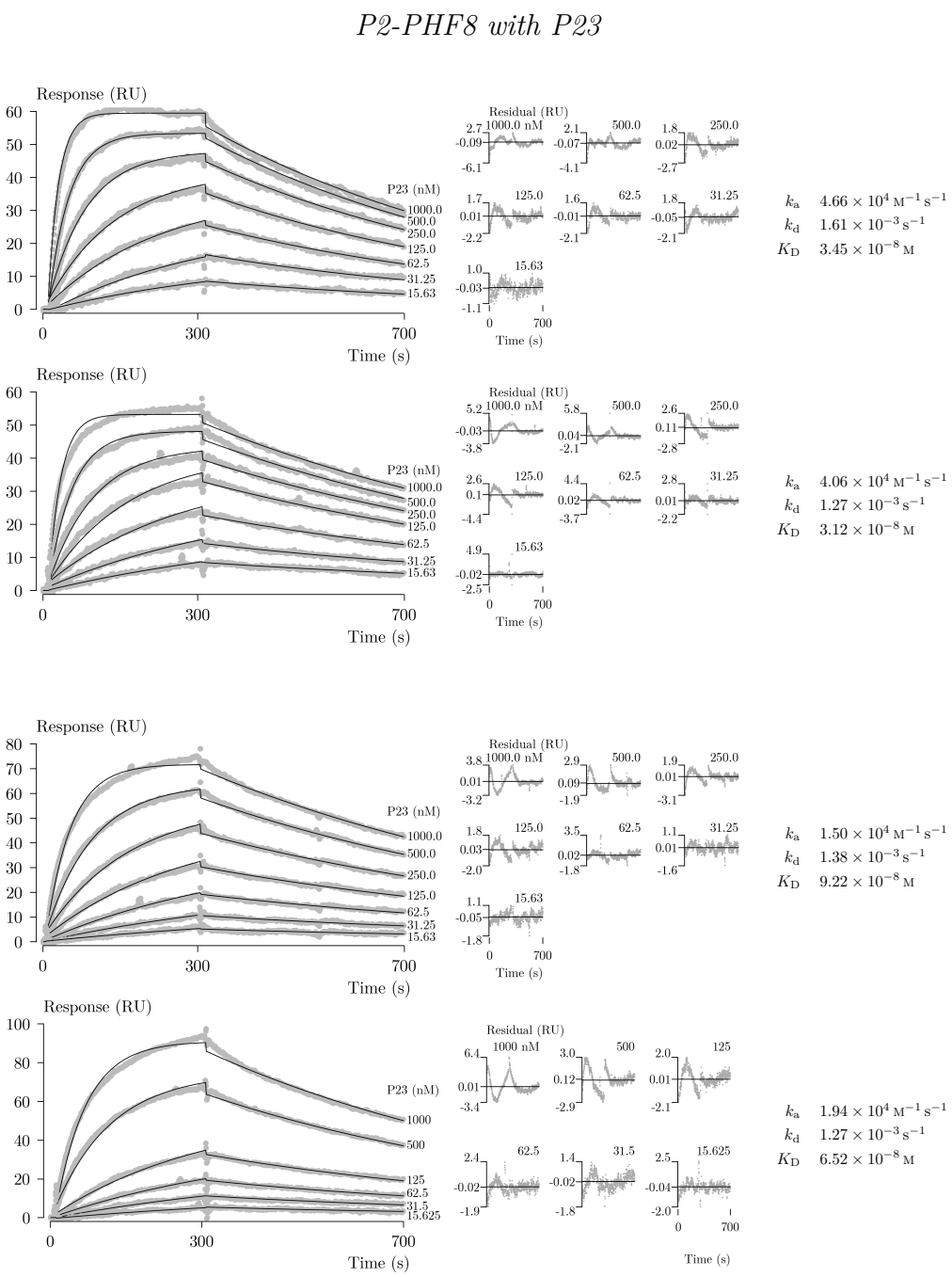


Figure 5.8: Duplicate SPR sensorgrams for P23 binding to (upper two traces) PHF8-PHD and (lower two traces) PHF8$_{1-447}$ captured by P2 which was immobilised to the sensor surface through amide coupling.

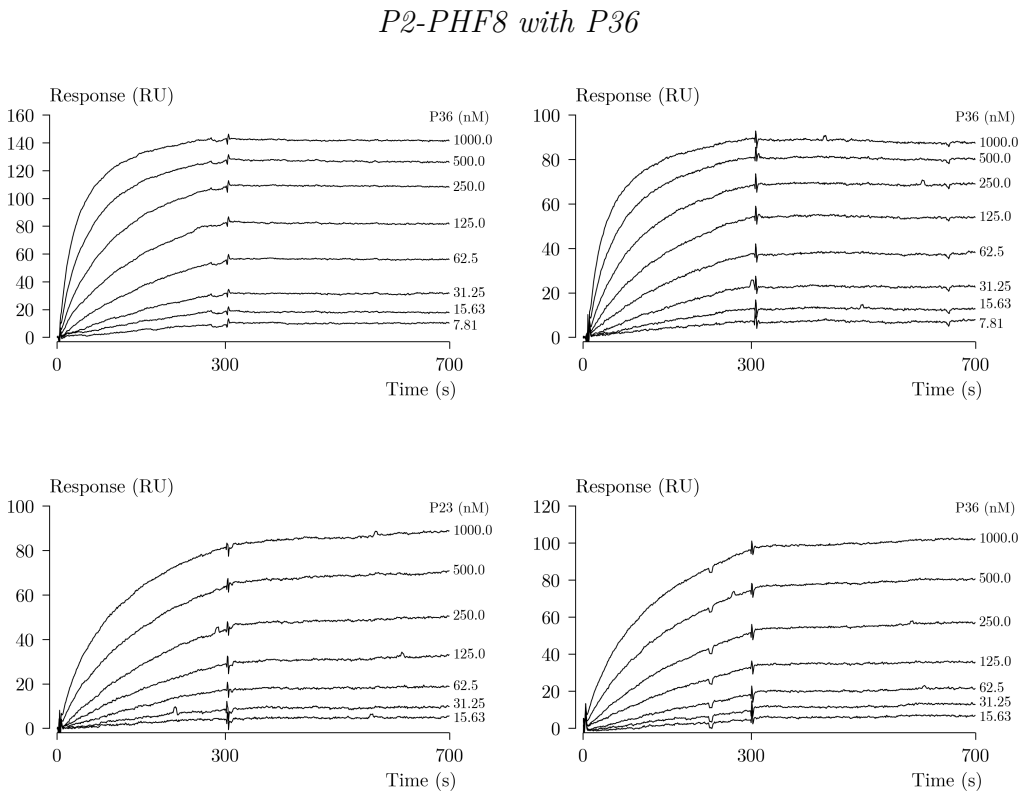


Figure 5.9: Duplicate SPR sensorgrams showing (top) P36 binding to PHF8-PHD captured by surface-immoblised-P2 and (bottom) P36 binding to PHF8$_{1-447}$ captured by surface-immobilised P2. Due to a very low off-rate accurate determination of kinetic parameters was not possible.

In size exclusion chromatograms, little complex forms when P23 and P36 are combined, just as there is little SPR response for the same combination. But when P2 and P36 are combined there is substantial complex formation, and the corresponding peak is shifted by nearly 3 min relative to the unbound peak – consistent with PHF8-PHD being in complex with more than one Fabs. Though the absorbance of two Fabs in complex is greater than one, for P2 and P36 the increase in complex and decrease in unbound antibody and antigen is, even qualitatively, more than for each antibody alone. The chromatogram of P3 and P36 is similar: there is a shift and greater height and lower retention time than the shoulder seen with PHF8-PHD and P36 or P3. The shift in retention time is less than that of P2 and P36, but the difference could be explained by the complex's geometry, conformational changes, or a rapid association and dissociation of antibody and antigen.

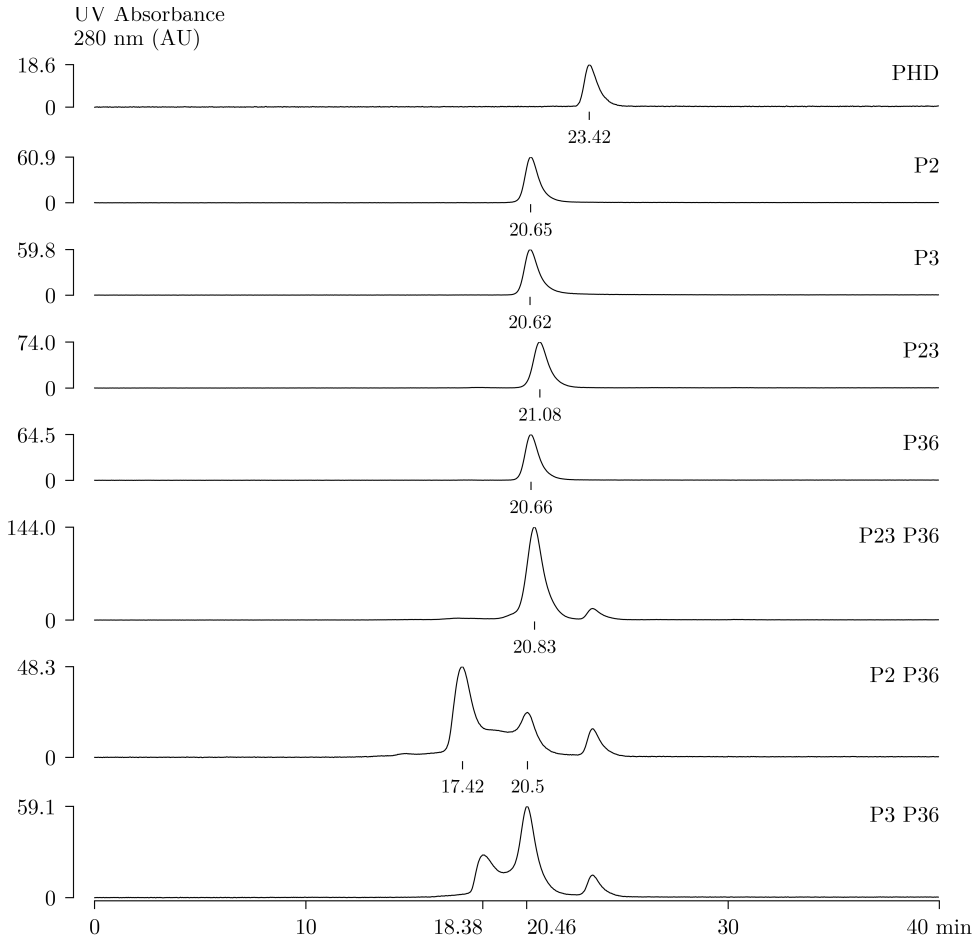


Figure 5.10: SEC chromatograms for the antibodies P2, P3, P23, and P36 combined with PHF8-PHD in the combinations indicated. The antibody concentrations were 20 µM *and the concentration of PHF8-PHD* 40 µM*. P23 and P36, whose binding is incompatible show only the slight shoulder associated with complex formation between a single antibody and PHF8-PHD, but P2 and P36, for instance, whose antigen-recognition was cooperative in an SPR assay, show much more complex forms at a higher molecular weight when the two antibodies are combined.*

Time

The amount of complex P36 forms with PHF8-PHD a concentration of 20 µM should be almost indistinguishable from the amount P23 forms. Even if the affinity of P36 for PHF8-PHD were 10 nM, and the affinity of P23 for PHF8-PHD were 120 nM (in reality they are even lower), and the seemingly permanent fraction of PHF8-PHD reserved for complex formation 10 % of

all molecules (it is likely even less), with antibody and antigen combined 1:1 at a concentration of 20 µM (an effective antigen concentration of 2 µM), for P36, the complex concentration would be

$$\begin{aligned}[A_bA_g] &= \frac{([A_b^T] + [A_g^T] + K_\mathrm{D}) - \sqrt{([A_b^T] + [A_g^T] + K_\mathrm{D})^2 - 4[A_b^T][A_g^T]}}{2} \\ &= \frac{(20 + 2 + 0.0096) - \sqrt{(20 + 2 + 0.0096)^2 - 4 \times 20 \times 2}}{2} \\ &= 2.00\,\mu\mathrm{M}\ (3\ \mathrm{sf.}),\end{aligned}$$

and for P23 the concentration would be:

$$\begin{aligned}[A_bA_g] &= \frac{([A_b^T] + [A_g^T] + K_\mathrm{D}) - \sqrt{([A_b^T] + [A_g^T] + K_\mathrm{D})^2 - 4[A_b^T][A_g^T]}}{2} \\ &= \frac{(20 + 2 + 0.162) - \sqrt{(20 + 2 + 0.162)^2 - 4 \times 20 \times 2}}{2} \\ &= 1.98\,\mu\mathrm{M}\ (3\ \mathrm{sf.}).\end{aligned}$$

And yet, much more complex forms when P36 is combined with PHF8-PHD (figure 3.18), than when P23 is combined with PHF8-PHD (figure 3.19). The reason, it appears, is the association of antibody and antigen in these traces has not yet reached equilibrium: over 29 h the P36 PHF8-PHD complex slowly accumulates:

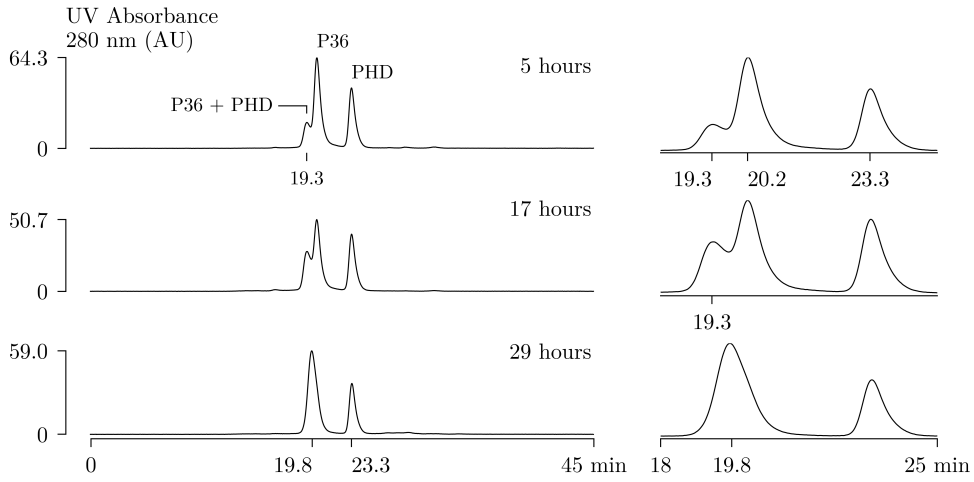


Figure 5.11: SEC chromatograms of P36 at a concentration of 10 µM *mixed with PHF8-PHD at* 50 µM *injected after* 5 h*,* 17 h*, and* 29 h*.*

Even after 29 h, complex formation is incomplete: the unbound P36 peak and its shoulder, have merged where their heights are equal, and appear as a single peak, slightly wider and with a retention time between the complex and unbound P36. The on-rate of P36 is too fast to support such a slow accumulation of complex, and since all antibody molecules are capable of binding in the few minutes it takes to record an SPR sensorgram, there must be a flux of PHF8-PHD from a state unseen by antibodies to one recognised. Since the on-rates of P36 and P23 are so similar, and the state of PHF8-PHD very likely to be the same, it is most likely P36's off-rate that supports its faster accumulation in complex. What remains is the nature of this state, and whether it is one sampled naturally, or part of a progression towards (partial) denaturation and aggregation.

EDTA

PHF8's PHD coordinates one zinc ion through cysteines 9, 11, 35, and histidine 32, and another through cysteines 24, 27, 51, 54.

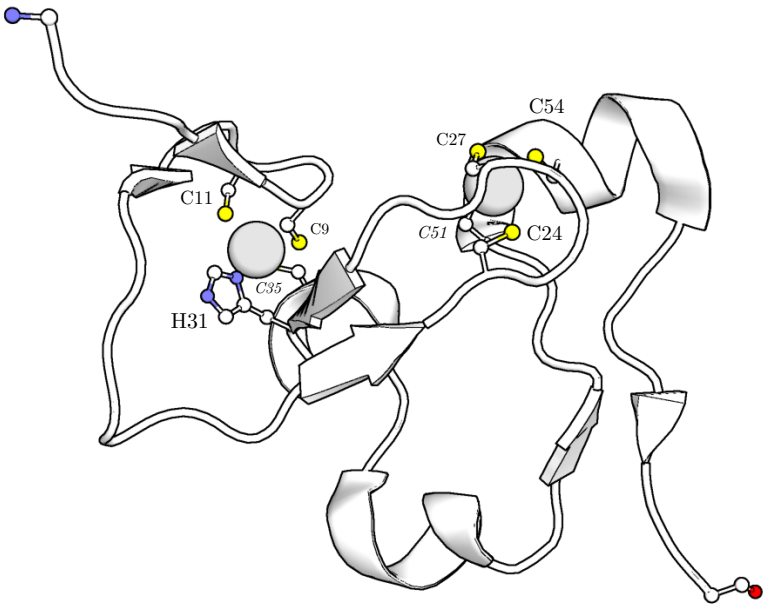


Figure 5.12: A structure of PHF8's PHD (PDB: 3KV4) showing (from left) the CCCH and CCCC motif that coordinates structural zinc. N-terminal nitrogen and C-terminal oxygen are indicated in blue and red respectively.

Zinc is likely integral to PHF8-PHD's structure as each one brings together regions of secondary structure that would otherwise be separated. But, no zinc was added during purification, so a slow disassociation of endogenous zinc from PHF8-PHD could be what underlies its transformation to a state recognised by the antibodies here. In support of this hypothesis, samples of P2 that happened to have been diluted into buffer containing 10 mM ethylenediaminetetraacetic acid (EDTA) and then buffer exchanged using a PD-10 desalting column into EDTA-free HBS, formed complexes much more readily than those not exposed to EDTA.

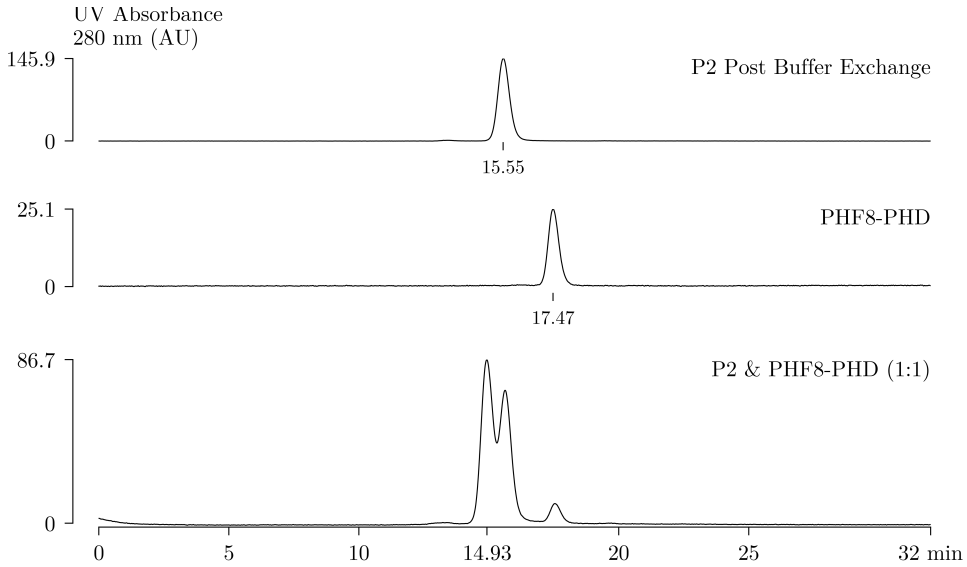


Figure 5.13: After buffer exchange from 100 mM *TRIS,* 10 mM *EDTA, pH 8 to EDTA-free PBS, P2 forms a reasonable amount of complex with PHF8-PHD even at a 1:1 stoichiometry.*

Residual EDTA remaining after buffer exchange is most likely to be the cause of this increase in complex formation, because if the buffer is exchange twice into fresh EDTA-free buffer, complex formation decreases.

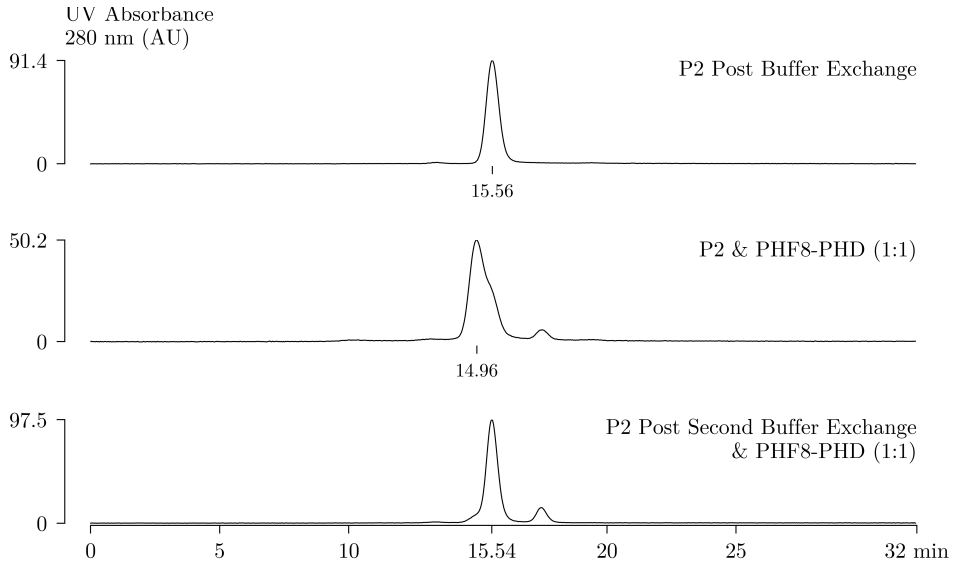


Figure 5.14: (P2, after buffer exchange from 100 mM *TRIS,* 10 mM *EDTA, pH 8 to EDTA-free HBS, more readily forms complex, but when buffer exchanged a second time into EDTA-free HBS the effect disappears, a sign residual EDTA is responsible.*

Some EDTA helps a complex form between PHF8-PHD and P2, but at higher concentrations it causes aggregation (figure 5.15). The complex between P2 and PHF8-PHD, appearing as a shoulder on P2's peak, increases with 0.5 equivalents of EDTA, but between 1 and 2 equivalents of EDTA there is little change. The major peak shifts to the complex, but the peak is also wider, probably because it is still a combination of unbound and bound P2. At higher EDTA concentration, the generation of high molecular weight species suggests aggregation is favoured over the formation of complexes. For P3, EDTA seems to do little besides causing aggregation; for P23 it is difficult to discern the complex from this aggregation; and for P36 even a 0.5 equivalents is too much (figures 5.15 and 5.16). There is a shoulder on the peak corresponding to unbound P36, and perhaps a small shoulder on that of P23, but both contain unbound antibody. For all antibodies, aggregation seems to compete with complex formation, and pushed by excess EDTA it seems the favoured process.

The peak area of each aggregate is visibly larger than that of PHF8-PHD, so, save a dramatic change in UV absorbance upon aggregation, the antibody must also be involved. The interaction could be non specific, but what is labelled aggregation in P2's traces has a lower retention time and more defined peak than the aggregates in traces for P3, P23, or P36. And with P36, the highest affinity antibody, aggregation begins at lower EDTA concentrations. So instead, the antibodies may be recognising legitimate epitopes, exposed when PHF8-PHD loses zinc, and then, aggregate as a complex with their partially denatured antigen.

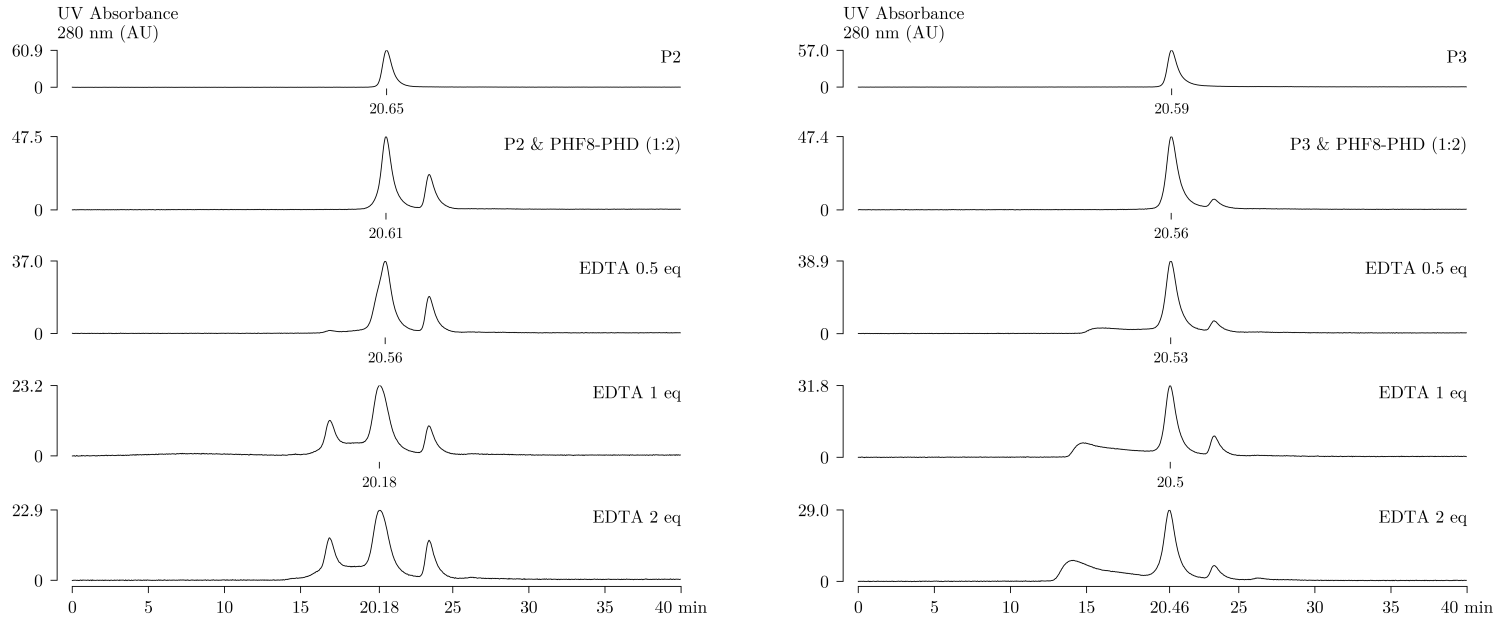


Figure 5.15: Size exclusion chromatograms of (left) P2 and (right) P3 with PHF8-PHD at a concentrations of 20 µM *and* 40 µM*, respectively, and with EDTA added at concentrations of* 10 µM*,* 20 µM*, and* 40 µM*.*

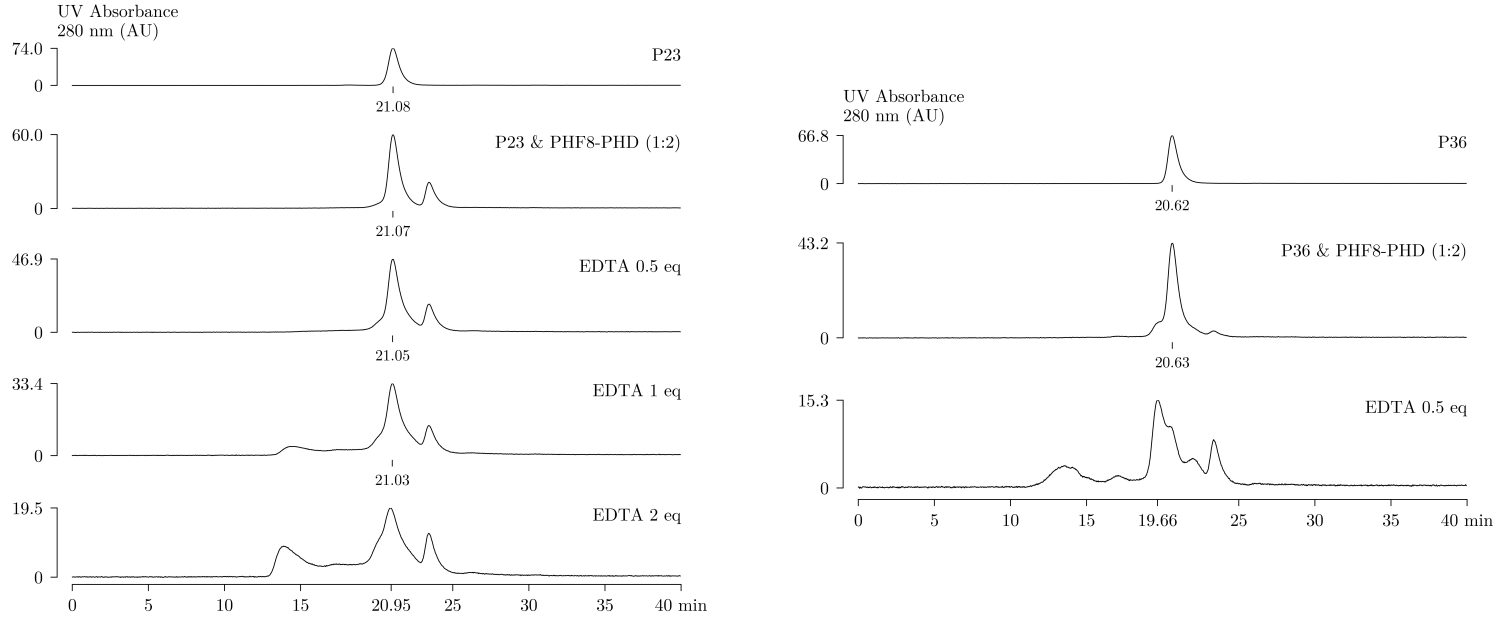
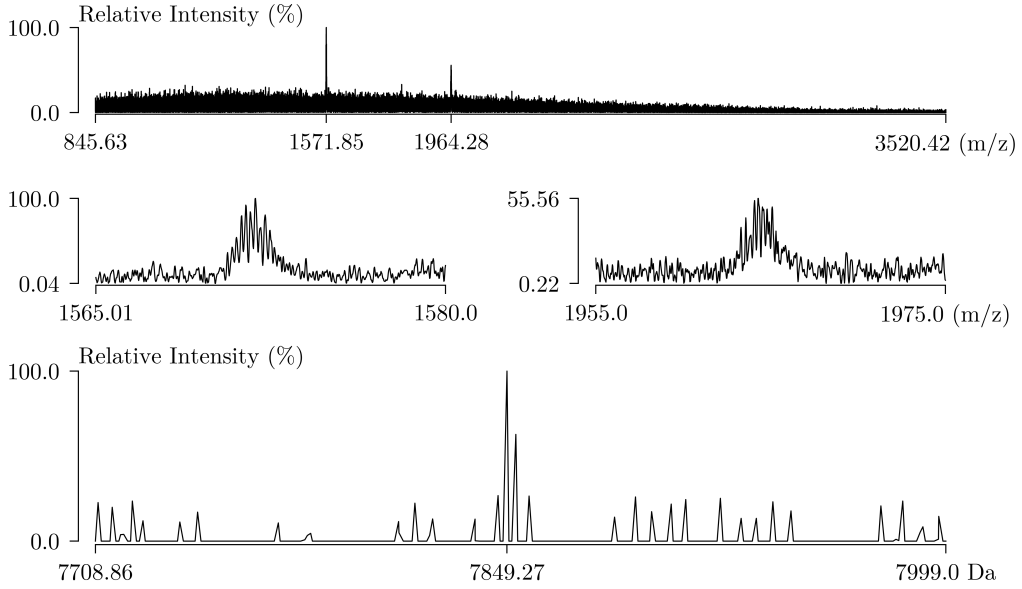


Figure 5.16: Size exclusion chromatograms of (left) P23 and (right) P36 with PHF8-PHD at a concentrations of 20 µM *and* 40 µM*, respectively, and with EDTA added at concentrations of* 10 µM*, and also* 20 µM *and* 40 µM *with P23.*

Zinc

The increase in complex formation seen at low concentrations of EDTA seems to serve the hypothesis that for PHF8-PHD to be recognised by P2, P3, P23, or P36 it must have lost one or more of its structural zinc ions. However, assayed by colourimetry, native mass spectrometry, and NMR, even after dialysis, there is no indication that PHF8-PHD has lost any zinc. When *p*-hydroxymercuriphenylsulfonic acid (PMPS) is added to a solution of PHF8-PHD, the propensity of sulphur to form mercaptide bonds will release zinc coordinated by cysteine. The zinc released can be quantified by adding 4-(2-pyridylazo) resorcinol (PAR) to the buffer, since the absorbance of the PAR-Zn complex, of 500 nm wavelength light, depends approximately linearly on concentration in the low micromolar range.[3,4,5,6] When saturated with PMPS the ratio of zinc to protein for duplicate measurements was 1.58 and 1.79. Their mean, 1.68 ± 0.15, within the error of protein concentration, is close to the expected two metal ions per molecule, and the deconvoluted native mass-spectrum of PHF8-PHD, its major peak at 7849.27 Da, confirms the metal is zinc, and again, that there are two per molecule of PHF8-PHD.



SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSIMKKRR

Monoisotopic Mass: 7722.52 u
Theoretical Mass: $7722.52 + 2 \times 63.93 = 7850.38$ u
Measured Mass: 7849.27 u

Figure 5.17: (Top) Native mass spectrum of PHF8-PHD, (middle) the same spectrum zoomed to areas containing peaks, (bottom) the deconvoluted mass spectrum. (Data courtesy of Dr. N Patel, UCB)

In addition to colourimetry and mass spectrometry, the Cβ chemical shifts of PHF8-PHD's cysteines are all below 32 ppm, which is a sign they are reduced.[7,8]

Cys.	Cβ (ppm)
9	26.11
11	27.79
24	30.03
27	29.56
35	28.40
51	26.92
54	27.17

Table 5.3: Cβ chemical shifts of PHF8-PHD's zinc coordinating cysteines are all less that 32 ppm *confirming they are reduced.*[7]

Though predicting zinc coordination from chemical shifts is non-trivial,[9] for these residues to remain cysteines in the absence of reducing agents suggests they are protected – most likely by coordinating the zinc whose presence is

already confirmed.

It took approximately 29 h for 10 % of PHF8-PHD to form a complex with P36 (figure 5.11). For a loss of structural zinc to drive this complex formation requires its dissociation be on a similar timescale. Faster and PHF8-PHD will aggregate; slower and complex formation will be too low. However, it is highly unlikely that zinc could be retained through purification in the quantities it is here if, under competition with P36, 10 % can be lost in 29 h. During purification, the 50 mL:4000 mL dialysis of PHF8-PHD for 16 h to 20 h should separate most zinc that happens to dissociate. Subsequent concentration steps, and gel filtration, would only further this loss. Losing zinc from even 10 % of PHF8-PHD molecules should be enough to introduce free (or oxidised) cysteine resonances into NMR spectra, and additional peaks to the native mass spectrum – neither of which were observed. What causes PHF8-PHD to become recognisable to antibodies is, therefore, not simply a loss of zinc. But EDTA at low concentrations assiting binding suggests PHF8-PHD without zinc resembles the form that is recognised. And that EDTA at high concentrations leads to aggregation could mean this state involves the partial denaturation implied by ELISA binding not being reflected in solution.

Cross-Reactivity

The high homology of PHF8's PHD with relatives PHD finger containing protein 2 (PHF2) and KDM7A limits antibodies that do not recognise all three proteins to only a few potential epitopes. By ELISA, P2 and P3 bind to PHDs of PHF8, PHF2, and KDM7A, but for P23 with PHF2 or KDM7A, and P36 with PHF2, binding signals are very low.

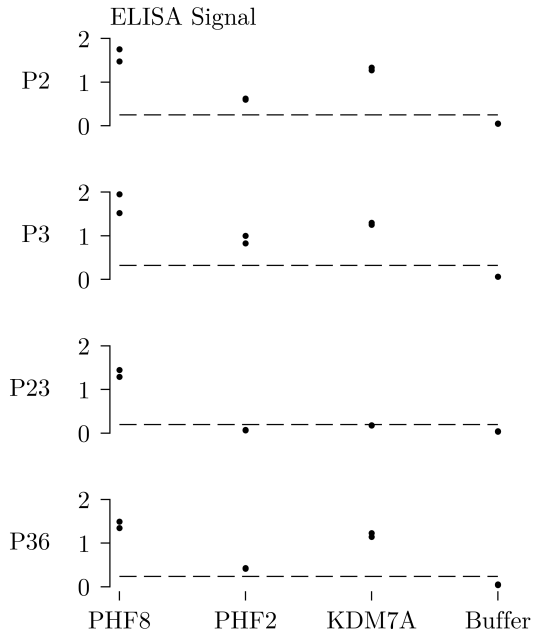


Figure 5.18: ELISA signals measured in duplicate for antibodies P2, P3, P23, and P36 and plastic-immoblised PHDs of PHF8, PHF2, KDM7A, and BSA negative control. The dashed horizontal line indicates five-times the negative controls signal used to differentiate genuine and non-specific interactions.

P36 recognises PHD's of both PHF8 and KDM7A but not that of PHF2, so potential epitopes are shortlisted to regions of sequence shared by PHF8 and KDM7A but not PHF2.

		1		2		3		4		
PHF8	SMASVPVYCLCRLPYDVTRFMIECD	MCQ	DWFHGSCVGVEE	EKAA	DID	L	YHCPNC	EVLHGPSIM	KKRR
PHF2	SMATVPVYCVCRLPYDVTRFMIECD	ACK	DWFHGSCVGVEE	EEAP	DID	I	YHCPNC	EKTHGKSTL	KKKR
KDM7A	SM----VYCVCRQPYDVNRFMIECD	ICK	DWFHGSCVGVEE	HHAV	DID	L	YHCPNC	AVLHGSSLM	KKRR

Since P36 recognises both PHF8 and KDM7A, it must tolerate the substitution of functionally equivalent residues in region 1, and therefore likely tolerates similar substitutions in PHF2 in regions 1 and 3. Similarly, in region 2 the sequence `EKAA`, in PHF8, can be replaced by *HHAV* in KDM7A, so unless a positive charge is required in position 2, or proline in position 4 significantly disrupts the secondary structure, these substitutions in PHF2 are also likely tolerated by P36. In region 4 position 2, valine, conserved in both PHF8 and KDM7A, is substituted for lysine, and a proline (serine in KDM7A), four residues later, is also substituted for lysine. Both are substitutions of a small hydrophobic residue for a large positively charged one and potentially disruptive enough to make region 4 the epitope for P36. Region 4 also begins directly after cysteines 51 and 54, two zinc coordinating residues which could explain the effect of zinc being removed on binding.

Although P36 and P23 could not simultaneously bind to PHF8-PHD in SPR assays, it does not necessarily imply they share an epitope. One antibody may be oriented so that it overlaps with the other, or the conformation recognised by one antibody may hide the epitope of the other. P23's specificity for only PHF8 could be caused by the functionally equivalent substitutions region 1 previously discounted, though a more likely epitope is in region 2 or 4, where the differences between PHF8, PHF2, and KDM7A are greater.

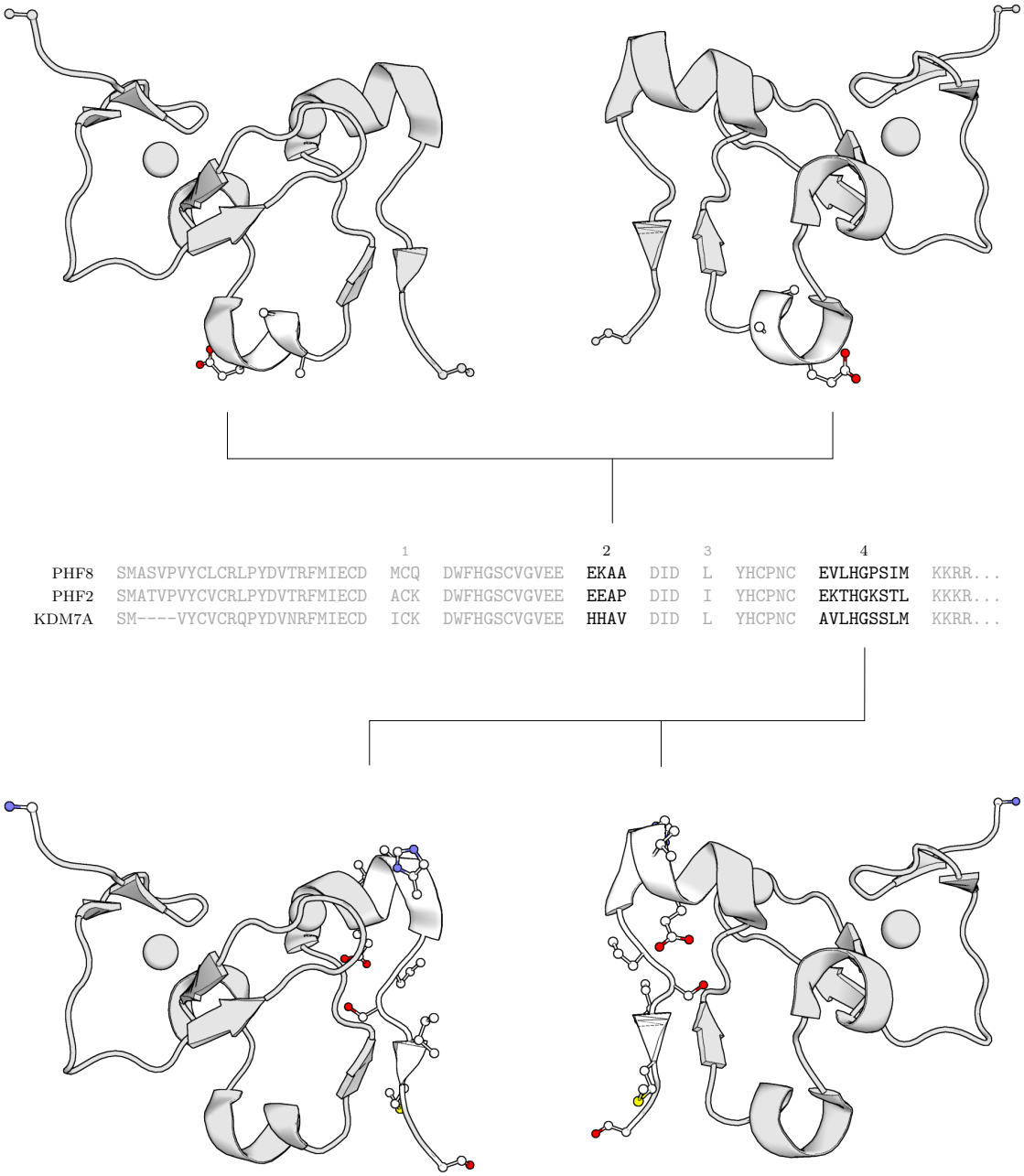


Figure 5.19: A crystal structure of PHF8 PHD (rendered from PDB: 3KV4)[1] with (top) region 2 and (bottom) region 4 highlighted in a ball and stick representation and rotated through 180°. *N-terminal nitrogen and C-terminal oxygen are indicated in blue and red respectively.*

Heat

EDTA accelerates the formation of complexes, almost certainly by seizing PHF8-PHD's structural zinc, such is its affinity for the metal ($K_D \sim 10^{-16}$).[10] But losing zinc cannot be the rate-limiting step to complex formation: the dissociation of zinc from PHF8-PHD is slower than complex formation, and high EDTA concentrations lead to aggregation not seen in its absence. A molecule of PHF8-PHD without zinc likely has its barrier to complex formation irreversibly lowered, whereas the mechanism underlying the slow complex formation of P36 and PHF8-PHD may be due to an irreversible change in PHF8-PHD, or the reversible sampling of a rare structural state. Though energy added as heat to the antibody-antigen reaction can favour dissociation, just as it can favour association, for P36 and PHF8-PHD the effect of incubation at 37 °C for 16 h is a net increase in complex formation.

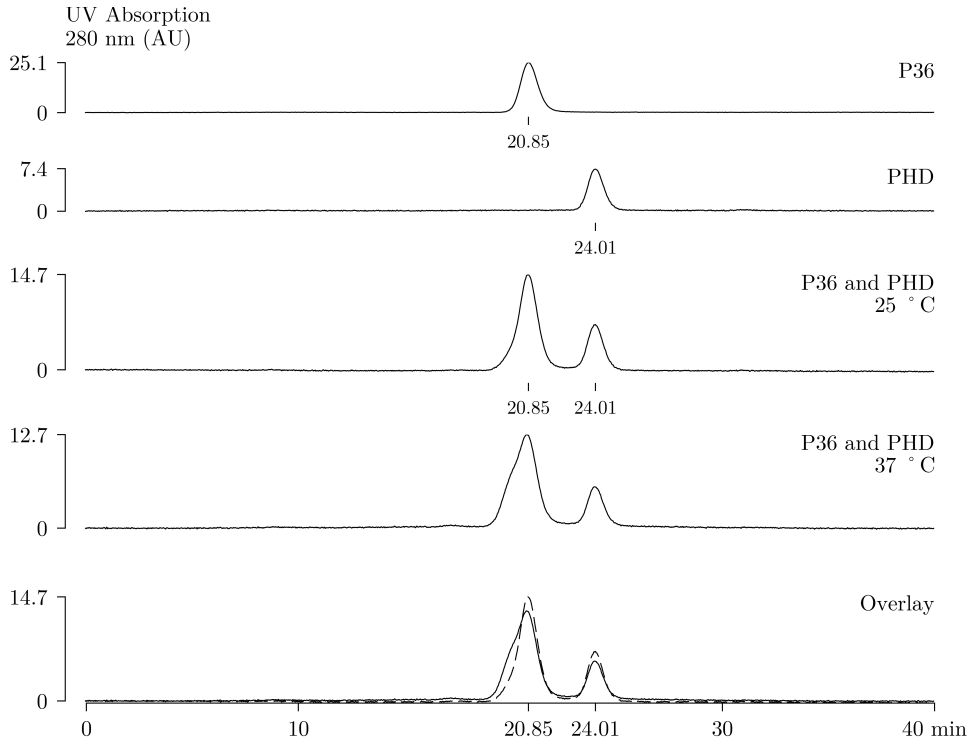


Figure 5.20: Size exclusion chromatograms of PHF8-PHD and P36 incubated together for 16 h *at either* 25 °C *or* 37 °C. *The complex appears as a shoulder on the unbound antibody's peak at* 20.83 min. *Overlaying the chromatograms corresponding to each temperature (bottom) show increased complex formation at* 37 °C *as a larger shoulder, and decreases in the peaks corresponding to unbound P36 and PHF8-PHD.*

The effect of heat on PHF8-PHD is, however, not permanent. When PHF8-PHD is incubated separately at 37 °C for 16 h and then combined at room temperature with P36, or P36 heated and added PHF8-PHD at room temperature the complex formation is slight, an asymmetrical shoulder on the peak of unbound P36, its size essentially the same as without heating.

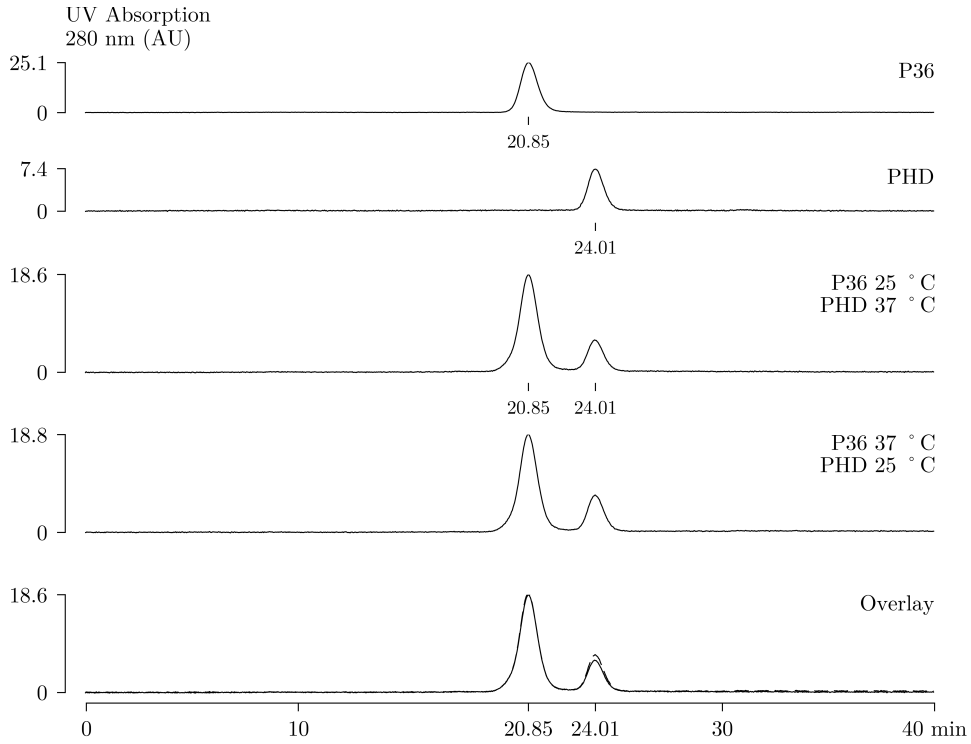


Figure 5.21: Size exclusion chromatograms of PHF8-PHD incubated for 16 h *at* 37 °C *and then combined with P36 which was incubated at* 25 °C*, and P36 incubated at* 37 °C *for* 16 h *and combined with PHF8-PHD which was incubated for* 16 h *at* 25 °C*. Overlaying the two chromatograms shows there is essential no different in the small amount of complex formed.*

That the effects of heat on PHF8-PHD appear reversible, but also assist complex formation, implies that PHF8-PHD becoming recognisable to antibodies at lower temperatures is also reversible, a product of sampling structural states whose energy barriers are (irreversibly) lowered by immobilisation on plastic or adding EDTA, or simply more easily approached at higher temperatures.

5.3 Conclusions and Future Work

The antibodies P2, P3, P23, and P36 have at least nanomolar affinity for PHF8-PHD, but in solution, at concentrations exceeding 20 µM, almost no complex forms. All antibody molecules appear capable of complex formation. As do all molecules of PHF8-PHD – but only a small fraction is in

a form that can bind immediately. On the time scale of SEC, SPR, and ELISA this fraction seems fixed, but the complex of P36 and PHF8-PHD accumulates much more slowly than its on-rate should allow, and therefore implies a progression of PHF8-PHD from a form unrecognised by P36 to one recognised.

Though PHF8-PHD seems capable of sampling such states reversibly, immobilisation on plastic or treatment with EDTA can help binding. Both likely disrupt PHF8-PHD's secondary structure – in fact, EDTA is one of the few ways to disrupt PHF8-PHD, a metalloprotein, in solution while ensuring the antibody remains unaffected. Partial denaturation may allow PHF8-PHD to better sample, or more easily adopt the conformations recognised by the antibodies here. In addition to conformational selection, recognition of PHF8-PHD appears to involve conformational changes on binding, implied by changes in off-rate when certain pairs of antibodies bind together – perhaps further evidence of conformational freedom in the recognition of PHF8-PHD.

It has been known for some time that adsorption of protein onto polymer surfaces disrupts structure, and it follows that antibodies screened on or isolated using antigens immobilised in this way may not bind in solution.[11] Indeed, the phenomenon has even been reported for a lysine demethylase, KDM5B.[12] What is implied here, however, is that some of these antibodies may be capable of capturing antigens as they briefly adopt a paritally denatured state during structural sampling. The ability of antibodies to stabilise various antigen conformations is often exploited,[13] but what there is little precedent for is stabilisation of such rarely sampled states. Perhaps the closest example is the stabilisation of viral proteins caught while "breathing".[14,15] However, whereas viral breathing reflects the motions of protein subunits exposing clefts in the complex multi-subunit structure of a viral envelope, the motion implied here is a subtle, probably local fluctuation, likely inherent in most proteins as they sample their various structural minima.

Though the slow time scale of binding between these antibodies and PHF8-PHD frustrates many common experimental approaches, isothermal titration calorimetry (ITC) could potentially yield beneficial information. If PHF8-PHD was supplied at high enough molar excess, there may be enough in a state appropriate to record binding isotherms, and the calculated stoichiometry, no doubt far from the 1:1 stoichiometry implied by SPR, would report the fraction of PHF8-PHD in an appropriate conformation. Combined with detailed binding timecourses measured by SEC, then at least some details of PHF8-PHD's structural equilibria could be determined. If the timescale of these motions was appropriate, NMR relaxation studies could help to rationalise these dynamics structurally. A complementary structural approach would be to forgo the complications of using intact PHF8-PHD, and instead attempt to crystallise antibody and epitope-derived peptide. Given the evidence that these are linear epitopes, peptides may be a reasonable substitute for PHF8-PHD, greatly simplifying subsequent investigation.

Even crystal structures of peptide-bound antibodies, or NMR studies of antigen dynamics will, however, need to be complemented by computation if they are to yield more than superficial insight. Most NMR studies only imply dynamics and must be fit into models to be practically useful.[16] And an antibody-bound peptide would need to be mapped on to the intact antigen's structure, or perhaps used to restrain molecular dynamics simulations of a complex between the antibody and intact antigen. But even a crystal or NMR of structure of intact PHF8-PHD with an antibody still avoids the details of how PHF8-PHD actually reaches its rare structural state. For this, currently computation, through molecular dynamics and, for instance, Markov state models of protein ensembles,[17] combined with experimental restraints, offers the most direct method for filling in PHF8-PHD's "invisible" states.

The kinetics of P36 are favourable enough to reflect conformational sampling by PHF8-PHD in the formation of antibody-antigen complexes. P23

and P2 have weaker affinity and require induced changes in PHF8-PHD, through EDTA, to admit their dependence on this sampling. P3 admits little besides binding simultaneously PHF8-PHD with P36 and P23. But to do so, P3 must recognise a similar rarely sampled conformation. Thus, these four antibodies derive their poor solution binding from specificity for very similar rarely sampled states of PHF8-PHD. Likewise, the many antibodies specific for PHF8's JmjC domain obtained from the same immunisation, by being similarly affected by poor solution binding and a dependence on antigen concentration are likely specific for rarely sampled structural states. Indeed, the similarity of partially denatured proteins and certain rarely sampled conformation may apply generally to proteins, and the many antibodies that struggle to bind outside ELISA.

References

[1] J. R. Horton et al. "Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases". In: *Nature structural & molecular biology* 17.1 (2010), pp. 38–43.

[2] GE Healthcare. "Biacore assay handbook". In: *Bio-Sciences AB, Uppsala, Sweden Google Scholar* (2012).

[3] V. Nicolaes et al. "Insights into the function of YciM, a heat shock membrane protein required to maintain envelope integrity in Escherichia coli". In: *Journal of bacteriology* 196.2 (2014), pp. 300–309.

[4] U. Jakob, M. Eser, and J. C. Bardwell. "Redox switch of hsp33 has a novel zinc-binding motif". In: *Journal of Biological Chemistry* 275.49 (2000), pp. 38302–38310.

[5] A. Von Arnim and X.-W. Deng. "Ring finger motif of Arabidopsis thaliana COP1 defines a new class of zinc-binding domain." In: *Journal of Biological Chemistry* 268.26 (1993), pp. 19626–19631.

[6] D. P. Giedroc et al. "Gene 32 protein, the single-stranded DNA binding protein from bacteriophage T4, is a zinc metalloprotein". In: *Proceedings of the National Academy of Sciences* 83.22 (1986), pp. 8452–8456.

[7] D. Sharma and K. Rajarathnam. "13C NMR chemical shifts can predict disulfide bond formation". In: *Journal of biomolecular NMR* 18.2 (2000), pp. 165–171.

[8] O. A. Martin et al. "Analysis of ^{13}Cα and ^{13}Cβ chemical shifts of cysteine and cystine residues in proteins: a quantum chemical approach". In: *Journal of biomolecular NMR* 46.3 (2010), pp. 217–225.

[9] G. J. Kornhaber et al. "Identification of zinc-ligated cysteine residues based on 13 Cα and 13 Cβ chemical shift data". In: *Journal of biomolecular NMR* 34.4 (2006), pp. 259–269.

[10] E. Martell. *Chemistry of the metal chelate compounds*. Prentice-Hall, Inc; New Jersy, 1952.

[11] C. Schwab and H. R. Bosshard. "Caveats for the use of surface-adsorbed protein antigen to test the specificity of antibodies". In: *Journal of immunological methods* 147.1 (1992), pp. 125–134.

[12] A. Wiuf et al. "Structure and binding properties of a cameloid nanobody raised against KDM5B". In: *Acta Crystallographica Section F: Structural Biology Communications* 71.10 (2015), pp. 1235–1241.

[13] C. Nizak et al. "Recombinant antibodies to the small GTPase Rab6 as conformation sensors". In: *Science* 300.5621 (2003), pp. 984–987.

[14] D. Haslwanter et al. "A novel mechanism of antibody-mediated enhancement of flavivirus infection". In: *PLoS Pathogens* 13.9 (2017), e1006643.

[15] K. A. Dowd et al. "A dynamic landscape for antibody binding modulates antibody-mediated neutralization of West Nile virus". In: *PLoS pathogens* 7.6 (2011), e1002111.

[16] A. J. Baldwin and L. E. Kay. "NMR spectroscopy brings invisible protein states into focus". In: *Nature chemical biology* 5.11 (2009), p. 808.

[17] J. D. Chodera and F. Noé. "Markov state models of biomolecular conformational dynamics". In: *Current opinion in structural biology* 25 (2014), pp. 135–144.

6 Conclusions and Future Work

That a protein never finishes folding and instead migrates endlessly among energy minima is often lost in drug discovery. Unless a protein is intrinsically disordered, its structural variation is usually distributed so narrowly around so few average conformations that for most purposes they can approximate the whole population. But this approximation undersells minor states of biological importance. Transient though minor states may be, they can represent the discrete steps in a protein's motion essential for function. It follows that stabilising such transient states may prove an attractive strategy for allosteric inhibition.

In the preceding chapters, generating antibodies for KDM4D and PHF8 has unexpectedly yielded significant numbers of antibodies selective for transient antigen states. Though the conditions that bore them are unknown, their properties support two simple hypotheses on their origin, both of which have potential application in developing novel therapeutic antibodies. How I have concluded this from testing the hypothesis that antibodies can, through the versatility of their binding, inspire small-molecule drug-discovery, requires some explanation. This work has revealed antibodies that manifest what is likely the first example of allosteric inhibition of KDM4D, and succeed where small molecules have struggled in blocking substrate recognition by the PHD of PHF8. However, these results have been possible not so much because of the binding versatility of antibodies, but in spite of it. Though it is a familiar trade-off in antibody discovery that a wider repertoire of interactions more readily finds distraction in epitopes irrelevant to function, here large numbers were lost to epitopes that didn't even seem always accessible. They would bind when their antigen was immobilised, but hardly at all when it was free in solution. Such contradictory observations are often dismissed as artefacts. But, the observations here were not

explained by 'weak' antibodies that exploit the avidity invited by antigen immobilised in closer proximity than they exist in solution, or an epitope being exposed by the partial denaturation that accompanies immobilisation on plastic. These antibodies genuinely did bind. But their epitope was transient; so versatile was their binding that they could favour sites only ever exposed in a fraction of the structural population.

Much still prevents antibodies from inspiring the design of small-molecule drugs. So the attraction of antibodies selective for transient states is that if generated for extracellular proteins they could have ready application to therapy. Essential to these application is understanding what has caused so many antibodies from a routine immunisation to be selective for transient protein states. Any hypothesis on their origin must account for five factors that characterise the antibody responses for KDM4D and PHF8 reported here:

1. Structural Integrity

   The protein that was used for immunisation was enzymatically active, and therefore contained at least some folded protein molecules.

2. Dominant Specificity

   Antibodies that appear to recognise dominant antigen states were part of the immune response (e.g. J11 and P13).

3. Transient Specificity

   At least a subset of the recovered antibodies recognise transient antigen states (e.g. P2 and P36).

4. Antigen Independence

   The behaviour indicting a preference for transient protein states was independent of both protein (KDM4D and PHF8), and domain (PHD and JmjC of PHF8.

5. Binding Cooperativity

   Between certain pairs of antibodies cooperative binding is observed

(e.g. P23 and P2, P36 and P2).

Specificity for transient antigen states may be by having superior immunogenicity or changes to the antigen's structure after immunisation exposing sites only exposed in transient states. Determining to what extent these are responsible requires knowing how the antigen has been presented to the immune system – itself dependent on the unknown effect of immunisation on the antigen's structural distribution. The following two hypotheses on the origins of transient state specificity take as their axiom the structure of the antigen after immunisation, specifically whether its form *in vivo* resembles that *in vitro*.

The Immunogenicity Hypothesis. *The transient structural states of KDM4D and PHF8 are more immunogenic than their dominant states, and hence an immune response is directed primarily towards transient states.*

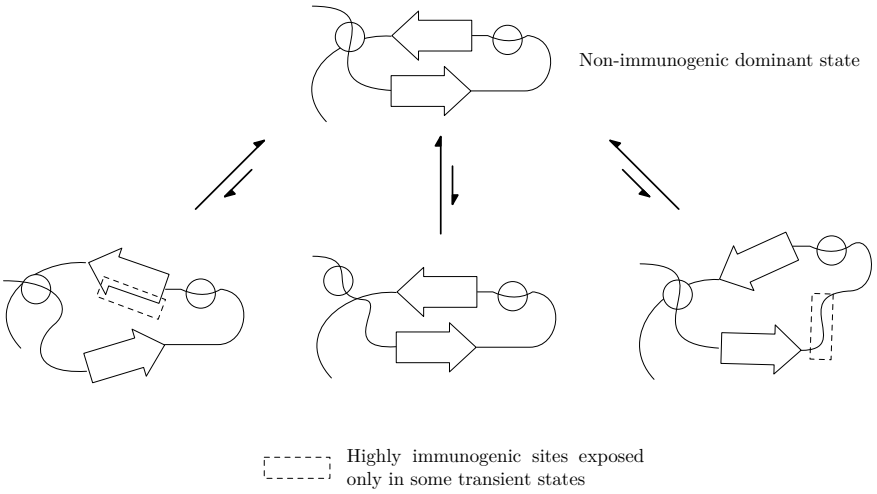


Figure 6.1: The immunogenicity hypothesis predicts that some sites exposed in the antigen's transient states are so highly immunogenic that despite representing only a fraction of the structural population, they elicit the strongest immune response.

If the antigen maintains its *in vitro* structural distribution after immunisation, then a minority of the structural population has received the majority of an immune response; which is to say, it is more immunogenic. The dominant state's similarity to the host has bought it protection from an immune response, but is lost when certain sites are exposed in rarely sampled transient states. Without competition from the dominant state, antibody generation is skewed disproportionately towards the transient state. By avoiding immune clearance, the dominant state is able to maintain continuous supply of the transient state through its structural equilibrium. For cooperativity to be observed between antibodies, there would need to be a structural relationship between their epitopes. Either the epitopes are both exposed in a common state, or stabilisation of one causes the other to be exposed more often. But most importantly, for the hypothesis to hold KDM4D and PHF8 must be similarly non-immunogenic in their dominant states. The similarity of the immunised proteins to their leporine counter-

parts is supportive of this condition, though immunogenicity is, of course, notoriously difficult to predict. If true it would suggest that what differentiates self from non-self immunologically can be as subtle as short-lived structural variation. And epitopes that would solicit an immune response are present in otherwise innocuous proteins, protected by only ever being exposed in a fraction of an already tiny native population. Immunisation saw this fraction grow larger than the threshold below which it is usually protected, and therefore stimulate and immune response. As a fail-safe mechanism it could have an evolutionary advantage: immune clearance of any native protein whose fold is aberrant at a sufficient concentration. It would also allow protein (mis)folding to be a cause of autoimmune diseases, and extend the hypothesis's consequences beyond just therapeutic antibody discovery.

The Sampling Hypothesis. *Antibodies raised against partially, or fully denatured KDM4D or PHF8 can recognise transient states of a folded antigen by exploiting epitopes exposed in the denatured and transient state, but not in the dominant state.*

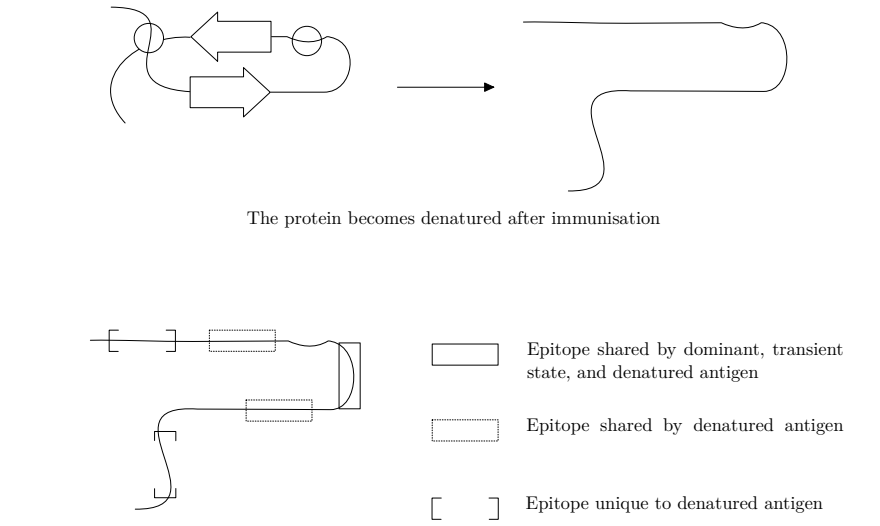


Figure 6.2: Under the structural hypothesis, the specificity of some antibodies for transient states comes from denaturation of their antigen after immunisation exposing sites unavailable in the antigen's dominant state, but available in rarely sampled transient states.

An antigen's structural distribution could be altered by immunisation. Were it to favour states that expose usually hidden antigen sites, their becoming antibody epitopes would logically follow. It is tempting to cite a

reverse in populations: the dominant state before immunisation becomes a minority after, and vice versa. But even at elevated temperatures the structural distribution of PHF8-PHD was resistant to changes dramatic enough to explain antibodies having such a strong preference for its transient states. This leaves only more permanent changes. Denaturation or proteolysis are both risks for an immunised protein that could expose sites usually unavailable. If antibodies were raised to an antigen adulterated in this way it is not surprising they would then struggle to bind the folded protein. They would have to exploit the relatively few epitopes shared by the folded and denatured protein. This specificity for transient antigen states suggests that through the antigen's natural dynamics many more epitopes may, however, be exposed. Although too fleeting to solicit the generation of antibodies when folded, the epitopes in transient states are exposed just frequently enough for antibodies raised against a similar epitope in the denatured protein to register binding. Cooperativity would imply there is even a degree of structural coordination between some epitopes exposed by these transient states. It is perhaps significant that both KDM4D and PHF8 were affected. Without there being any evidence that they are more dynamic than other proteins it could suggest these dynamics are common and thus easily exploited by antibodies. Though the computation evaluation of a protein's full motions is currently intractable, even a conservative interpretation of the hypothesis could mean a folded protein exposes sites shared with its (partially) denatured form regularly.

There are at least two reasons why determining which hypothesis is correct is non-trivial. First, it is likely impossible to assure that the antigen's structural distribution is maintained after immunisation. Even slight changes could influence antibody generation by exposing usually hidden antigen sites. Second, both hypotheses can simultaneously be true. Structural changes that present rare epitopes don't prevent the presence of highly immunogenic but rarely exposed sites, and such structural changes could even help to boost their relative immunogenicity by making them more

prominent in the structural population. The following three tests are proposed as the minimum necessary to deconvolute the contribution of immunogenicity and sampling to the generation of antibodies specific for transient states.

1. Immunisation of the denatured antigen

While it may be difficult in many cases to guarantee the structural distribution of a folded protein after immunisation, a denatured one is unlikely to refold. Under the sampling hypothesis denaturing the antigen before immunisation should not effect the antibodies that are generated. Denaturation having no effect on, or preventing antibodies being generated against transient states, is evidence for and against the sampling hypothesis, respectively.

2. Immunisation of a mutated antigen

The transient state immunogenicity hypothesis relies on the transient antigen states exposing highly immunogenicity of sites. A test of this hypothesis needs to prove their exposure is due to structural sampling of the folded protein and not simply unfolding. It must be designed so that temporary exposure of immunogenic sites in transient states yields different antibodies than constant exposure with the denatured protein. A transient state's relative stability is often due to only a few favourable intramolecular interactions which when mutated effectively removes the state from the structural population. A destabilising mutation far from the epitope of interest (to make antibodies responding to immunogenicity introduced by the mutation itself obvious) could be used to hide an immunogenic site only in the *folded* protein. Thus, if the immunogenicity hypothesis holds, no antibodies should be generated to the putative immunogenic site, but if antibodies are generated then the antigen must have undergone a structural change after immunisation that exposes a site hidden in the folded protein.

3. Epitope analysis

Both the immunogenicity and structural hypothesis could also be special

cases of a more general hypothesis: the epitopes exposed in transient protein states are indeed more immunogenic than those exposed in the dominant state, but not so immunogenic they can generate antibodies without permanent structural change exposing them. Immunisation of denatured protein (test 1) would produce antibodies against transient states, as would immunisation with an antigen mutated to hide certain epitopes (test 2). But this result could be indistinguishable from the two hypotheses acting concurrently but independently. That is, some degree of denaturation exposing a fraction of epitopes and immunogenicity and sampling of genuine states exposing others. Unless every immunogenic epitope is subjected to test 2, which may be impractical, the only indication of whether each hypothesis acts independently or dependently, is in the distribution of antibody epitopes. The sampling hypothesis sees antibodies generated without bias against the entire denatured antigen. Those that bind the folded protein are presumably only a subset. But under the immunogenicity hypothesis antibodies binding the folded protein are everything, and their epitopes concentrated among a few immunogenic sites. Of course all epitopes are necessarily immunogenic, and there is some subjectivity in discerning between these two cases. Still, both hypotheses operating independently should produced antibodies with epitopes throughout the protein, whereas dependence between the hypotheses will narrow the distribution of these epitopes to obviously fewer sites, and likely see the diverse antibodies targeting the same sites.

The form antibody discovery for transient states takes follows from the extent to which an antigen obeys either of the proposed hypotheses. Under the immunogenicity hypothesis, temporary exposure of highly immunogenic sites makes some of an antigen's transient states more immunogenic than its dominant states. Generating antibodies against such transient states would simply require immunisation with enough antigen for the fraction of immunogenic transient states to provoke an immune response. Were protected immunogenic sites contained in most apparently non-immunogenic

proteins, they would all be potential targets of the technique. In fact, it would be one of the few ways to engage protein such protein with antibodies. Under the sampling hypothesis generating antibodies for transient states would only require immunisation of the denatured protein. The condition is that the transient states must expose sites hidden by the dominant state, but available in the denatured protein. And it is a property that could easily be shared by other proteins.

Only raising antibodies to the transient states of different proteins will show if stabilising such states will result in function-modification frequently enough to be a practical method of drug discovery. The relationship between transient states and the protein dynamics that govern catalysis and molecular recognition is, however, hopeful at least for allosteric inhibition. In fact, inhibition may have already been observed. Among anti-PHF8 antibodies that manifest behaviour characteristic of transient state specificity were some that also appeared to inhibit catalysis. At the time it was discounted. Enzymatic assays are fickle, and inhibition found in artefacts often more readily than specific interactions. However, that 'disingenuous' binding has been proved genuine could mean this observation of inhibition was too. For antibodies that only recognise a fraction of the antigen population to also be inhibitors only requires that that fraction be responsible for catalysis. The antibody must recognise a state also sampled in the dynamics necessary for catalysis, and by stalling this motion act as an inhibitor. Thus though they appear to engage only a minority of the structural population these antibodies could still be effective inhibitors, and such inhibitors may already exist.

The antibody-antigen interaction is usually thought of as an equilibrium of association and dissociation, but is really a coupling of equilibria: association and dissociation of an antibody-antigen complex is between two molecules themselves in complex structural equilibria. It becomes obvious when the antigen is a transient structural state. Association needs not just a chance collision, but also one that has reached the appropriate confor-

mation. And dissociation is not just a loss of complex, but the antigen's return to a dominant state – becoming, once again, unrecognisable to its antibody. The complex formed in equilibrium depends on the frequency of the antigen's sampling being matched with the affinity of antibodies specific for it. This is what gives the antibodies here their contradictory behaviour. In the few hours they were incubated, just enough complex formed for assays that amplify the binding signal (such as ELISA) to report binding, and those that do not (such as SEC) to apparently report none. Partial unfolding of proteins immobilised on plastic can also cause exactly this assay behaviour. Binding is only observed when the antigen is immobilised on a surface and, so the interpretation goes, a non-native epitope is exposed. I wonder if given the time, how many of these antibodies would show, like the antibodies here, that what masquerades as selectivity for denatured protein is actually a genuine preference for transient states.

And perhaps therein lies the greater lesson of this work for drug discovery. The observation I have presented here, potentially of general significance to drug discovery, was almost lost to the approximation that proteins are essentially static structures. These approximations persist because the underlying science of drug discovery, particularly molecular recognition, is not understood well enough to meet the relentless demand for therapeutics. It is a demand that has been met largely through technological advance based on empirical observation, rather than scientific understanding. Though it would be irresponsible to sacrifice therapeutic discovery for purely scientific investigation, a more sustainable model for drug discovery may be regularly turning its own technological advances to examining the limits in its scientific foundations. The results here came from antibody discovery on targets that would never make realistic therapeutic targets, and for a cause obviously beyond drug discovery's current technical reach. And yet this work has demonstrated that the manipulation of protein dynamics often neglected in drug discovery presents new therapeutic applications for antibodies, and the opportunity to better understand the scientific foundations

that constrain rational drug design.

7 Materials and Methods

7.1 Subcloning

Avi-tag

PCR was performed using Phusion High-Fidelity DNA Polymerase (New England BioLabs) with Phusion GC buffer and 1 % (v/v) DMSO. Denaturing, annealing, and elongation conditions were 95 °C (30 s), 63 °C (30 s), and 72 °C (1 min) respectively, repeated for 30 cycles. The PCR products were purified using a PCR purification kit (Qiagen) according to the manufacturer's instructions. The pNIC-Bio3 (SGC, Oxford) vector was digested at room temperature using BsaI (New England BioLabs). Ligation independent cloning (LIC) sites were generated by incubating the PCR product for 30 min at 22 °C with T4 DNA polymerase (New England BioLabs) and dCTP, and for the digested plasmid using the same protocol but replacing dCTP with dGTP. The "ligation" step was performed by incubating the PCR products and digested vector at room temperature for 30 min. The primers used for KDM4D were:

FWD	`5'-TACTTCCAATCCATGGAAACTATGAAGTCT-3'`
REV	`5'-TATCCACCTTTACTGCTCAGCTCTTGGCTGGCTGGTACC-3'`

And the primers used for PHF8 were:

FWD	`5'-TACTTCCAATCCATGAGCGATAAAATTATTCAC-3'`
REV	`5'-TATCCACCTTTACTGCTGTTCTGTTGGAAGATGTC-3'`

Cloning PHF8-PHD Constructs

Constructs of PHF8-PHD were cloned from a plasmid encoding the PHF8 immunisation construct using the LIC procedure detailed above but substituting a N-terminal his-tag thioredoxin vector (pNH-TrxT).

7.2 Protein Purification

Tranformation by Heat Shock

Approximately 25 µL to 50 µL of an appropriate strain of chemically competent bacterial cells was thawed on ice and 0.5 µL to 2 µL of a solution containing the plasmid to be transformed added and incubated with the cells on ice for 30 min. The cells were then incubated in a 42° water bath for 45 s before being put back on ice directly. After 2 min, up to 50 µL of the cells was spread across the suface of LB agar containing an appropriate antibiotic. The plates were inverted and incubated at 37 °C overnight.

Bacterial Culture

BL21 (DE3) cells were transformed by heatshock with an appropriate plasmid and cultured overnight at 37 °C on LB agar supplemented with a kanamycin at 50 µg mL^{-1}. A single colony was used to innoculate LB also supplemented with kanamycin and incubated overnight at 37 °C with shaking at 125 rpm. 10 mL of the stationary phase culture was used to innoculate each litre of LB or terrific broth (TB), again supplemented with kanamycin. The cultures were incubated at 37 °C with shaking at 125 rpm until the OD_{600} value reached 0.4 for LB and 2 for TB. They were then cooled to 20 °C and shaking reduced to 70 rpm to 120 rpm. Once the temperature stabilised (when OD_{600} was approximately 0.6 for LB or 3 for TB) the cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG), added to a final concentration of 0.5 mM. The next day, after approximately 16 h, the cultures were centrifuged at 4500 rpm for 10 min in a floor standing

centrifugre (Beckman). After discarding the supernatant the bacterial cell pellet was transferred to a plastic bag, and stored at −20 °C.

Amendments

1. *$PHF8_{1–447}$ was cooled to* 16 °C *during expression.*

Culture for Biotinylation

An appropriate vectors was transformed into BL21 (DE3) cells previously transformed with a BirA (biotin ligase) encoding plasmid and starter cultures grown at 37 °C overnight in LB supplemented with 50 µg mL^{-1} kanamycin, 50 µg mL^{-1} ampicillin, and 50 µg mL^{-1} streptomycin. These were transferred to 1 L LB cultures containing kanamycin and streptomycin. Cells were grown at 37 °C until the OD_{600} reached 0.6 at which time 10 mM biotin in 10 mM bicine (pH 8.0) was added to a final biotin concentration of 500 µM. The cells were then induced, incubated, and harvested as already described.

Purification

Cell pellets were resuspended in lysis buffer (50 mM HEPES, 250 mM NaCl, 5 % (v/v) glycerol, and 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.4) and lysed by several passes through a constant cell disrupter operating at a pressure of 1500 kPSI. The cell lysate was centrifuged at 15 000 rpm in a floor standing centrifuge (Beckman, rotor: JLA16.1) at 4 °C for at least 40 min. The lysate was then passed through nickel-(nitrilotriacetic acid) (Ni-NTA) resin (GE Healthcare) using a gravity flow column, and the resin washed in approximately 10 column of lysis buffer supplemented with immidazole to a concentration of 10 mM to 30 mM. The his-tagged protein was eluted in 10 mM HEPES, 250 mM NaCl, 5 % (v/v) glycerol, 0.5 mM TCEP, and 300 mM Imidazole, pH 7.4. The his-tag was cleaved by overnight incubation with TEV protease at 4 °C while the 50 mL of

eluted protein was dialysed in gel filtration buffer (10 mM HEPES, 250 mM NaCl, 5 % (v/v) glycerol, and 0.5 mM TCEP, pH 7.4) using 3.5 kDa cut-off SnakeSkin Dialysis Tubing (ThermoFisher Scientific). The cleaved his-tag, uncleaved protein, and any background retained through washing was removed by passing the dialysed protein through a 5 mL Ni-NTA "his-trap" column (GE Healthcare) on an ÄKTA prime (GE Healthcare). The flow-through was concentrated and further purified by gel-filtration using an ÄKTA Express with, according to molecular weight, either a SuperDex 16/60 S200 or S75 column. Aliquots of 100 µL were flash frozen with liquid nitrogen or a mixture of ethanol and dry ice.

Amendments

1. *TCEP was left out of the gel filtration buffer for PHF8-PHD and the PHD of PHF2 and KDM7A*
2. *TCEP was also left out of buffers of any proteins that were to be used with antibodies*
3. *The rebinding step was repeated twice for PHF2's PHD*

7.3 Immunisation Constructs

KDM4D

SMETMKSKANCAQNPNCNIMIFHPTKEEFNDFDKYIAYMESQGAHRAGLAKIIPPKEWKARE
TYDNISEILIATPLQQVASGRAGVFTQYHKKKKAMTVGEYRHLANSKKYQTPPHQNFEDLER
KYWKNRIYNSPIYGADISGSLFDENTKQWNLGHLGTIQDLLEKECGVVIEGVNTPYLYFGMW
KTTFAWHTEDMDLYSINYLHLGEPKTWYVVPPEHGQRLERLARELFPGSSRGCGAFLRHKVA
LISPTVLKENGIPFNRITQEAGEFMVTFPYGYHAGFNHGFNCAEAINFATPRWIDYGKMASQ
CSCGEARVTFSMDAFVRILQPERYDLWKRGQDRAVVDHMEPRVPASQEL

PHF8

SMASVPVYCLCRLPYDVTRFMIECDMCQDWFHGSCVGVEEEKAADIDLYHCPNCEVLHGPSI
MKKRRGSSKGHDTHKGKPVKTGSPTFVRELRSRTFDSSDEVILKPTGNQLTVEFLEENSFSV
PILVLKKDGLGMTLPSPSFTVRDVEHYVGSDKEIDVIDVTRQADCKMKLGDFVKYYYSGKRE
KVLNVISLEFSDTRLSNLVETPKIVRKLSWVENLWPEECVFERPNVQKYCLMSVRDSYTDFH
IDFGGTSVWYHVLKGEKIFYLIRPTNANLTLFECWSSSSNQNEMFFGDQVDKCYKCSVKQGQ
TLFIPTGWIHAVLTPVDCLAFGGNFLHSLNIEMQLKAYEIEKRLSTADLFRFPNFETICWYV
GKHILDIFRGLRENRRHPASYLVHGGKALNLAFRAWTRKEALPDHEDEIPETVRTVQLIKDL
AREIRLVEDIFQQN

7.4 Antibody Recovery

B-Cell Culture

B-cells isolated from rabbit lymph nodes were cultured in RPMI 1640 medium (ThermoFischer Sciences) supplemented with 8.7 % (v/v) fetal bovine serum (Sigma-Aldrich), 1.7 % (v/v) HEPES buffer 1 M stock solution, pH 7.4 (Sigma-Aldrich), 1.7 % (v/v) L-glutamine 200 mM solution (ThermoFischer Scientific), 0.86 % (v/v) penicillin-streptomycin solution (ThermoFischer Scientific), 0.17 % (v/v) normacin 50 mg mL^{-1} solution (Invivogen), and 0.09 % (v/v) 2-mercaptoethanol (55 mM in Dulbecco's phosphate buffered saline (DPBS)) (Gibco). This media, filtered and pre-incubated at 37 °C, was used to re-suspend frozen rabbit B-cells. The cells were cultred in 96-well plates, in a volume of 200 µL, and a density of approximately 5×10^3 cells per well, with an additional 5×10^4 EL4 feeder cells per well. The cells were cultured for 5 d to 7 d at a temperature of 37 °C and in an atmosphere of 5 % CO_2.

FMAT

50 µL of Superavidin beads (Pierce) per assay plate were centrifuged at 10 000 rpm in a benchtop centrifuge for 3 min and resuspended in 4 mL of PEG blocking buffer (PBS containing 1 % (w/v) PEG 4000). 1 µg of protein

per assay plate was added to the streptavidin beads and incubated with gentle inversion for 1 h at room temperature. The supernatant was then removed by centrifugation at 13 000 rpm for 30 s and the beads resuspended in 2 mL of PEG block per assay plate. Dylight-649 conjugated anti-Rabbit IgG (Jackson ImmunoResearch) was added such that the final dilution in each well would be 1:5000 (2.4 µL per plate) and the container wrapped in aluminium foil. 20 µL of the assay mixture was dispensed into each well of a 96-well plate containining 20 µL of B-cell supernatant. After incubating the plates for 1.5 h in darkness, the signal level was read using a MirrorBall FMAT plate reader (TTP LabTech).

Fluorescent Foci Assay

15 µL of Streptavidin coated magnetc beads (New England Biolabs) per sample were washed using a magnetic tube rack, bt resuspension three times in the B-cell media previously described. 1 µg of biotinylated protein was added for each sample, and the two were incubated at room temperature with gentle inversion for 1 h. The beads were then washed again three times in B-cell media and finally resuspended in 120 µL of B-cell media per sample.

Frozen B-cells were recovered from 96-well culture plate by adding 300 µL of B-cell media pre-warmed to 37 °C to mix and transfer the contents to another vessel. This was repeated twice more to collect all material. The samples were then centrifuged at 2000 rpm for 2 min, turned through 180° in so the inside wall of each vessel now faced outside, and centrifuged again at 2000 rpm for 2 min. The media was removed by pipette until approximately 60 µL remained.

For every sample, 120 µL of the previously prepared bead mixture was combined with the 60 µL B-cell suspension, and 60 µL of FITC-conjugated anti-rabbit IgG antibody (Jackson ImmunoResearch) at a 1:300 dilution in B-cell media. Approximately 90 µL was spotted on to glass microscope slides coated in SigmaCote (Sigma-Aldrich), and then covered in mineral oil (Sigma Aldrich) such that there was no exposure to air and no chance of

evaporation during incubation. The slides were the incubated at 37 °C for 60 min and cells isolated by micromanipulator according to the procedure described in the text.

Reverse Transcriptase PCR

Reverse transcriptase PCR was performed using a SuperScript IV kit (ThermoFisher Scientific). In a biosafety cabinet, single B-cells frozen in approximately 0.5 µL of media were thawed by adding 20 µL of reaction mixture (10 µL DEPC-treated water, 4 µL 5× SSIV buffer, 1 µL 100 mM DTT, 1 µL 50 µM oligo $d(T)_{20}$, 1 µL 10 % NP-40, 1 µL 10 mM dNTP mix (10 mM each), 1 µL ribonuclease inhibitor 40 U μL^{-1}, 1 µL SuperScript IV (ThermoFischer Scientific)). The samples were incubated in a PCR block at 50 °C for 10 min, and then 80 °C for 10 min to inactivate the enzyme. The mixture was then stored at −20 °C or used directly in the next step.

Antibody Recovery

Heavy and light chain variable regions were recovered separately through two rounds of PCR using the RT-PCR reaction mixture from the previous step. The primary PCR mixture contained 1 µL of the previous RT-PCR reaction mix, 10.5 µL DPEC-treated water (ThermoFischer), 12.5 µL KOD Hotstart Master mix (Merck), and 1 µL of a stock (for the heavy chain amplifications) containing the following primers at the concentrations indicated:

Heavy	FWD	2.5 µM	`5'-AAGCTTACGCTCACCATGGAGACTGGGCTGCGCTGG-3'`
Heavy	REV	2.5 µM	`CAGGTCACGGTCACTGGCTC`
Heavy	REV	2.5 µM	`GCCYTCTAGATGMMTGCT`

and the following primers for the light chain reaction:

Light	FWD	2.5 µM	`5'-CGCGCGAAGCTTCGAAGCCACCATGGACAYGAGGGCCCCCACTCAG-3'`
Light	FWD	0.25 µM	`CGCGCGAAGCTTCGAAGCCACCATGAACAYGAGGGCCCCCACTCAG`
Light	REV	2.5 µM	`CGCCACACACACACGATGGTGACTG`

The PCR reaction conditions were: 96 °C (2 min), 40 cycles of 96 °C (15 s), 55 °C (15 s), 68 °C (1 min), and finally 68 °C (5 min) followed by 4 °C until collection with a 105 °C lid temperature maintained throughout.

1 µL of the of this reaction mixture was combined with 9 µL DEPC-treated water (ThermoFischer), 12.5 µL KOD Hotstart Master mix (Merck), and 2.5 µM of a stock solution containing, for the heavy chain, the following primers at the concentrations indicated:

Heavy	FWD	2.5 µM	5'-CTGCAGTCACCGTCCTTGACACGAAGCTTCGAAGCCACCATGGAGACTGGGC-3'
Heavy	REV	2.5 µM	ACTGATGGAGCCTTAGGTTGCCCACTCGAGACGGTGACCAGGGTGCC
Heavy	REV	0.25 µM	ACTGATGGAGCCTTAGGTTGCCCACTCGAGACGGTGACCAGGGTVCC

and the following primers for the light chain amplification:

Light	FWD	2.5 µM	5'-GTCTTTTCTGCAGTCACCGTCCTTGACACGAAGCTTCGAAGCCACCATG-3'
Light	REV	2.5 µM	ATGAGGACAGTAGGTGCAACTGGCGTACGTTTGACCACCACCTCGGTCCCTC
Light	REV	0.25 µM	ATGAGGACAGTAGGTGCAACTGGCGTACGTTYGACCACCACCTCGGTCCCTC
Light	REV	0.25 µM	ATGAGGACAGTAGGTGCAACTGGCGTACGBBYGACHACCAYCTCGGTCCCTC

The PCR conditions for the second heavy chain amplification were: 96 °C hold 2 min, 40 cycles of 96 °C (15 s), 60 °C (15 s), 68 °C (50 s), and finally 68 °C (3 min) followed by 10 °C until collection. And the PCR conditions for the second light chain PCR reaction were 96 °C hold 2 min, 40 cycles of 96 °C (15 s), 55 °C (15 s), 68 °C (50 s), and finally 68 °C (3 min) followed by 10 °C until collection. Samples were stored at −20 °C until needed.

Subcloning Antibody Variable Regions

Purified antibody heavy and light chain variable regions, amplified by PCR in the previous step, were digested using HindIII and XhoI (for heavy chains) and HindIII and BsiWI (for light chains) (New England BioLabs). The digested inserts were purified using an agarose gel, and ligated using T4 DNA ligase (New England Biolabs) into a mammalian expression vector such that the insert was preceded by an HCMV and followed by an appropriate rabbit

Fab constant region (CH-1 or CL-1) and his-tag. XL-1 Blue competent E. coli cells (Agilent) were then transformed by heat shock with 1 µL to 2 µL of the neat ligation mixture. Transformed cells were cultured on LB agar supplemented with the antibiotic kamamycin (Sigma), to which the plasmid confers resistence, at a concentration of 50 mg/mL. The transformed plasmid was then amplified by miniprep of single colonies and sequenced.

Recombinant Antibody Expression

Plasmids containing a heavy chain and light chain were mixed and used to transfect HEK293 cells (ExpiHEK cells, ThermoFisher Scientific), using an ExpiFectamine 293 transfection kit (ThermoFisher Scientific) according to the manufacturer's instructions. Following addition of the "enhancer" reagent, 12 h after transfection, the cells were incubated for 5 d to 7 d at 37 °C in an atmosphere of 5 % CO_2 and with constant shaking at 125 rpm. Cells were first transfected on a 30 mL scale, and later at 0.5 L to 1 L for structural studies.

Antibody Purification

HEK293 cells were separated from their antibody-containing supernatants by centrifugation at 5000 rpm for 1 h to 3 h in a floor standing centrifuge (Beckman). The supernatant was then filtered through a 0.2 µm filter (Merck) to remove cell debris. As the Fab constructs used here had C-terminal his-tags on the heavy chain, the supernatants were passed through a 5 mL Nickel Excel column (GE Healthcare) (or two connected in series) using an ÄKTA prime. Non-specifically binding proteins were removed by passing approximately 10 column volumes of PBS with the NaCl concentration adjusted to 500 mM through the column, followed by a similar volume of PBS containing approximately 5 µM Imidazole. The antibody was then eluted using PBS supplemented with imidazole to a concentration of 300 mM. Aggregates, and occasionally dimers were removed by gel filtration

with a SuperDex 16/60 S75 column (GE Healthcare) on an ÄKTA FPLC (GE Healthcare) with the buffer being simultaneously being exchanges in either PBS or HBS.

7.5 Analysis

ELISA

An antigen was immoblised directly on the surface of MaxiSorp ELISA plates (NUNC) by incubating 50 µL (96-well plates) or 10 µL (384-well plates) of an antigen at a concentration of 1 µg mL^{-1} to 10 µg mL^{-1} in carbonate coating buffer (28.6 mM $NaHCO_3$ and 71.4 mM Na_2CO_3, pH 9.6) for either 2 h to 4 h at a temperature of 37 °C or 16 h at room temperature. The plates were washed with PBS containing 0.1 % (v/v) Tween-20 (Sigma-Aldrich) using a micro plate washer (BioTek). For all subsequent incubation steps the solution volume was matched to that use for antigen immobilisation. Each well was blocked for at least 1 h with blocking buffer (PBS with 1 % (w/v) BSA). After washing, an antibody containing solution (B-cell supernatant, or recombinant Fab diluted in blocking buffer to a concentration of 10 µM), was incubated in each well for 1 h. The plates were washed and either horse radish peroxidase (HRP)-conjugated goat IgG anti-rabbit Fc fragment-specific antibody (Jackson ImmunoResearch) or HRP-conjugated goat IgG anti-rabbit F(ab')$_2$-specific antibody (Jackson ImmunoResearch) at a 1:5000 dilution in blocking buffer, was incubated in each well for 1 h. After washing 1-Step Ultra TMP-ELISA substrate solution (ThermoFisher Scientific) was added to each well, and after no more than 5 min incubation, the reaction between HRP and TMP was quenched with 100 µL of a 2 % (w/v) NaF solution.

Analytical Size Exclusion Chromatography

Samples were analysed by size exclusion chromatography using either a Superdex 200 10/300 Increase GL column or a Superdex 75 10/300 Increase GL column (GE Healthcare). The running buffer was either HBS or PBS with NaCl added to a final concentration of 250 mM to prevent non-specific interactions with column material. Flow rates were adjusted such that the experiment run time was 35 min to 45 min depending on retention times and column back pressure.

Affinity Determination by SPR

All SPR experiments were performed using a BIAcore 3000 (GE Healthcare), at a temperature of 25 °C, and with HBS containing 0.05 % (v/v) P20 surfactant (HBS-P) (GE Healthcare) as the running buffer (EDTA containing buffers were avoided as KDM4D and PHF8 are metelloproteins). Goat $F(ab')_2$ anti-rabbit Fab (Jackson Immunoresearch) was immobilised on the surface of a CM5 sensor chip (GE Healthcare) using an amine coupling kit (GE Healthcare) according to the manufacturer's instructions. The capture level was approximately 3000 RU to 4000 RU. For affinity measurements a Fab was first captured, and its antigen injected at varying concentrations at a flow rate of 30 µL min^{-1} to reduce mass-transport effects. The association and dissociation time were 5 min and 10 min, respectively. The Fab was captured such that the subsequent response after injection of the antigen remained less than 100 RU. The sensor surface was regenerated between runs by injecting 10 µL each of 40 mM HCl, 5 mM NaOH, and 40 mM HCl once again. Data were processed using the BIAeval software (GE Lifesciences), assuming a 1:1 Langmuir binding model and locally fitting R_{max} to account for differences in Fab capture between runs.

PHF8-PHD SPR Binding Assay

Streptavidin (Jackson ImmunoResearch) was immobilised on the surface of a CM5 sensor chip (GE Healthcare) using an amine coupling kit (GE Healthcare) to approximately 500 RU to 1000 RU. A biotinylated $H3_{1–15}$K4me3 (Peptide Protein Research) dissolved in deionised water was injected until the baseline had increased by 10 RU to 30 RU. Measurements of PHF8-PHD affinity were performed by injecting PHF8 diluted to various concentrations in HBS-P. Inhibition of the interaction between PHF8-PHD and $H3_{1–15}$-K4me3 peptide was judged by incubating and antibody and PHF8-PHD and then injecting the mixture.

Activity Assay

Demethylation of an $H3_{1–15}$K4me3 peptide (Peptide Protein Research) by KDM4D or an $H3_{1–21}$K4me3K9me2 peptide (GL Biochem) by PHF8 was monitored using an Agilent RapidFire High-Throughput Mass Spectrometry System connected to an Agilent Q-TOF (Target Discovery Institute, University of Oxford). The final assay concentrations for each reagent were: 2-oxyglutarate 10 µM, ferrous ammonium sulfate hexahydrate 50 µM, L-ascorbic acid 100 µM, peptide 10 µM and protein 300 nM. An enzyme solution containing 4× the final assay concentration of 2-oxyglutarate, ferrous ammonium sulfate hexahydrate, L-ascorbic acid, and peptide, and a protein solution of equal volume containing 4× the final concentration of protein was prepared separately in 50 mM MES buffer (pH 7.0). A 2× solution of Fab was incubated with the protein mixture for 30 min, and the reaction then started by adding the 4× substrate solution. The reaction was quenched by addition of 10 % (v/v) formic acid to a final concentration of 1 % (v/v). For PHF8 assays were carried out at 37 °C using an incubator, and for KDM4D all assays were carried out at room temperature.

PAR/PMPS Assay

PHF8-PHD was diluted to 5 µM to 10 µM in a buffer containing 100 µM PAR and 40 mM HEPES-KOH, pH 7.5. PMPS prepared as a stock solution was added and the absorbance at 500 nm monitored with a spectrophotometer. When the absorbance had saturated, the zinc concentration was inferred using an extinction coefficient of 66 000 M^{-1} cm.

7.6 Phagemid Display Library Construction

RNA Extraction

RNA was extracted from 1×10^7 B-cells per source using an RNAeasy-plus minikit (Qiagen) according to the manufacturer's instructions.

Reverse Transcriptase PCR

1 µg of total RNA was added to 1 µL of Oligo-dT (20 µM), 1 µL of dNTPs (10 mM), and DEPC-treated water added to a volume of 13.5 µL. The reaction mixture was incubated at 65 °C for 5 min and then put on ice for 5 min. To the cooled reaction, a mixture of 4 µM 5× FS buffer, 1 µL DTT (0.1 mM), 0.5 µL rRNasin (40 U μL^{-1}, and 1 µL of SuperScript III RT (200 U μL^{-1}) (ThermoFisher Scientific) was added and the new mixture incubated at 50 °C for 60 min, 70 °C for 15 min, and then held at 4 °C.

Antibody Amplification

Antibody variable regions were amplified from cDNA generated by reverse transcriptase RT-PCR in two PCR steps. The first step used the same primers as the first step of variable region amplification from single B-cell previously detailed. 1 µL of cDNA was added to 16 µL of DEPC-treated water, 2.5 µL of 10× KOD polymerase buffer, 1 µL $MgSO_4$ (25 mM), 2.5 µL dNTPs (2 mM each). In a PCR block the reaction was heated to 96 °C for 2 min, cycled 40 times through 96 °C (15 s), 55 °C (15 s), and then 68 °C

(1 min); finally the reaction was incubated at 68 °C for 5 min, and held at 4 °C.

The product was subjected to a second PCR to further amplify the variable regions and add overhangs in preparation for a pull-through PCR. Each of the following forward primers, specific for heavy-chain variable regions, was mixed separately with the reverse primer at a concentration of 20 µM.

FWD	5'-CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCCATYRADWTGAC-3'
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCMGYCGYVCTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGASVTNGTGATGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCSSWDRTGMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGAYVYBGTGMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGMTGTYGWGMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGAYGTBGWGATGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCCATCSTGMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCCMTSSTGMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGWCTATGATATGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCMTTSGAKWTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCCGYCGYRMTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCAACRTCGTGHTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGGCCAAGTGCTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCGCNCAARTGSTGAC
FWD	CAGGCGGAGGTGGCTCTGGCGGTGGCCTTAGCCAGYYTGTGCTGAC
REV	GTGATGGTGATGCGCGGCCGCCACGGGCGTACGTTYGACSACCACCTYGG

2 µL of each primer mixture was then used in a separate PCR reaction by combining it with 1 µL of the previous reaction, 15 µL of DEPC-treated water, 2.5 µL 10× KOD polymerase buffer, 2.5 µL dNTPs (2 mM each), 1.5 µL $MgSO_4$ (25 mM), 1 µM KOD Polymerase (Merck), and subjecting the reaction mixture to the following thermal cycle: 95 °C (5 min), 30 cycles of 95 °C (10 s), 58 °C (10 °C), 68 °C (30 s), 68 °C (10 min), and a hold at 4 °C. This was repeated for the following light chain specific primers using the same PCR conditions, and doing a separate reaction for each forward-reverse primer pair.

FWD	5'-GCGGCCCAGCCGGCCATGGCCCAGGAGCWRCTGAAGGAGTC-3'
FWD	GCGGCCCAGCCGGCCATGGCCCWGTCVHTRGAGGAGTCCGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGYMCYTGRWGGAGTCCGG
FWD	GCGGCCCAGCCGGCCATGGCCSAGCAGCTGGWGSAGTCCGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGSAGBTGRAGSAGTCCGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGTCGKTGGAGGARKCCGG
FWD	GCGGCCCAGCCGGCCATGGCCGAGCAGCTGGTGGAGTCCGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGGARCWRCTGGWGGMGTC
FWD	GCGGCCCAGCCGGCCATGGCCCAGAGCSTGGARGAAAGCGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGAGTCTTGAGGAATCTGG
FWD	GCGGCCCAGCCGGCCATGGCCCAGTCGYTGGWGGAGTCYVG
FWD	GCGGCCCAGCCGGCCATGGCCCAGTCDVTGRAGGAGTCCGR
FWD	GCGGCCCAGCCGGCCATGGCCCAGSAGCAGCTGGWGGAGTY
REV	CCAGAGCCACCTCCGCCTGAACCCGACTCCCCACTCGAGACGGTGACSAGGGTSCC

Pull-Through PCR

The heavy and light chains were combined as an scFv fragment and restriction sites added by pull-through PCR. 10 ng to 15 ng of amplified heavy chain variable region DNA, and 10 ng to 15 ng of amplified light chain variable region DNA were added to 2.5 µL 10× KOD polymerase buffer, 2.5 mM dNTPs (2 mM each), 1.5 µM $MgSO_4$ (25 mM), 1 µL of KOD DNA polymerase (Merck), 1 µL each of the following 10 µM primer stocks:

FWD	CATGACTACCGCGGCCCAGCCGGCCATG
REV	GTGATGGTGATGCGCGGCCGCCACGG

and water to adjust the volume to 25 µL. The reaction was heated in a PCR block to 95 °C for 5 min, cycled 10 times through: 95 °C (15 s), 61 °C (15 s), and 68 °C (1 min); finally the reaction was incubated at 68 °C for 10 min, and then held at 4 °C.

Ligation and Transformation

The scFv fragments generated in the previous step were purified using an agarose gel and gel purification kit (Qiagen). The purified products were digested with SfiI (10 U μL^{-1} (New England Biolabs), and NotI-HF (20 U μL^{-1}), and purified once again using an agarose gel. The digested inserts were then ligated into a digested phagemid vector (pTAPMID) with

T4 DNA ligase ($400\,\mathrm{U\,\mu L^{-1}}$) (New England Biolabs). After an overnight incubation at 16 °C the ligation reaction was purified using a Qiaquick PCR Cleanup column (Qiagen). The purified ligated phagemids, were then used to transform bacteria by electroporation. 25 µL of TG-1 electrocompetent cells (Lucigen) was combined every 1 µL of ligated DNA and transformed in a 2 mm curvette, using a BioRad Electroporator with capacitance set to 25 µF, resistance 200 Ω, and voltage 3000 V. After electroporation the cells the manufacturer-supplied recovery media was added and the mixture spread on LB agar plates supplemented with carbenicillin ($100\,\mathrm{\mu g\,mL^{-1}}$ and 1 % (w/v) glucose to grow overnight at 37 °C.

Phage Rescue

5 mL to 10 mL of LB was poured over the surface of a large culture plate and the attached colonies resuspended by gentle scraping of the surface. A portion of the suspension was then diluted into 50 mL to 250 mL of LB, supplemented with $100\,\mathrm{\mu g\,mL^{-1}}$ carbenicillin and 1 % (w/v) glucose, such that the OD_{600} value measured 0.1. The remaining resuspended material was frozen at −80 °C after adding 25 % (v/v) glycerol. The culture was incubated at 37 °C with shaking at 125 rpm until the $OD_{1–447}$ value reached 0.6, at which point the culture was infected with M13 helper phage (ThermoFisher Scientific) at a multiplicity of infection of 20. The cultures were left at 37 °C without shaking for 1 h and then centrifuged at 4000 rpm for 10 min. The supernatant was discarded, and the cells resuspended in an equal volume of 2YT media supplemented with carbenicillin ($100\,\mathrm{\mu g\,mL^{-1}}$), kanamycin ($50\,\mathrm{\mu g\,mL^{-1}}$). The cultures were incubated overnight at 30 °C with shaking at 125 rpm. The next day the phage-containing supernatant was separated by centrifugation at 8000 rpm for 20 min, and the phage purified by PEG precipitation.

Phage Panning (Plastic)

The inner surface of a plastic Immunotube (NUNC) was coated in with an antigen by incubation with 500 µL of a solution of 1 µg mL^{-1} antigen solution in PBS for 1 h at 37 °C. The tube was washed with PBS and 500 µM of phage, blocked by being combined with an equal volume of PBS containing 2 % (w/v) BSA (Sigma Aldrich) and 4 % (w/v) skim milk powder (Sigma Aldrich), incubated for no longer than 1 h at room temperature. The phage was removed and the immunotubes washed, in round one, with PBS containing 0.1 % (w/v) Tween-20 (Sigma Aldrich) five times, and then with PBS until there were no traces of Tween-20. In round two, the tubes were washed 20 times with PBS containing 0.1 % (w/v) Tween-20 and then PBS alone. Centrifugation of the tubes briefly at approximately 500 rpm forced any remaining liquid from the walls and can be collected with a pipette. 500 µL of 10 mM HCl (Sigma Aldrich) was added to the tube to elute any bound phage and left for 8 min with occasional agitation by vortexing. The 500 µL acid eluate was transferred to a new vessel containing an equal volume of 1 mol dm^{-3} TRIS-HCl, pH 8.0 (Sigma Aldrich) for neutralisation. The eluted phage was then added to 5 mL of XL-1 Blue cells (Agilent) that had reached log phase. The cells were to incubated at 37 h for 0.5 h to 1 h without shaking, so they could become infected. The cells were centrifuged, and their supernatant discarded, the cells resuspended in the dead volume of liquid that remained and then spread on LB agar containing 100 µM carbnecillin and 0.5 % (w/v) glucose. After an overnight incubation at 37 °C, phage was rescued using the general method already mentioned in round one and starting with round two, monoclonal rescue where individual colonies were picked, rescued and screened by phage ELISA.

Phage Panning (Solution)

A biotinylated antigen was diluted to a concentration of 500 nM in 250 µL to 500 µL of phage blocked as in the previous method. The two were incubated

for up to 1 h, when 50 µL of magnetic streptavidin-coated beads (Invitrogen) were added to the mixture to capture the biotinylated antigen and any bound phage. After 5 min to 10 min the beads were gathered using a magnetic rack and the phage removed by pipetting. During round one 1 mL of PBS containing 0.1 % (w/v) Tween-20 was used to wash the beads three times, and during round two the number of washes was increased to between 5 and 10. In both cases the beads were washed with PBS until there were no obvious traces of Tween-20 (2 to 4 times). Phage was eluted by incubating the beads in 10 mM HCl, with occasional vortexing, for no more than 8 min. Beads and acidic elution were separated using the magnetic wrack and the elution neutralised with an equal volume of 1 M TRIS-HCl. Infection then proceeded as it did for panning on plastic immobilised antigens.

Chemical Biotinylation

PHF8-PHD was chemically biotinylated using a Lightning-Link Ranpid Biotin Type A kit (Innova Biosciences) according to the manufacturers instructions.

Monoclonal Phage Rescue

A single colony containing a phagemid vector was picked from solid media and used to innoculate LB supplemented with carbenicillin (100 µg mL^{-1}) and 1 % (w/v) glucose. Most often 1 mL of media was used in a 48-well block. The culture was incubated overnight at 37 °C with shaking (600 rpm for 48-well blocks). These cultures, which reach stationary phase overnight, were then diluted into fresh media such that the final OD_{600} was 0.1, and rescue continued according to the general phage rescue protocol, but instead of phage precipitation, the supernatants themselves were kept for use in phage ELISA.

Phage ELISA

Performed according to the ELISA procedure already described except the blocking buffer was PBS containing 1 % (w/v) BSA (Sigma Aldrich). Phage was blocked in for at least 1 h prior to incubation by mixing equal volumes of phage-containing supernatant and PBS with 2 % (w/v) BSA and 4 % (w/v) skim milk powder (Sigma Aldrich). For the reveal step an HRP-conjugated monoclonal anti-M13 antibody (ThermoFisher Scientific) was used.

Supervisors

Professor Paul Brennan (Oxford)
Professor Chris Schofield (Oxford)
Dr. Alastair Lawson (UCB)

Protein Purification

Dr. Sebastian Mathea, *Oxford*
Eidarus Salah, *Oxford*
Kamal Abdul-Azeez, *Oxford*
Prashant Mori, *UCB*
Dr. Carl Doyle, *UCB*

Animal Handling

Sophie Shaw, *UCB*

Antibody Discovery

Dale Starkie, *UCB*
Dr. Dan Lightwood, *UCB*

Spectroscopy

Dr. Terry Baker, *UCB*
Dr. Alister Henry, *UCB*
Dr. Oliver Durrant, *UCB*

Phage Display

Dr. Anthony Scott-Tucker, *UCB*
Dr. Michael Wright, *UCB*

Demethylation Assays

Dr. Radoslaw Nowak, *Oxford*

NMR

Dr. Christine Prosser, *UCB*
Dr. Tim Claridge, *Oxford*
Dr. Taekyoung Yu

Crystallography

Dr. Oliver Durrant, *UCB*

Mass Spectrometry

Georgina Berridge, *Oxford*
Dr. Nisha Patel, *UCB*

Acknowledgements

Firstly I would like to thank the Clarendon Fund, from which I have received generous financial support through a Clarendon Scholarship. In addition to my supervisors, I would like to thank specially the scientists who have taken the time to teach and supervise my laboratory work. At Oxford, Dr. Sebastian Mathea, Eiduras Salah, and Kamal Abdul-Azeez have helped throughout my project. At UCB I have been lucky to have worked with Dale Starkie, who supervised antibody discovery, Dr. Anthony Scott-Tucker, who supervised phage display and most molecular biology, Dr. Oliver Durrant, who supervised both crystallography and spectroscopy, Prashant Mori who supervised protein purification, Dr. Christine Prosser who allocated a substantial amount of NMR time for me, Dr. Xiaohua Zhou, and Dr. Anastasios Spiliotopoulos. The advice of Dr. Dan Lightwood, Dr. Michael Wright, Dr. Alistair Henry, and Dr. Terry Baker at various points during my project has also been much appreciated. Finally, I would like to thank my industrial supervisor, Dr. Alastair Lawson, who has kindly hosted me at UCB for the last two years.

Experiments Performed by Other People

Immunisation: Sophie Shaw, *UCB*

X-ray diffraction: Diamond Light Source

Native MS: Dr. Nisha Patel, *UCB*