

# SCIENTIFIC REPORTS



OPEN

## Intracellular immunization against HIV infection with an intracellular antibody that mimics HIV integrase binding to the cellular LEDGF protein

Leyuan Bao<sup>4</sup>, Clare Hannon<sup>2</sup>, Abimael Cruz-Mignoni<sup>1,3</sup>, Denis Ptchelkine<sup>3</sup>, Mei-yi Sun<sup>4</sup>, Ami Miller<sup>1</sup>, Wilawan Bunjobpol<sup>1</sup>, Camilo E. Quevedo<sup>1</sup>, Mariliza Derveni<sup>1</sup>, Jennifer Chambers<sup>1</sup>, Alison Simmons<sup>4</sup>, Simon E. V. Phillips<sup>3,5</sup> & Terence H. Rabbitts<sup>1</sup>

Preventing the protein-protein interaction of the cellular chromatin binding protein Lens Epithelium-Derived Growth Factor (LEDGF) and human immunodeficiency virus (HIV) integrase is an important possible strategy for anti-viral treatment for AIDS. We have used Intracellular Antibody Capture technology to isolate a single VH antibody domain that binds to LEDGF. The crystal structure of the LEDGF-VH complex reveals that the single domain antibody mimics the effect of binding of HIV integrase to LEDGF which is crucial for HIV propagation. CD4-expressing T cell lines were constructed to constitutively express the LEDGF-binding VH and these cells showed interference with HIV viral replication, assayed by virus capsid protein p24 production. Therefore, pre-conditioning cells to express antibody fragments confers effective intracellular immunization for preventing chronic viral replication and can be a way to prevent HIV spread in infected patients. This raises the prospect that intracellular immunization strategies that focus on cellular components of viral integrase protein interactions can be used to combat the problems associated with latent HIV virus re-emergence in patients. New genome editing development, such as using CRISPR/cas9, offer the prospect intracellularly immunized T cells in HIV+ patients.

Protein-protein interactions (PPIs) are important inside cells for transcription, signaling and formation of complex organelles and a range of normal cellular activities such as maintenance, control of cell division, and quiescence. If these become abnormal in disease states, the modulation of the PPIs can be important for therapy. An important mediator of MLL (mixed lineage leukaemia; HGNC nomenclature KMT2A) protein function and of HIV productive infection is LEDGF (Lens Epithelium-Derived Growth Factor), (HGNC nomenclature PSIP1, also known as PC4 and SFRS1 interacting protein) functions by binding MLL but also Human Immunodeficiency Virus type 1 (HIV-1) productive infection through interaction with HIV-1 Integrase (IN). The main functions of LEDGF are in transport and tethering other proteins to chromatin. The N-terminal portion of the protein carries a nuclear localization sequence that binds to chromatin, and another domain called the Integrase Binding Domain (IBD) that binds to MLL and Menin (MEN1) reviewed in<sup>1</sup>.

Advances in anti-retroviral therapy (ART) have transformed HIV/AIDS from a deadly pandemic to a treatable but long-term chronic disease<sup>2</sup>. ART has contributed to reduction in HIV replication at the individual level and in subsequent transmission at a population level. This has raised optimism that Treatment-as-Prevention (TasP) may arrest the global pandemic by 2020. A barrier to these goals remains the ability of the virus to become

<sup>1</sup>Weatherall Institute of Molecular Medicine, MRC Molecular Haematology Unit, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK. <sup>2</sup>School of Clinical Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Rd, Cambridge, CB2 0SP, UK. <sup>3</sup>Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot, OX11 0FA, UK. <sup>4</sup>Weatherall Institute of Molecular Medicine, MRC Human Immunology Unit, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DS, UK. <sup>5</sup>Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK. Correspondence and requests for materials should be addressed to T.H.R. (email: [terence.rabbitts@imm.ox.ac.uk](mailto:terence.rabbitts@imm.ox.ac.uk))

resistant, necessitating the development of ever more sophisticated methods to target replication and prevent emergence of resistant strains. The recent use of inhibitors of HIV-1 IN for treating HIV-1 infection has proven to be very beneficial in this regard, with three HIV integrase strand transfer inhibitors (INSTIs) (raltegravir (RAL), elvitegravir (EVG) and dolutegravir) in clinical use. INSTI-based regimens are favoured for ART-naïve HIV-infected persons, based on improved tolerability and side effect profiles, better drug–drug interaction profile and high genetic barrier to resistance<sup>3</sup>. Despite this, resistance has emerged even to these drugs and other methods to target HIV IN such as targeting multimerization of the enzyme or cellular proteins required for IN function present attractive new avenues.

Current anti-HIV treatments require daily administration of a cocktail of drugs and compliance in this is a key feature of failure<sup>4</sup>. Further any new drugs that target HIV proteins, such as HIV IN binders, will also require long term drug regimes. A further critical point in HIV treatment is the latency of the virus in leucocyte pools that can re-emerge in an unpredictable way. The use of biological reagents against infections has been proposed and in particular intracellular antibody fragments<sup>5</sup>. This concept developed into the idea of intracellular immunization<sup>6,7</sup> where intracellular antibody fragments delivered into cells could interfere with specific disease protein functions<sup>7</sup>. In this respect, the LEDGF-HIV IN interaction represents an attractive therapeutic target<sup>8</sup> since this mediates viral integration into transcriptional active regions<sup>9</sup> thereby allowing maintenance of latent HIV for subsequent rounds of viral production<sup>8–10</sup>. The development of small molecule inhibitors of LEDGF and HIV IN for clinical use is in progress<sup>11,12</sup> but are still at the early stages<sup>13</sup> and the required potency of compounds that will interfere with this PPI may not be readily achieved as PPIs are hard to block by small compounds<sup>14</sup>. Peptides that inhibit HIV replication have been developed<sup>15</sup> as an alternative to compounds. Intracellular single domain antibodies offer a further alternative to small molecules for blocking PPIs in cells since these can readily be selected from diverse libraries and bind to specific target antigens within cells with high affinity.

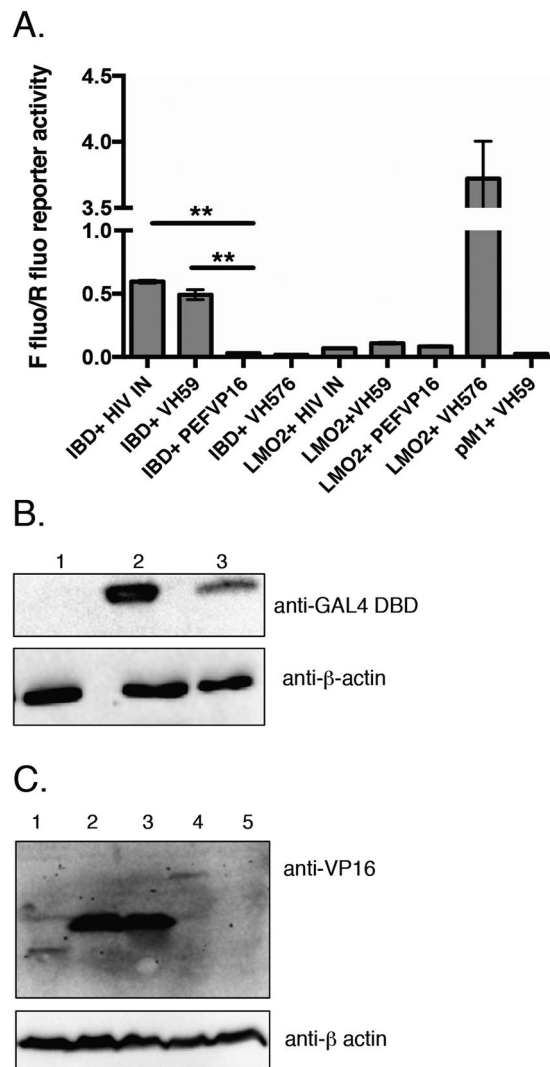
We have expanded on the concept of intracellular immunization by targeting LEDGF/PSIP1 rather than targeting the HIV integrase itself. Our data have been obtained using an intracellular single domain antibody (iDAb), identified by intracellular antibody capture (IAC) which is an in-cell selection method allowing scFv or iDAbs to be identified from diverse intracellular antibody libraries<sup>16</sup>. We show that a T cell line, constitutively expressing an anti-LEDGF/PSIP1 intracellular single domain antibody (iDAb), has a marked reduction in HIV p24 production in HIV-1 infected cells, reflecting interference with infected virus replication.

## Results

**Isolation of a heavy chain variable region single domain antibody binding to LEDGF IBD.** We have used IAC technology<sup>16</sup> to isolate a heavy chain variable region single domain binding to the IBD of the chromatin binding factor LEDGF. This technology uses *in vivo* (in yeast) screening with a two-hybrid strategy to probe diverse naïve human VH iDAb libraries. These libraries comprise a VH3 subgroup framework sequence and are randomized, by a PCR method, at the complementary determining regions (CDR) 3 giving individual libraries with the randomized CDR3 length of between 2 and 15 amino acids. Libraries with different lengths of CDR3 were screened with the IBD bait and clones that grew in the absence of histidine, due to interaction of the IBD bait and a VH-VP16 prey segment permitting transcription of the *HIS3* gene, were selected. Three clones were characterized further (herein called VH59, VH62 and VH65) originating from the libraries with the longest CDR3 (14 or 15 amino acids) and one clone, designated VH59, that originates from the library with CDR3 length 14, was selected on tryptophan, leucine and histidine knockout (WLH-minus) plates at the highest concentration of 30mM 3-amino-1,2,4 triazole (3-AT). The ability of VH59 to bind LEDGF IBD was also confirmed using the yeast system by comparing its binding to both full length LEDGF and the IBD as well as comparing VH62 and VH65 binding properties (Supplementary Figure S1). VH59 binds to both baits, unlike VH62 and VH65 (Supplementary Figure S1B, grid sectors 2, 3, 5, 7, 8 and 10). VH576, an anti-LMO2 iDAb, and Y6, an anti-Ras iDAb were used as negative controls. VH576-LMO2 and Y6-KRAS pairs were used as positive controls. VH59 was taken forward for detailed study.

Yeast expressing the IBD and VH59 sustain growth at 30 mM 3-AT, reflecting a high interaction affinity (Supplementary Figure S1C). This feature generally signifies sufficient affinity to support interaction in the more stringent setting of mammalian cell cytoplasm. CHO cells were transfected with mammalian versions of the two hybrid expression vectors, encoding a GAL4 DNA binding domain (DBD)-IBD fusion and a variety of VH domains fused to the VP16 activation domain (Fig. 1A) with luciferase reporter plasmids. A GAL4-DBD-LMO2 bait was used as a negative control and HIV IN was used as a known partner of the LEDGF IBD domain. VH59 showed similar luciferase activity to HIV integrase when co-transfected with LEDGF IBD bait whereas neither prey protein showed luciferase activity with non-relevant bait LMO2 (Fig. 1A). Western analysis confirmed the expression of IBD and LMO2 baits using an anti-Gal4 DBD antibody and the expression of prey proteins VH59, VH576 and HIV integrase was confirmed by Western blotting with anti-VP16 antibody (Fig. 1B,C). Comparisons of derived protein sequences of VH59 with previously identified iDAbs (anti-LMO2 VH576 and anti-RAS VHY6) are shown in Supplementary Figure S2A and the VH CDR3 is boxed. An alignment of the CDR3 regions of VH59, VH62 and VH65 proteins is shown in Supplementary Figure S2B and the 14 amino acid segment of VH CDR3 that was randomized in the original iDAb library is overlined.

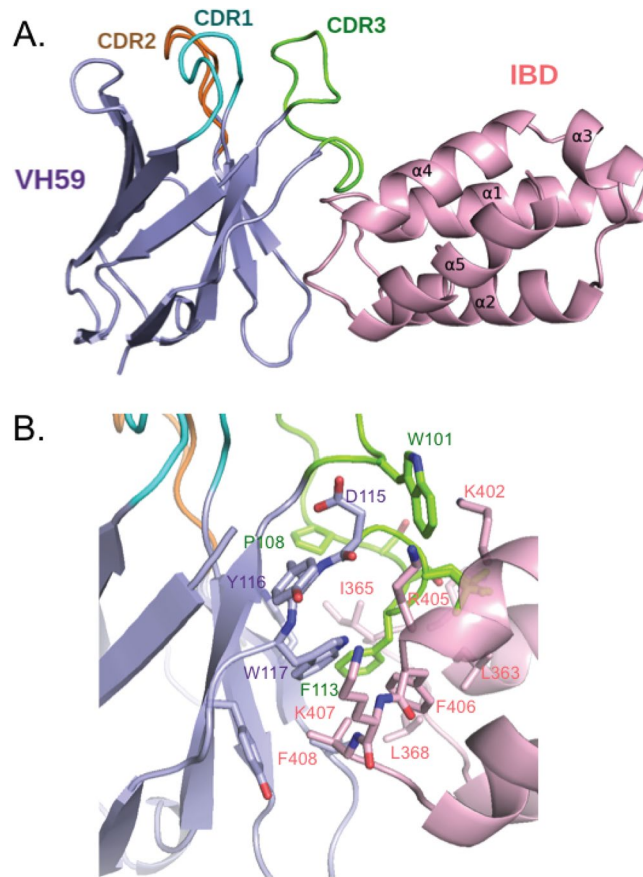
**Structural analysis of the LEDGF IBD-VH59 complex.** We determined the crystal structure of the IBD-VH59 complex to obtain detailed information on the mode of binding of VH59 to IBD. The structure was determined at a resolution of 1.7 Å by molecular replacement (Supplementary Table 1). We used the structure of the single domain VH anti-RAS intracellular antibody, determined in our previous study (PDB: 2UZI), and the IBD domain crystal structure from its complex with catalytic core domain of HIV-1 integrase (CCD) of HIV IN (PDB: 2B4J) as the search models. Good electron density allowed all four chains of the two independent VH59-IBD complexes, chains H/D and A/E, in the asymmetric unit to be built. The final refined model includes



**Figure 1.** Interaction of LEGDF IBD and anti-IBD VH59 in mammalian cells. The interaction of the LEGDF IBD with the iDab VH59 was tested by mammalian two-hybrid analysis using a luciferase reporter system in CHO cells. LEDGF IBD or LMO2 control baits were expressed as fusions with GAL4 DBD (pM1 vector) and various VH iDabs or HIV IN were expressed as prey fusions with the VP16 activation domain (pEF-VP16 vector). **(A)** Luciferase reporter activity in transfected CHO cells. Forty-eight hours after transfection, cells were lysed and luciferase activities were measured. Firefly/Renilla luciferase activity was calculated for each transformation pair (\*\*IBD and HIV IN or iDab VH59 interaction p-value < 0.01, n = 3). LMO2 bait and the anti-LMO2 VH576 were included as a positive interaction controls. **(B)** Western blotting using anti-GAL4 DBD antibody (upper panel) with protein extracts of mock transfected cells (lane 1), CHO cells transfected with pM1-IBD (lane 2) or pM1-LMO2 (lane 3); Protein levels were normalized by probing with an anti-β-actin antibody (lower panel). **(C)** Western blotting using anti-VP16 antibody (upper panel) with protein extracts of CHO cells transfected with pEFVP16 (lane 1); pEFVP16-VH59 (lane 2); pEFVP16-VH576 (lane 3); pEFVP16-HIV IN (lane 4); mock transfection (lane 5). Protein levels were normalized by probing with an anti-β-actin antibody (lower panel).

all amino acid residues (1–126) for both VH59 chains, and 347–425, 345–424 for IBD respectively. The full IBD construct encoded residues 345–431 (345–346 were introduced by the cloning procedure), and 14 residues are therefore disordered and invisible in the electron density maps. The intra-domain disulphide bridges are formed in only 30% and 60% of the two independent VH59 molecules respectively, presumably reflecting the intracellular reducing environment where the original selection was carried out and the complex was expressed, and leading to further minor disorder in the structure. The structure was refined to an R factor of 0.187, with an  $R_{\text{free}}$  of 0.251 (Supplementary Table 1). The R factor difference is higher than expected at this resolution, probably due to the missing IBD termini and disorder around the VH59 disulphide bridges.

The two complexes are essentially identical, and one of these (chains H and D) is shown in Fig. 2A. There are, however, some differences between the two VH59 chains in the region of CDR3 not in contact with IBD, and some surface loops, resulting from close contacts with neighbouring molecules in the crystal lattice, but these

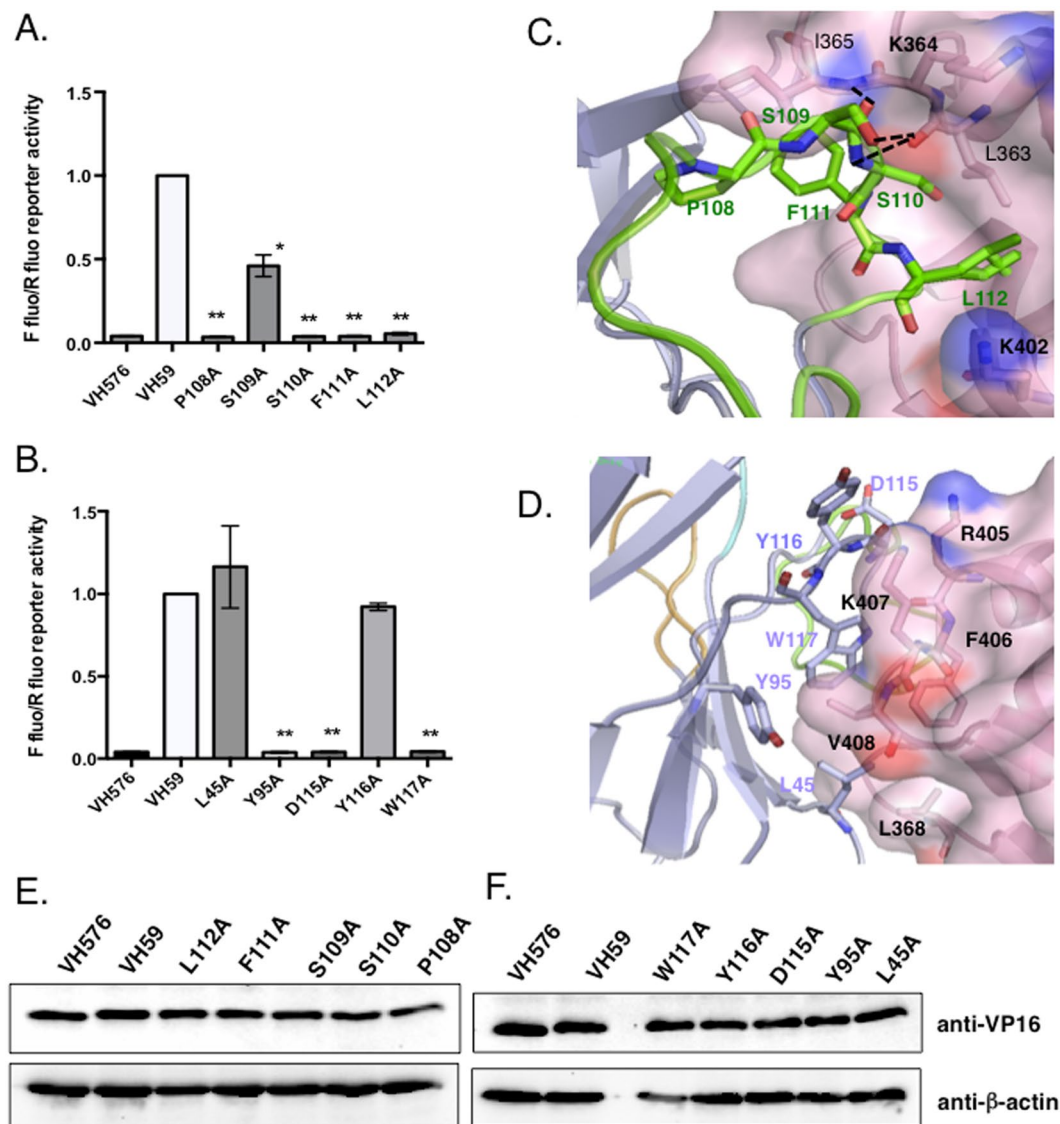


**Figure 2.** Structure of the LEDGF IBD and iDAb VH59 complex. The crystal structure of the complex between the IBD of LEDGF and the single domain antibody VH59. Panel A: Overall structure of the IBD-VH59 heterodimer: the LEDGF IBD subunit in pink, VH59 in blue. Panel B: The interface in IBD-VH59 heterodimer showing key interacting amino acids; the CDR1 of VH59 is shown in cyan, CDR2 in orange and CDR3 in green.

do not affect the VH59-IBD interface. The LEDGF IBD domain is very compact and contains five helices, while VH59 has a classical immunoglobulin  $\beta$ -sandwich fold acting as a scaffold for the three CDRs (or hypervariable loop regions), that normally mediate interactions with target antigens. Superposition of the IBD on the solution structure of IBD (PDB: 1Z9E), determined by NMR, and those from the other IBD complexes used as molecular replacement search models, showed that VH binding does not affect the overall conformation of the LEDGF IBD (Supplementary Figure S3).

The binding site for VH59 on the IBD surface is mainly formed by the two loops that connect helices 1 and 2 (loop1–2 amino acids 364–368) and helices 4 and 5 (loop4–5 amino acids 405–408) of IBD (Fig. 2), with additional contacts to Leu363 and Lys402 at the C-termini of helices 1 and 4, respectively. The binding site on VH59 does not correspond to a normal antibody combining site, with IBD binding to one side of the  $\beta$ -sandwich, rather than interacting with the three CDR at the end of the domain. Only the CDR3 of the iDAb makes contact with IBD, unlike many antibody-antigen interactions where two or three CDR of each domain are involved. The C-terminal region of the long CDR3 turns away from the other CDRs to form a loop (amino acids 108–114) that wedges between the two IBD loops to form the major part of the interface. Additional contacts are made by amino acids 95 and 115–118 either side of CDR3, that formally belong to the framework region, and the interface is completed by contacts from amino acids 45–47 in another framework region, the  $\beta$ -strand preceding CDR2. The overall interaction surface on VH59 is largely hydrophobic and covers an area of  $700\text{\AA}^2$ . The IBD surface is mostly complementary, but the complex leaves a central cavity between the molecules, cut off from bulk solvent. This cavity has space for several water molecules, but only one or two, respectively, are ordered in the two independent complexes in the crystal. The IBD-VH59 interactions include four hydrogen bonds, all of which are with CDR3 and mostly via main chain donors and acceptors. (Fig. 3C,D; Supplementary Table 2). The large number, and hydrophobic nature, of the interacting amino acids suggest strong interaction between the proteins, consistent with the observation that IBD-VH59 elutes as single peak on size exclusion chromatography (Supplementary Figure S4).

The key interacting amino acids of VH59 site identified in the crystal structure of the complex were confirmed by mutagenesis. Single mutations of VH59 CDR3 amino acids were made and cloned into pEF-VP16 (Fig. 3, Supplementary Figure S2). These were co-transfected into CHO cells with a pM1 clone expressing the GAL4-DBD-IBD and both a luciferase and Renilla reporter plasmid. The binding site on the VH59 is mainly formed by amino acids with large hydrophobic side chains and mutating those to amino acids with a short



**Figure 3.** Anchor amino acids in iDAb VH59 determined by alanine scanning. Key amino acids of iDAb VH59 CDR3 and framework region were mutated to alanine using site-directed mutagenesis. pEFVP16-VH59 or mutant VH59 were co-transfected with pM1-LEDGF IBD plasmid into CHO cells. After forty-eight hours of transfection, cells were lysed and luciferase activity was measured. Firefly and Renilla luciferase activities were determined for each VH59 CDR3 mutation (panel A) or framework mutation (panel B) p-values are shown as \* $p < 0.05$ , \*\* $p < 0.01$ ,  $n = 3$ . Key amino acids of VH59 CDR3 (panel C) and framework region (panel D) binding to the IBD are indicated in the crystal structure. Western blotting was used to determine expression of VH59 or mutant VH59 fusion proteins with anti-VP16 antibody (panels E and F)<sup>41,42</sup>.

aliphatic side chains would be expected to lead to decreased interaction, loss of stereo-complementarity and reduction of the interface area. Indeed, amino acid mutations P108A, Y95A, F111A, L112A and W117A significantly reduced the interaction of VH59 with the IBD (Fig. 3A,B). P108 makes no direct contacts to IBD, but the conformational restriction imposed by proline probably helps stabilize the observed CDR3 loop conformation. Its replacement by alanine could alter the conformation, and the additional flexibility could increase the entropic barrier to binding. The impact of additional direct contacts made to the backbone of the protein and H-bond mediated contacts was confirmed by the IBD mutants S109A, S110A and D115A as they also showed reduced interaction with VH59. This observation underlines the importance of a local conformation for specific recognition between the two proteins. Expression of all the mutated VH59 proteins is comparable to un-mutated VH59 (Fig. 3E,F).

We have measured the  $K_d$  using SPR and obtained a value of around  $1 \mu\text{M}$ . We have assessed this with scFv and Dab using biotinylated IBD on the SPR chip and with a second protein preparation of scFv using GST-IBD on the SPR chip (Supplementary Figure S5). We believe that reflects the difference in protein structure and flexibility between periplasmic expressed recombinant scFv, which will have S-S bonds<sup>17</sup>, and the intracellular expressed

protein, which will not. The structural analysis of VH in the co-expressed IBD-VH shows a framework interaction of the antibody fragment with IBD, and only partial oxidation of the S-S bridges. The restraints placed on the VH structure by S-S bridges are likely to reduce its affinity for IBD compared to the reduced form, consistent with oxidation being incomplete in the crystal structure despite exposure to oxidising conditions during purification and crystallization. This would also be true for the sample used for SPR-based K<sub>d</sub> measurement, so this would represent a lower limit for affinity.

**VH59 and HIV IN catalytic core domain share same binding site on the IBD.** In order to determine whether VH59 interferes with the binding of partner proteins to IBD, we have made structural comparisons to other known IBD complexes<sup>18</sup>. Alignment of IBD-VH59 and IBD-HIV IN catalytic core domain (CCD) structures (PDB: 2B4J) revealed a significant overlap of binding sites of VH59 and HIV-1 integrase CCD proteins on the IBD surface (Fig. 4). The VH59 binding site covers a broad area on the IBD protein surface that encompasses the whole CCD binding site (Fig. 4A) and extends further to the neighboring area (the overall buried surface for IBD-VH59 is 1391 Å<sup>2</sup> compared to 1280 Å<sup>2</sup> for IBD-CCD (Fig. 4B)). Indeed, binding of VH59 and CCD are very similar as most of IBD-IN CCD contacts are made to the same amino acids of IBD as those of IBD-VH59 (Fig. 4C highlights these common amino acids in a comparison of human and mouse IBD; Supplementary Table 2).

The structural alignment of IBD-VH59 and IBD-IN allowed identification of the amino acids on the surface of IBD that interact with both to HIV IN and VH59. As expected, mutations of these amino acids (L363A, I365A, L368A, F406A and V408A) resulted in a significant decrease in the interaction of the IBD with both VH59 (Fig. 5A) and IN-CCD (Fig. 5B).

**VH59 inhibits HIV replication of CD4 T cells by competing HIV integrase binding to LEDGF IBD.** The co-incident binding of VH59 to the LEDGF protein at the same amino acids as the HIV IN suggests that VH59 could be a biological inhibitor of HIV by interfering with HIV genome integration. A mammalian three hybrid system was used to assess whether VH59 could compete with the binding of HIV integrase to LEDGF IBD. The interaction of HIV IN-VP16 and the DBD-IBD was assayed in the CHO-luciferase reporter assay (Fig. 6A) and this interaction was not affected by co-expression of a non-relevant VH (anti-LMO2 VH576). Conversely, co-expressing the HIV IN-VP16 and Gal4-IBD proteins with either free VH59 or HIV IN resulted in about 90% inhibition of luciferase production due to competition. Western blotting confirmed the expression of VH59 and VH576 proteins in this assay (Fig. 6B).

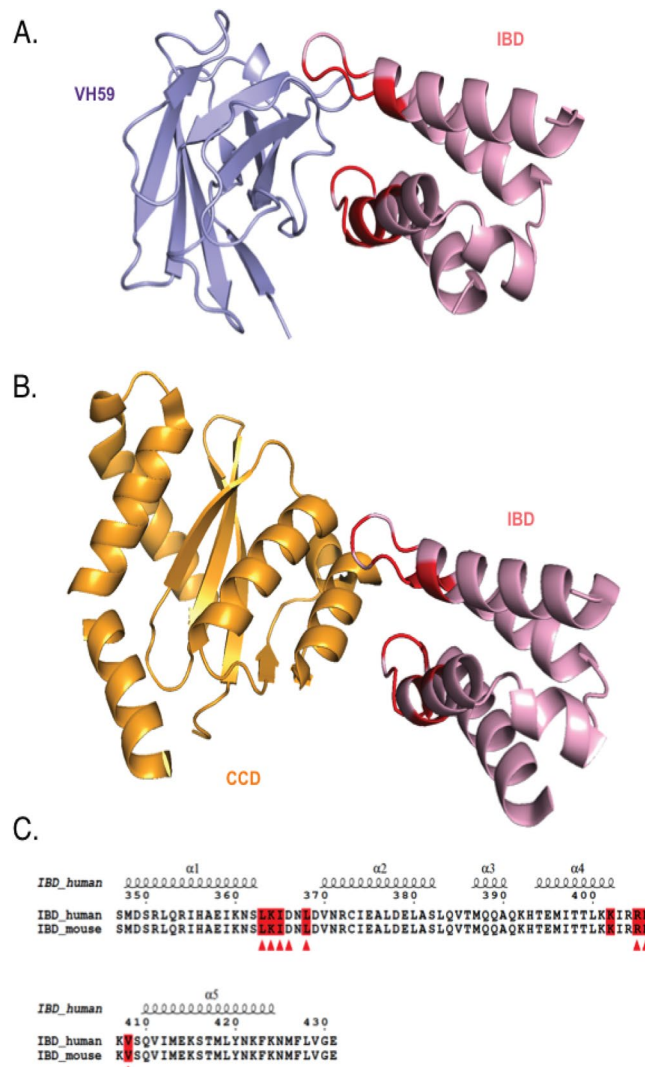
In HIV-1 infections, CD4-positive T cells express IN that interacts with LEDGF for nuclear import and chromatin tethering of the pre-integration complex, facilitating integration of the viral DNA into the host genome<sup>19</sup>. Since VH59 acts as a competitor of LEDGF-HIV IN interaction, we tested if this competitive binding ability would have an inhibitory effect on HIV replication. We used an HIV infection assay in which VH59, or anti-LMO2 VH576 as a control, was expressed from a lentivirus vector in the CD4-expressing Jurkat cells and effects on HIV replication was measured by determining production of the p24 capsid protein (Fig. 6C). Increasing levels of HIV virus were used to infect VH59 or VH576 transduced Jurkat cells and p24 expression level was determined after 5 days. We observed that the p24 levels in VH59 expressing cells were significantly lower than the controls at all HIV virus titres tested. In cells infected with the highest viral titre, presence of VH59 suppressed the expression of p24 by 50% compared to p24 in cells expressing the non-relevant VH iDAb or the empty lentiviral vector (\*\*P < 0.01). The inhibitory effect of VH59 on the interaction of LEDGF and HIV IN translates into an inhibitory effect on the production of HIV virus if cells have been pre-loaded with VH59 before HIV exposure. This means that the VH59-expressing cells have effectively undergone intracellular immunization.

## Discussion

**An intracellular domain antibody binding to the HIV IN binding domain of LEDGF.** Intracellular antibody capture technology (IAC)<sup>20,16</sup> was developed to allow selection of antibody fragments from diverse libraries in the reducing environment of cells to allow their binding to intracellular target proteins. It was originally developed with scFv selection<sup>20</sup> and later using single variable region iDabs<sup>16</sup>. Diverse libraries of VH segments, based on a human VH3 subgroup sequence, were generated with diversity at the CDR3 region for screening in yeast against antigens of choice<sup>16</sup>. The binding affinities of initial hits can be improved as required by an *in vivo* affinity maturation protocol<sup>21</sup>. The iDAb described here (VH59) was developed using IAC selection of a diverse VH library with CDR3 length 14 amino acids.

Blocking the PPI between LEDGF and HIV IN can affect HIV replication<sup>15</sup> and small molecule drugs that work on the LEDGF-HIV IN interface could potentially be important for therapy. However, those in development are still not in clinical use owing to potency issues, probably because the flat and rather featureless interaction surface represents a poor target for small molecules that bind more strongly in clefts. Compounds that bind to either HIV IN<sup>11,13,22</sup> or to LEDGF/p75<sup>23</sup> have been isolated and are progressing to clinical use. However, macromolecules (such as iDabs), that bind either LEDGF or HIV IN, would also have potential for therapeutic application in early HIV infections and implementation should be possible by coupling new genome editing methods with using new delivery methods<sup>24,25</sup> to allow expression of iDabs in CD4+ T cells.

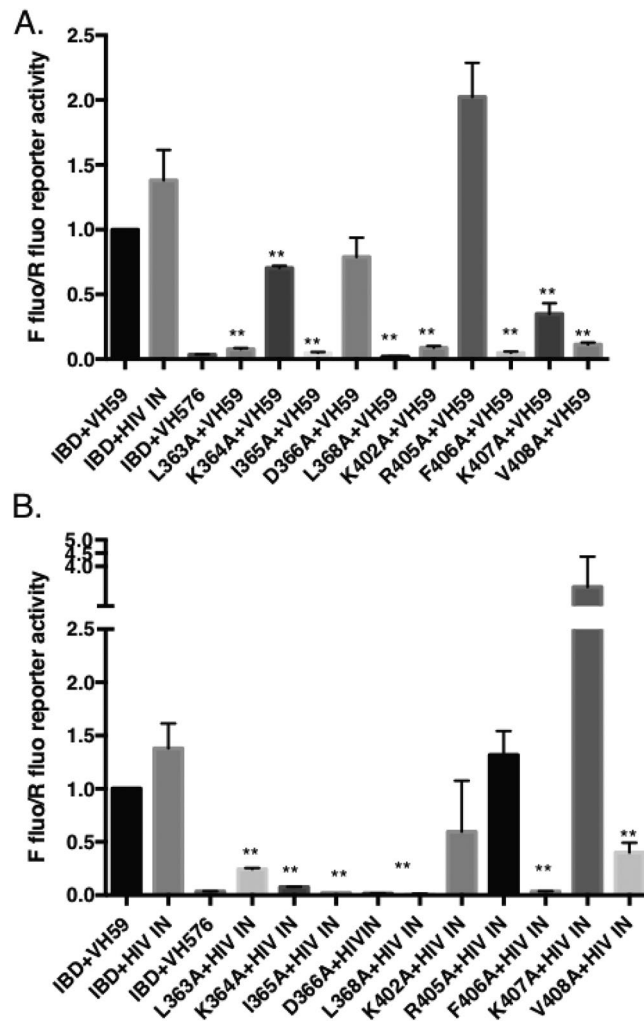
Several intracellular antibodies have been described against HIV proteins, including structural proteins, TAT, REV, reverse transcriptase and integrase (reviewed in<sup>26</sup>). In our study, we have selected a heavy chain variable region iDAb, using IAC screening that binds the cellular protein LEDGF and proved to compete with the binding of HIV IN to LEDGF. The crystal structure of the IBD-VH59 complex shows a large area of interaction between the binding partners, as often observed in antibody-antigen complexes, with a particularly high proportion of hydrophobic interactions and few hydrogen bonds. More striking is the observation that IBD is not bound by all three CDR loops, but lies alongside VH59 domain. This unusual antigen binding site is not unprecedented for



**Figure 4.** Comparison of LEDGF IBD binding modes with iDAb VH59 and HIV integrase. The crystal structure of the VH59-LEDGF IBD complex determined in this paper is shown in panel A while panel B shows the HIV integrase-LEDGF IBD complex crystal structure (PDB: 2B4J). Panel C shows the co-occurrence of amino acids mediating VH59 binding (highlighted in red) and the binding sites of HIV IN (marked with arrows) in sequence alignment, using ESPript 3.0<sup>43</sup> of mouse and human LEDGF IBD domains indicating the five helical regions.

VH complexes, and is similar to that observed for the complex between the anti-LMO2 VH576 and its LMO2 target<sup>27</sup>, where CDR3 is the only hypervariable loop involved, in addition to contacts within the VH framework amino acids. This is probably the result of the IAC screening procedure, where the VH59 protein sequence was derived from a diverse VH library with 14 amino acids of CDR3 randomized while CDR1 and CDR2 were not<sup>16</sup>. The framework interactions can therefore provide additional possible binding surfaces and compensate to some extent for the lack of CDR1 and CDR2 diversity. In addition, the framework region on the face of the VH59  $\beta$ -sandwich adjacent to CDR3 is hydrophobic, since this is the surface that would be buried in the interface to VL in a normal antibody. This is, therefore, complementary to the hydrophobic nature of the interaction site on IBD.

**VH59 mimics HIV IN binding sites on LEDGF.** Our intracellular antibody binds the cellular protein LEDGF, and the binding site of HIV IN on the IBD of LEDGF has been mapped onto two loops on the surface<sup>28</sup>. The crystal structure indicated that VH59 binds to LEDGF IBD at an almost identical position as the integrase. The conformation of the IBD in the context of the full length LEDGF protein is not known although it must be accessible to integrase and is very likely to be the same as in the LEDGF IBD-IN crystal structure. It is therefore also likely to bind VH59. The crystal structure analysis, and site-directed mutagenesis of VH59 and IBD, show common binding amino acids on LEDGF IBD that interact with VH59 and HIV IN (Fig. 5). This explains why VH59 competes with the binding of HIV integrase with LEDGF. The property of inhibiting LEDGF IBD-IN interaction makes the VH59 iDAb a potential biological inhibitor of HIV infection by interfering with HIV replication.

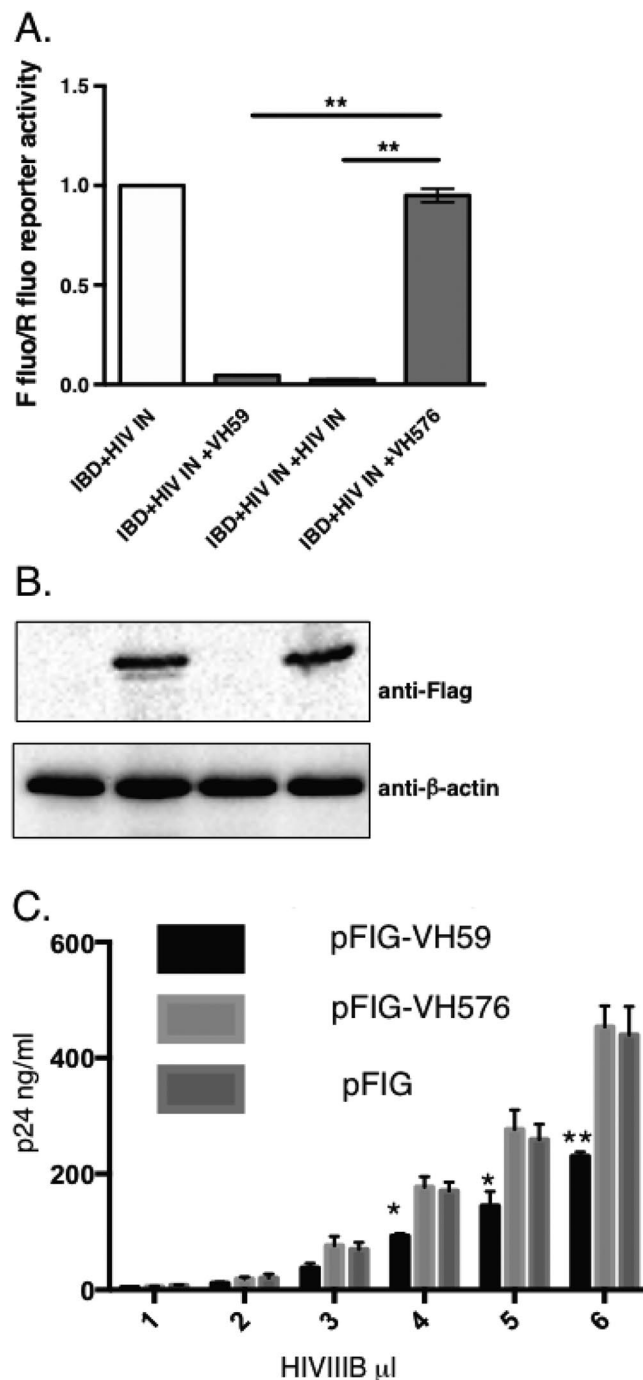


**Figure 5.** Interaction of mutant LEDGF IBD with VH59 and with HIV IN. Key amino acids of the LEDGF IBD were mutated to alanine using mutagenesis. Expression vectors encoding the mutant proteins were co-transfected in CHO cells with either an expression vector encoding VH59 (panel A) or HIV IN (panel B) together with a Firefly luciferase reporter plasmid and Renilla luciferase expression plasmid for normalization. Forty-eight hours after transfection, cells were lysed and luciferase activities were measured. The interaction of the IBD with wild-type VH59 or HIV IN were used as positive controls and between the IBD and an anti-LMO2 VH (VH576) as a negative control.

The probability of resistance development to VH59 should be low, as it targets a cellular protein. I could be wrong but I didn't think LEDGF had enzymatic activity(?) by contrast with integrase inhibitors. iDABs have a larger binding surface area unlike HIV integrase inhibitors, such as raltegravir that are also susceptible to mutations, to which the HIV genome is prone<sup>29</sup>. Our data show that binding of an iDAB to the LEDGF surface, rather than the HIV genome-encoded integrase, is effective and this protein is unlikely to acquire mutations in view of its function. While *Legdf/Psip1* gene in mouse knockouts show perinatal lethality<sup>30</sup>, these knock-out mice have normal immune systems and TALEN-based depletion of human CD4+ T cells show no deleterious effects of losing LEDGF<sup>31</sup>. In an *in vivo* implementation of LEDGF-iDAB binding in T cells, this should not produce a phenotype other than preventing HIV productive genome integration.

**The intracellular domain antibody is an inhibitor of HIV.** LEDGF is a transcription co-factor that interacts with several proteins, traffics into the nucleus and binds strongly to chromatin (Christ and Debysers, 2013). This protein has functions in MLL-leukaemias mediated by binding of MLL and Menin to IBD helix 5<sup>32</sup>. The interaction of HIV IN with LEDGF is needed for the uptake of the viral pre-integration complex into nuclei of infected cells for chromatin binding and viral genome integration<sup>8</sup>. In view of this, there is great interest in developing inhibitors to block LEDGF and HIV-1 protein interactions for potential therapy of HIV infection. As an alternative to small compounds, we have described a new inhibitor of the interaction of HIV IN and LEDGF with the ability to interfere with the HIV virus replication cycle. This is the first intracellular antibody that binds this cellular protein rather than an HIV encoded protein. Using *in vitro* HIV IIB infection of CD4 T cells as a model, we show that VH59 inhibits virus production measured by decreased p24 capsid protein levels in the





**Figure 6.** VH59 disrupts the interaction of LEGDF IBD with HIV IN in Jurkat T cell inhibiting the production of viral capsid protein p24. A mammalian LEDGF IBD bait vector (in pM1 to make a GAL2 DBD-IBD fusion) was co-transfected into CHO cells with the prey vector expressing HIV IN (in pEF-VP16 to make an integrase-VP16 activation domain fusion) plus the Firefly and Renilla luciferase reporter plasmids either alone or with an expression vector expressing VH59, HIV IN or an anti-LMO2 VH576. Forty-eight hours after transfection, cells were lysed and luciferase activity was measured. Firefly/Renilla luciferase activity was calculated for each transformation pair (\*\*  $P < 0.01$ ,  $n = 3$ ) (panel A). The VH protein levels were estimated using an anti-FLAG tag antibody and normalized using an anti-β-actin antibody (panel B). Jurkat cells were infected with a lentivirus expressing the anti-IBD iDAb VH59, a lentivirus expressing anti-LMO2 VH576 or lentiviral vector alone. Transduced cells were sorted by flow cytometry detection of GFP expression. These cells were infected with escalating doses of HIV IIB virus and after 5 days, supernatants were collected for p24 protein expression determination by ELISA (Panel C).

supernatant. Our use of iDAb VH59 shows that by pre-loading target cells with an antigen-specific intracellular single domain antibodies, we can effectively immunize against HIV-1 infection. This type of intracellular immunization has potential for use in treating HIV-1 infected individuals through stem cell genome editing approaches wherein all subsequently progeny CD4<sup>+</sup> T cells will express the anti-viral VH and thus would be impervious to continued HIV-1 infection. This would permit healthy T cell re-population. However, for this intracellular immunization technology to become clinically feasible, the uptake of the therapeutic in a cell-specific manner is dependent upon *in vivo* delivery system advances<sup>24</sup> (although this is also true of other biological therapies currently in development).

**Conclusions and Future Prospective.** We show that blocking LEDGF interaction with HIV IN in cells constitutively expressing an intracellular single domain antibody binding to LEDGF interferes with HIV replication. This is distinct from intracellular antibodies targeting viral proteins and small molecule inhibitors of LEDGF:HIV IN and indicates that anti-LEDGF iDAb expression in CD4<sup>+</sup> T cells will be without deleterious effects *per se* whilst influencing HIV replicative processes.

One reason why anti-HIV intracellular antibodies have not progressed to clinical use is the difficulty of their delivery to target cells in patients. Our paper shows that adding an intracellular antibody expression gene to the genome gives immunity to these cells. This is ‘intracellular immunization’<sup>7</sup> as the cell has prior production of an single domain antibody to target a viral antigen. We envisage that cell-specific antibody fragment expression will be possible when the new methods of genome editing (viz. CRISPR/cas9-mediated genome integration) are achieved to specifically express iDAbs in T cells or macrophages (the targets of HIV infection or reservoirs of latent virus). Intracellular immunization has potential for therapeutic application in early HIV infections and implementation should be possible using new delivery technologies that are becoming feasible<sup>24,33,34</sup>. Such a gene editing approach has recently been used in the correction of mutations in human embryos<sup>35</sup> and promises to provide the means to stably alter cells to express exogenous genes such as those encoding intracellular antibodies. These technologies that will facilitate intracellular immunization will be alternatives to small molecule drugs that may lack potency and to which resistance can emerge.

## Materials and Methods

**Intracellular antibody capture (IAC) selection of VH iDAbs.** A construct encoding the LEDGF IBD (amino acids 345–431) was cloned into the yeast bait vector pBTM116 and transfected into L40 yeast. Eleven single domain VH iDAb diverse libraries, with randomized CDR3 loops ranging from 5 to 15 amino-acids<sup>16</sup> were transfected into the LEDGF IBD yeast strain and clones showing IBD-VH interaction were identified by growth of yeast on plates lacking tryptophan, leucine and histidine<sup>16</sup>. Three VH clones were identified and were selected for study (designated VH59, VH62 and VH65).

**Mammalian two-hybrid and three-hybrid analysis.** The LEDGF IBD coding region was cloned into the bait expression vector pM1, and VH59 and VH576 (an anti-LMO2 iDAb) were cloned into the prey expression vector pEF-VP16<sup>16</sup>. For mammalian two hybrid analysis, bait and prey vectors were transfected into CHO cells using lipofectamine 2000 (Invitrogen) together with the Firefly luciferase reporter pG5-Fluc and control Renilla luciferase vector pRL-CMV. Forty-eight hours after transfection, Firefly and Renilla luciferase enzyme activities were measured using the Dual-Luciferase reporter assay system (Promega) according to the manufacturer’s instructions. Data were analyzed using a Lucetta™ Luminometer (Lonza). The relative reporter activity was calculated as a ratio of Firefly luciferase activity/Renilla luciferase activities.

For three-hybrid competitor analysis, competitor cDNA sequences were cloned into the pEF-BOS-cyto vector (Invitrogen) and bait, prey and competitor plasmids were co-transfected into CHO cells together with pG5-Fluc and pRL-CMV reporter vectors. The luciferase activities were measured as for the mammalian two hybrid assay. Statistics were carried out using paired T-test. \* indicate  $P < 0.05$ , \*\* indicate  $P < 0.01$ .

**LEDGF IBD and VH59 protein co-expression in *E. coli*.** LEDGF IBD and VH59 iDAb were co-expressed in *E. coli* C41(DE3) cells. The LEDGF IBD coding region (amino acids 347–431) and the VH59 sequence were cloned into the pRK-172 vector in-frame with an N-terminal 6 × his-tag and a TEV protease recognition site with the two cistrons linked by an internal ribosome binding site (RBS). The plasmid containing IBD-VH59 sequences was transformed into C41(DE3) cells and grown in 50 ml LB medium with 100 µg/ml ampicillin for 24 hours, and subsequently added to 1 L LB medium. Protein expression was induced at OD = 0.2 by addition of 0.5 mM IPTG and grown at 37 °C for 5 hours. Bacteria were collected and sonicated in 20 mM Tris-HCl pH8, 250 mM NaCl, 20 mM imidazole, 0.5% NP40, 5 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Proteins were purified using nickel agarose beads (Invitrogen) and bound proteins were eluted in 20 mM Tris-HCl pH 8, 250 mM NaCl, 300 mM imidazole, EDTA-free protease inhibitor cocktail. HIS-tagged TEV protease (1.5 mg/ml) was added at a ratio of 1:100 and dialysed against 20 mM Tris-HCl pH 8, 250 mM NaCl. Protease and free HIS-tag were removed using nickel agarose beads (Invitrogen) and unbound proteins were eluted in 20 mM Tris-HCl pH 8, 250 mM NaCl, 300 mM imidazole, EDTA-free protease inhibitor cocktail. The final samples were concentrated using Amicon Ultra-15 centrifugal filter devices with a 10kDa cut off (Millipore, MA, USA). The proteins were further purified by gel filtration on a HiLoad Superdex 75 10/300GL column (GE Healthcare, Uppsala, Sweden) in a buffer containing 10 mM Tris-HCl, pH 8, 150 mM NaCl at a flow rate of 1 ml/min. Fractions corresponding to the protein complex were pooled and concentrated to 6 mg/ml for crystallization trials. Protein concentrations were determined by Bradford assays. Protein purity was analyzed by SDS-PAGE stained with Coomassie Brilliant Blue.

**VH59-IBD crystallization and structural determination.** Crystallization experiments were carried out in 96-well Greiner plates using 200 nanolitre droplets. Crystallization plates were stored at 4 °C and examined at regular intervals. Crystals of the IBD-VH59 complex suitable for X-ray diffraction grew in 25.0% w/v Polyethylene Glycol 3350, 0.1 M Tris pH 8.5. Crystals were cryo-protected with Paratone and flash-cooled in liquid nitrogen. X-ray diffraction data were collected at station I04-1 of the Diamond Light Source (DLS, Oxfordshire). A total of 1800 diffraction images were collected, each with an oscillation range of 0.2°. The IBD-VH59 crystals belong to space group P1, with unit cell dimensions  $a = 35.1$ ,  $b = 41.4$ ,  $c = 58.7$  Å,  $\alpha = 104.0$ ,  $\beta = 96.1$ ,  $\gamma = 100.5$ °. The structure was solved by molecular replacement. The asymmetric unit contains two copies each of IBD and VH proteins. An initial search was done in PHASER and included two molecules of VH<sup>36</sup>. Additional searches performed in MolRep allowed correct localization of the two missing molecules of IBD<sup>37</sup>. Model building was carried out using Coot<sup>38</sup> and refinement was done with REFMAC<sup>39</sup>. Data collection and refinement statistics are shown in Supplementary Table 1.

**SDS-PAGE and Western analysis.** CHO cells transfected with different bait and prey vectors were lysed in 2% SDS/PBS, and boiled in sample buffer (100 mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 2% (w/v) SDS and 2% (v/v)  $\beta$ -mercaptoethanol). 30  $\mu$ g of protein were fractionated on 12% denaturing acrylamide gels. For Western blotting, proteins were transferred to polyvinylidene difluoride (PVDF) membranes, blocked with 5% non-fat milk and incubated with either primary antibody mouse anti-VP16, mouse anti-Gal4 DBD (Santa Cruz Biotech) or mouse anti- $\beta$ -actin (Sigma-Aldrich) (the latter was for loading control). The membranes were washed three times with 10 mM Tris, pH 7.5, 50 mM NaCl and 0.1% (v/v) Tween-20 and incubated with horseradish peroxidase-labeled anti-mouse antibody (Sigma-Aldrich). A ChemiDoc MP enhanced chemiluminescence imaging system (GE Healthcare, Uppsala, Sweden) was used for data collection.

**Preparation of GST-IBD and scFv-59.** GST-IBD: cDNA encoding the integrase binding domain, residues 345–431 of LEDGF/p75, was cloned into the pGEX-2T vector in-frame with the C-terminal Glutathione-S transferase (GST). The pGEX-GST-IBD plasmid was transformed into *E. coli* C41 (DE3) cells for protein expression. Bacterial cells were cultured in LB media containing the appropriate antibiotics at 37 °C, 225 rpm until an OD<sub>600</sub> of 0.6 was reached. Protein expression was induced by the addition of 0.1 mM IPTG (isopropyl 1-thio-beta-D-galactopyranosid) and the cells were incubated overnight at 16 °C, 225 rpm. Cells were harvested by centrifugation and the pellet was resuspended in lysis buffer (10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, 2.7 mM KCl, 140 mM NaCl, pH 7.3) containing EDTA-free protease inhibitor cocktail tablets (Roche, Germany). Cells were lysed using a cell disruptor (Constant Systems Ltd., UK) at 25,000 psi, 4 °C and the sample was clarified by centrifugation at 50,000  $\times$  g for 1 hour. The cell lysate was incubated with glutathione Sepharose 4B (GE Healthcare, UK) for 2 hours at 4 °C. Soluble GST-IBD was eluted using 50 mM Tris, 10 mM reduced glutathione (GSH), 0.1 ZnSO<sub>4</sub>, pH 8.0. The concentrated protein sample was further purified by gel filtration using a HiLoad 16/600 Superdex 75 column (GE Healthcare, UK) in a buffer containing 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 8.0. The purified GST-IBD protein was concentrated to 0.5 mg/ml for SPR experiments.

VH59-scFv: was cloned into the pSANG10 vector in-frame with C-terminal 6xHis and FLAG tags. Plasmid DNA was transformed into *E. coli* C41 (DE3) cells for protein expression. A single colony was used to inoculate 50 ml of 2xTY containing 2% glucose and 50  $\mu$ g/ml kanamycin, which was grown overnight at 30 °C, 225 rpm. The overnight seed culture was diluted 1:100 in 4  $\times$  1 L of 2xTY containing 0.1% glucose and 50  $\mu$ g/ml kanamycin. The cultures were grown at 37 °C, 225 rpm until an OD<sub>600</sub> of 0.8 was reached. Protein expression was induced by the addition of 1 mM IPTG (isopropyl 1-thio-beta-D-galactopyranosid) and the cells were incubated overnight at 30 °C, 225 rpm. Cells were harvested by centrifugation and protein extracted by periplasmic osmotic shock. The cell pellets were resuspended in TES buffer (50 mM Tris, 1 mM EDTA, 20% sucrose, pH 8.0) containing EDTA-free protease inhibitor cocktail tablets (Roche, Germany) and 100  $\mu$ g/ml lysozyme (50 ml/L culture). Following a 10 minute incubation on ice, the cells were centrifuged at 4000 rpm for 10 min at 4 °C. The supernatant was retained and the pellet was resuspended in an ice-cold 5 mM MgSO<sub>4</sub> solution containing protease inhibitors and 100  $\mu$ g/ml lysozyme (50 ml/L culture). Both supernatants were pooled and the mixture were clarified by centrifugation at 11,000 rpm for 45 minutes. The periplasmic extract was incubated with Ni-NTA Agarose (QIAGEN, UK) for 2 hours at 4 °C. The soluble VH59-scFv protein was purified from the beads using 50 mM Tris, 400 mM imidazole, 500 mM NaCl, pH 8.0. The protein was dialysed into PBS, pH 8.0 using SnakeSkin Dialysis Tubing with a MWCO of 3,500 Da (Thermo Fisher Scientific, UK). The VH59-scFv protein was concentrated to 1.5 ml and purified further by gel filtration using a HiLoad 16/600 Superdex 75 column (GE Healthcare, UK) in PBS, pH 8.0. Fractions containing pure VH59-scFv were pooled and concentrated to 0.2 mg/ml for SPR experiments.

**Surface plasmon resonance.** A BIAcore T200 (GE Healthcare) was used to measure the affinity of the VH59-scFv binding to GST-IBD. A polyclonal goat anti-GST antibody (GE Healthcare) was immobilised on a CM5 sensor chip (GE Healthcare) by the amine coupling method and the GST-IBD protein was immobilized via the antibody. To immobilize the antibody on CM5 chip, the chip was first activated by injecting 100  $\mu$ l mixture of 0.2 M EDC (N-ethyl-N-(dimethylaminopropyl) carbodiimide hydrochloride) and 0.05 M NHS (N-hydroxysuccinimide) at 10  $\mu$ l/min flow rate. 100  $\mu$ g/ml anti-GST antibody in 10 mM sodium acetate, pH 5.0 was injected at 5  $\mu$ l/min for 900 sec and immobilized until 9,000–12,000 RU. After immobilization the chip was immediately inactivated by injecting 1 M ethanolamine, pH 8.5 at 10  $\mu$ l/min for 600 sec. The chip was kept running in the HBS buffer (GE Healthcare) comprising 10 mM HEPES, pH 8.0, 150 mM NaCl. 5  $\mu$ g/ml recombinant GST and GST-IBD proteins were injected into flow cell Fc1 (GST) and Fc2, Fc3 and Fc4 (all GST-IBD). All proteins were trapped through the immobilised anti-GST antibody. The GST-IBD protein was immobilized up to 2,500 RU in Fc2, 4,000 RU in Fc3 and 9,300 in Fc4 via a customized method using The BIAcore evaluation 2.1 software.

The protein was injected at 5  $\mu\text{l}/\text{min}$  for sufficient time to reach the desired level of immobilization. Before the scFv dose response experiment was carried out, the running buffer was replaced by 1X PBS pH 8. A dose response experiment was set up with a series of dilutions of the initial concentration for the VH59-scFv (all in PBS pH 8). The dilution series were 6.1, 3.1, 1.53, 0.76, 0.38, 0.19, 0.095, 0.048, 0.024, and 0.0 uM and in triplicate in a polypropylene 96 well plate. The protein was injected at 30  $\mu\text{l}$  for 50 sec at 20  $\mu\text{l}/\text{min}$  flow rate over flow cells Fc1, Fc2, Fc3 and Fc4. The corresponding kinetic rate constants,  $K_{\text{on}}$  and  $K_{\text{off}}$ , were determined using steady-state affinity and the Fit Kinetic analysis of the compounds' binding affinities, assuming 1:1 ligand-protein stoichiometry.

**Site directed mutagenesis.** Primers for amplification of mutant forms of the anti-IBD iDab VH59 or LEDGF IBD were designed using the GeneArt Primer and Construct Design online tool. Template DNA was amplified with KOD polymerase enzyme (Novagen, Nottingham, UK), PCR products were digested by Dpn I at 37°C for 1 hour and ligated into respective backbone vector followed by transformation into XL-1 blue *E. coli*. Positive clones were confirmed by DNA sequencing (Source Bioscience, Nottingham, UK).

**Lentivirus infection of Jurkat T cells and measurement of p24 viral capsid protein.** Anti-LEDGF VH59 and anti-LMO2 VH576 were cloned into the pFIG lentiviral vector<sup>40</sup> to yield pFIG-VH59 or pFIG-VH576, co-expressing VH and a green fluorescence protein (GFP) reporter. These plasmids or pFIG empty vector were co-transfected with the packaging plasmid p8.91 and the envelope plasmid pMD2.G into HEK293T cells. The transfected cells were cultured for 2 days and lentivirus particles were harvested from the culture supernatant. Jurkat cell lines expressing VH59 or VH576 were established by infecting cells with lentivirus in the presence of 8  $\mu\text{g}/\text{ml}$  protamine. Cells were cultured for 3 hours, fresh medium was added and cultured for further 2 days. Transduced cells expressing GFP were separated by flow cytometry. The GFP-positive Jurkat cells expressing VH were seeded at a density of  $2.5 \times 10^6$  cells/ml in 96-well plates at 40  $\mu\text{l}/\text{well}$ . HIV-1 IIIB virus, was added at serial dilutions (6  $\mu\text{l}$  virus equal to MOI 0.1), mixed with cells and cultured for 90 min. Cells were centrifuged and washed to remove free virus prior to re-suspension in RPMI, 10% FCS. Cells were cultured for 5 days prior to collection of supernatant and measurement of p24 gag protein by ELISA (Advanced BioScience Laboratories Inc).

The co-ordinates of the IBD-VH59 complex have been submitted to the Protein Data base and assigned number 5N88.

## References

- Slany, R. K. The molecular biology of mixed lineage leukemia. *Haematologica* **94**, 984–993, <https://doi.org/10.3324/haematol.2008.002436> (2009).
- Wainberg, M. A. AIDS: Drugs that prevent HIV infection. *Nature* **469**, 306–307, <https://doi.org/10.1038/469306a> (2011).
- Miller, M. M., Liedtke, M. D., Lockhart, S. M. & Rathbun, R. C. The role of dolutegravir in the management of HIV infection. *Infect Drug Resist* **8**, 19–29, <https://doi.org/10.2147/IDR.S58706> (2015).
- Schaefer, K. L. The importance of treatment adherence in HIV. *Am J Manag Care* **19**, s231–237 (2013).
- Rondon, I. J. & Marasco, W. A. Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. *Annu Rev Microbiol* **51**, 257–283, <https://doi.org/10.1146/annurev.micro.51.1.257> (1997).
- Baltimore, D. Gene therapy. *Intracellular immunization*. *Nature* **335**, 395–396, <https://doi.org/10.1038/335395a0> (1988).
- Biocca, S. & Cattaneo, A. Intracellular immunization: antibody targeting to subcellular compartments. *Trends Cell Biol* **5**, 248–252 (1995).
- Christ, F. & Debyser, Z. The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy. *Virology* **435**, 102–109, doi:S0042-6822(12)00483-7 [pii].
- Ciuffi, A. *et al.* A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* **11**, 1287–1289, <https://doi.org/10.1038/nm1329> (2005).
- Llano, M. *et al.* An essential role for LEDGF/p75 in HIV integration. *Science (New York, N.Y.)* **314**, 461–464, <https://doi.org/10.1126/science.1132319> (2006).
- Christ, F. *et al.* Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol* **6**, 442–448, [http://www.nature.com/nchembio/journal/v6/n6/supplinfo/nchembio.370\\_S1.html](http://www.nature.com/nchembio/journal/v6/n6/supplinfo/nchembio.370_S1.html) (2010).
- Christ, F. & Debyser, Z. HIV-1 integrase inhibition: looking at cofactor interactions. *Future Med Chem* **7**, 2407–2410, <https://doi.org/10.4155/fmc.15.162> (2015).
- Demeulemeester, J. *et al.* LEDGINS, non-catalytic site inhibitors of HIV-1 integrase: a patent review (2006–2014). *Expert Opin Ther Pat* **24**, 609–632, <https://doi.org/10.1517/13543776.2014.898753> (2014).
- Higuero, A. P., Jubb, H. & Blundell, T. L. Protein-protein interactions as druggable targets: recent technological advances. *Curr Opin Pharmacol* **13**, 791–796, <https://doi.org/10.1016/j.coph.2013.05.009> (2013).
- Desimie, B. A. *et al.* Phage display-directed discovery of LEDGF/p75 binding cyclic peptide inhibitors of HIV replication. *Mol Ther* **20**, 2064–2075, <https://doi.org/10.1038/mt.2012.132> (2012).
- Tanaka, T. & Rabbitts, T. H. Protocol for the selection of single-domain antibody fragments by third generation intracellular antibody capture. *Nat Protoc* **5**, 67–92, <https://doi.org/10.1038/nprot.2009.199> (2010).
- Berkmen, M. Production of disulfide-bonded proteins in *E. coli*. *Protein Expression & Purif* **82**, 240–251 (2012).
- Cherepanov, P., Ambrosio, A. L. B., Rahman, S., Ellenberger, T. & Engelman, A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17308–17313, <https://doi.org/10.1073/pnas.0506924102> (2005).
- Cherepanov, P. *et al.* HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* **278**, 372–381, <https://doi.org/10.1074/jbc.M209278200> (2003).
- Visintin, M., Tse, E., Axelson, H., Rabbitts, T. H. & Cattaneo, A. Selection of antibodies for intracellular function using a two-hybrid *in vivo* system. *Proc Natl Acad Sci USA* **96**, 11723–11728 (1999).
- Zeng, J., Li, H. C., Tanaka, T. & Rabbitts, T. H. Selection of human single domain antibodies recognizing the CMYC protein using enhanced intracellular antibody capture. *J Immunol Methods* **426**, 140–143, <https://doi.org/10.1016/j.jim.2015.08.009> (2015).
- Serrao, E. *et al.* Fragment-based discovery of 8-hydroxyquinoline inhibitors of the HIV-1 integrase-lens epithelium-derived growth factor/p75 (IN-LEDGF/p75) interaction. *J Med Chem* **56**, 2311–2322, <https://doi.org/10.1021/jm301632e> (2013).
- Dawei, Z., Hongqiu, H., Mengmeng, L., Zhixia, M. & Shunxing, G. A Novel Assay for Screening Inhibitors Targeting HIV Integrase LEDGF/p75 Interaction Based on Ni(2+) Coated Magnetic Agarose Beads. *Scientific reports* **6**, 33477, <https://doi.org/10.1038/srep33477> (2016).
- Stewart, M. P. *et al.* *In vitro* and *ex vivo* strategies for intracellular delivery. *Nature* **538**, 183–192, <https://doi.org/10.1038/nature19764> (2016).

25. Wang, M., Glass, Z. A. & Xu, Q. Non-viral delivery of genome-editing nucleases for gene therapy. *Gene Ther* **24**, 144–150, <https://doi.org/10.1038/gt.2016.72> (2017).
26. Lobato, M. N. & Rabbitts, T. H. Intracellular antibodies as specific reagents for functional ablation: future therapeutic molecules. *Curr Mol Med* **4**, 519–528 (2004).
27. Sewell, H. *et al.* Conformational flexibility of the oncogenic protein LMO2 primes the formation of the multi-protein transcription complex. *Scientific reports* **4**, 3643, <https://doi.org/10.1038/srep03643> (2014).
28. Cherepanov, P. *et al.* Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol* **12**, 526–532, <https://doi.org/10.1038/nsmb937> (2005).
29. Blanco, J. L., Varghese, V., Rhee, S. Y., Gatell, J. M. & Shafer, R. W. HIV-1 integrase inhibitor resistance and its clinical implications. *J Infect Dis* **203**, 1204–1214, <https://doi.org/10.1093/infdis/jir025> (2011).
30. Sutherland, H. G. *et al.* Disruption of Lcdgt/Psip1 results in perinatal mortality and homeotic skeletal transformations. *Mol Cell Biol* **26**, 7201–7210, <https://doi.org/10.1128/MCB.00459-06> (2006).
31. Fadel, H. J. *et al.* TALEN knockout of the PSIP1 gene in human cells: analyses of HIV-1 replication and allosteric integrase inhibitor mechanism. *J Virol* **88**, 9704–9717, <https://doi.org/10.1128/JVI.01397-14> (2014).
32. Huang, J. *et al.* The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* **482**, 542–546 (2012).
33. Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat Rev Drug Discov* **16**, 387–399, <https://doi.org/10.1038/nrd.2016.280> (2017).
34. Marschall, A. L. *et al.* Delivery of antibodies to the cytosol: debunking the myths. *MAbs* **6**, 943–956, <https://doi.org/10.4161/mabs.29268> (2014).
35. Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature*. <https://doi.org/10.1038/nature23305> (2017).
36. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206> (2007).
37. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25, <https://doi.org/10.1107/S0907444909042589> (2010).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501, <https://doi.org/10.1107/S0907444910007493> (2010).
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255, <https://doi.org/10.1107/S0907444996012255> (1997).
40. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science (New York, N.Y.)* **295**, 868–872, <https://doi.org/10.1126/science.1067081> (2002).
41. Lefranc, M. P. Immunoglobulin and T cell receptor genes:IMGT and the birth and rise of immunoinformatics. *Front. Immunol.* **5**, eCollection (2014).
42. Lefranc, M. P. *et al.* IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* **27**, 55–77 (2003).
43. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320–W324, <https://doi.org/10.1093/nar/gku316> (2014).

## Acknowledgements

This work was supported by grants from the Medical Research Council (MR/J000612/1), Bloodwise (12051) and Wellcome Trust (099246/Z/12/Z). We would like to thank Dr. Steve Carr and Dr. Ray Owens (OPPF) for the useful suggestions of protein crystallization and structure determination. We would also like to thank Dr Fusinita van den Ent for the HIV integrase clone.

## Author Contributions

Originator of project & designed the study: T.H.R. Conducted experiments: L.B., C.H., A.C.M., D.P., A.M., C.Q., W.B., M.y.S., M.D., J.S.C. Performed data analysis: L.B., C.H., A.C.M., D.P., J.S.C., A.S., S.E.V.P., T.H.R. Wrote or contributed to the writing of the manuscript: All Authors.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16742-2>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017