

Interrogating HIV integrase for compounds that bind- a SAMPL challenge

Thomas S. Peat · Olan Dolezal · Janet Newman ·
David Mobley · John J. Deadman

Received: 14 November 2013 / Accepted: 25 January 2014 / Published online: 16 February 2014
© Springer International Publishing Switzerland 2014

Abstract Tremendous gains and novel methods are often developed when people are challenged to do something new or difficult. This process is enhanced when people compete against each other-this can be seen in sport as well as in science and technology (e.g. the space race). The SAMPL challenges, like the CASP challenges, aim to challenge modellers and software developers to develop new ways of looking at molecular interactions so the community as a whole can progress in the accurate prediction of these interactions. In order for this challenge to occur, data must be supplied so the prospective test can be done. We have supplied unpublished data related to a drug discovery program run several years ago on HIV integrase for the SAMPL4 challenge. This paper describes the methods used to obtain these data and the chemistry involved.

Keywords HIV integrase · Structure based drug design · SPR · Crystallography · SAMPL challenge

Introduction

The AIDS epidemic has caused over 32 million deaths and over 33 million people are currently infected with HIV

(WHO data, <http://UNAIDS.org>). Small molecule therapeutics against several different protein targets of the HIV virus have been developed over the past two decades [1–4] culminating in a drug combination regimen referred to as Highly Active AntiRetroviral Therapies (HAART) used to treat AIDS. However, these therapies only slow the replication of the virus in patients and new forms of the virus have appeared that are resistant to all the drugs to date [5, 6], so there is a continuing need for new drugs. The integrase enzyme (IN) is critical to the viral life cycle as it is required for the integration of viral DNA into the host chromatin, which in turn is required for formation of new copies of the virus. Integrase performs two catalytic functions termed “3’ processing”-cleaving two nucleotides off of the viral cDNA in a sequence-specific manner to generate “sticky ends” and the “strand transfer reaction”-which covalently attaches, or integrates, the cleaved viral cDNA into human genomic DNA, in a non-sequence-specific manner. There are already three molecules [2] that block the catalytic site in the integrase catalytic core domain (CCD) that have been approved for human use. The structure of HIV integrase has been studied in detail [7–10], and it is most often found as a dimer or tetramer in solution, with the tetrameric form thought to be the active form in the cell. Several pockets in the CCD have been identified to which small molecules have been shown to bind and inhibit enzymatic activity [11–13]. In addition to the CCD, integrase also contains N-terminal and C-terminal domains that are important for DNA binding (to both the viral DNA and cellular DNA).

The integrase protein forms a complex called the pre-integration complex (PIC) with viral DNA and several cellular factors from the host [10] [14]. The cellular protein Lens Epithelium Derived Growth Factor (LEDGF/p75) is normally part of the PIC complex and LEDGF/p75 has a

Electronic supplementary material The online version of this article (doi:[10.1007/s10822-014-9721-7](https://doi.org/10.1007/s10822-014-9721-7)) contains supplementary material, which is available to authorized users.

T. S. Peat (✉) · O. Dolezal · J. Newman
CSIRO, Parkville, VIC, Australia
e-mail: tompeat@hotmail.com

D. Mobley
UC Irvine, Irvine, CA, USA

J. J. Deadman
Chemocopeia Pty Ltd, Melbourne, VIC, Australia

conserved integrase binding domain (IBD, residues 347–429) [15]. LEDGF/p75 facilitates association between the PIC complex with the host chromosome, and in-vitro measures have shown the enhancement of binding of HIV integrase to DNA in the presence of LEDGF/p75 to be 30–50 fold [16]. The PWWP domain (a Pro-Trp-Trp-Pro motif found in proteins of nuclear origin) of LEDGF/p75 has been shown to have a role in localizing the site of integration to genes, such that a PWWP domain deletion mutant leads to a loss of interaction with condensed chromatin and reduced viral replication [17]. When HIV integrase is in a dimeric form, a pocket is created that binds to the IBD loop (residues 362–369). Peptides derived from the sequence of the IBD loop have been shown to bind to HIV integrase with micromolar affinity [15] and to compete with IBD binding. The structures of several peptide complexes have been determined and several key interactions are shared with small molecules that bind in this pocket [18]. In addition, full length integrase and LEDGF/p75 make further interactions through the N-terminal domain of integrase [19, 20]. It has recently been shown, through time of addition studies, that blocking formation of the integrase–LEDGF/p75 complex may prevent effective viral maturation, blocking the formation of competent viral particles [21]. Consistent with the role of LEDGF/p75 in the progression and pathogenesis of HIV is an association that has been shown between polymorphisms in the *psp1* gene that codes for the LEDGF/p75 protein and variation in serologic levels of HIV virus [22].

At the commencement of our project in 2007, we were interested in identifying molecules that bound to any site on the CCD. Indeed as a result of the fragment screen we

identified molecules that bound to the ‘fragment’ pocket [23], another pocket that we described as the Y3 site [13] and the LEDGF/p75 binding site [24] (Fig. 1). This year’s SAMPL4 challenge is based on the compounds that were developed to bind to HIV integrase during the intermediate stage of the program conducted by CSIRO and Avexa Ltd. Various aspects of this program have been published previously [13, 18, 23, 24] but the compounds selected for this challenge are in general not incorporated in these previous publications. The compounds were developed from a fragment screen run against the Maybridge fragment library and compounds were found in multiple sites on HIV integrase.

Experimental methods

N-terminally hexa-His tagged CCD IN (residues 50–210) containing the mutations C56S, F139D and F185H (core3H) was cloned into the *E. coli* expression vector pET28b(+) (Novagen) and expressed and purified with the His-tag retained on the protein. Compounds were tested for affinity using SPR on a GE Healthcare Biacore T200 machine essentially as described previously [23]. Briefly, the ‘minimally biotinylated’ core3H protein was captured onto a Streptavidin chip surface in SPR capture buffer by injecting at 5 μ L/min for 5 min over a single flow cell, typically resulting in immobilization of approximately 2,300 RU of target protein. All SPR binding experiments were performed at 20 °C in SPR binding buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 0.05 % (v/v) Tween20, 5 mM DTT, 10 mM $MgCl_2$, 5% (v/v) DMSO). Small

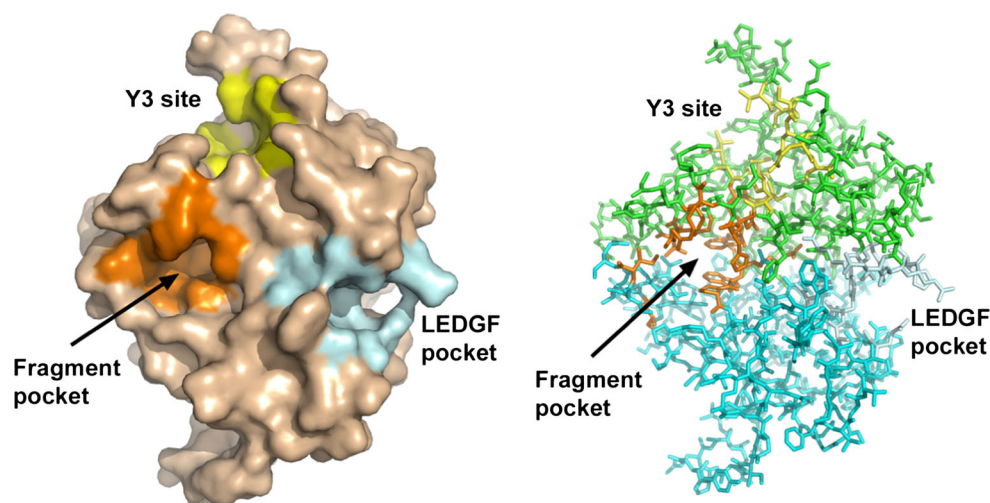


Fig. 1 To the left is a surface representation of the HIV integrase CCD dimer with three of the binding sites highlighted in different colors: the fragment pocket in orange, the Y3 site in yellow and the LEDGF/p75 binding pocket in light blue. To the right in the same orientation is the HIV integrase CCD dimer represented with the

residues as sticks, with the same three sites highlighted in the same colors, but with the individual protomers of the dimer colored in green and cyan. This shows that the fragment pocket and the LEDGF/p75 binding sites are at the dimer interface whereas the Y3 site is found in each of the individual protomers

molecule stock solutions were serially diluted (twofold) in SPR binding buffer and injected for 30 s contact time at 60 $\mu\text{L}/\text{min}$ and then allowed to dissociate for 60 s. The entire concentration series for each compound was performed in triplicate. Raw sensorgram data were processed, solvent corrected and double referenced using Scrubber software (BioLogic Software, Australia). For easy comparison between data sets, all experiments were normalized using a normalization formula of Giannetti et al. [25]. To determine the binding affinity (K_D), responses at equilibrium for each analyte were fitted to a 1:1 steady state affinity model available within Scrubber.

All crystals were grown in the Collaborative Crystallisation Centre (C3, CSIRO) using the core3H CCD construct of the integrase protein, purified using a Ni-column, dialysis and gel filtration. The protein at 5.5 mg/mL concentration was crystallized in 1.6–2.0 M ammonium sulfate in sodium acetate buffer, pH 5.0 to 5.8 at 100 mM concentration. Crystals formed in 3–10 days at 20 °C and were determined to be in spacegroup P31. Crystals were soaked with compounds for 12–48 h prior to data collection at the Australian Synchrotron (at the MX-1 beamline) in a cryo-buffer containing 100 mM sodium acetate pH 5.5, 1.75 M ammonium sulfate, 25% (v/v) ethylene glycol and 5% (v/v) DMSO. Data were indexed using Mosflm [26], scaled using SCALA [27] and molecular replacement was done using Phaser [28]. The models were re-built manually using Coot [29] and refined using Refmac [30]. The compounds were fit into density using Afitt (OpenEye Scientific Software). Some compounds were found, based on inspection of OpenEye Shape toolkit shape overlays onto the refined structures, to be in higher energy conformations, and these were refit into the density and subsequently re-refined (using Refmac with dictionaries from Afitt).

Compounds were considered to be ‘non-binders’ based on two basic criteria: 1) if the binding affinity (K_D) estimated by SPR was greater than 2 mM the compound was considered a non-binder for SPR; 2) if there was no or ambiguous electron density, it was considered to be a non-binder for crystallography. True non-binders had both no electron density and $K_D > 2$ mM. In reality, few compounds that did not bind by SPR were soaked into crystals as the SPR was almost always done prior to the crystallographic work and we tended to focus on the most promising compounds for the project.

Discussion

Background chemistry: Concurrent to our work, several groups designed small molecule inhibitors of the LEDGF-integrase interaction [12, 31–33].

Starting from a catechol scaffold that is well known to chelate metals in the active site of HIV integrase, a series of benzamides were synthesized and led to *N*-(cyclohexylmethyl)-2,3-dihydroxy-5-(piperidin-1-ylsulfonyl) benzamide, that had an IC_{50} of 8 μM and $\text{CC}_{50} > 40$ μM [32]. Recently, activity against infection of whole cells was reported and this compound also bound at the active site and inhibited strand transfer with an IC_{50} of 19 μM [32].

In one study, a pharmacophore based on the pocket of integrase to which the IBD binds was used to screen a virtual set of 160,000 compounds and from further docking results, 25 compounds were purchased and assayed in an AlphaScreen bio-assay [12], leading to the identification of a lead series of 2-(quinolin-3-yl)acetic acids as inhibitors [12]. One compound, CHIBA-3003 [34] [Fig. 2, compound 4], which inhibited the interaction of LEDGF with integrase in an AlphaScreen assay with an IC_{50} of 35 μM was identified from a pharmacophore based on the interactions of the residues of the IBD (I365, D366 and L368). It was predicted from docking that the phenolic hydroxyl of CHIBA-3003 could make a hydrogen bond to the backbone NH of Q168, while the carboxylate made hydrogen bonds to the backbone NHs of E170 and H171. This latter interaction is comparable to that of the side chain carboxylate of the IBD residue D366 as seen in the crystal structure [11].

One lead compound of this series (compound 6* [20] [12], also called CX04328 (compound 1, Fig. 2) had an

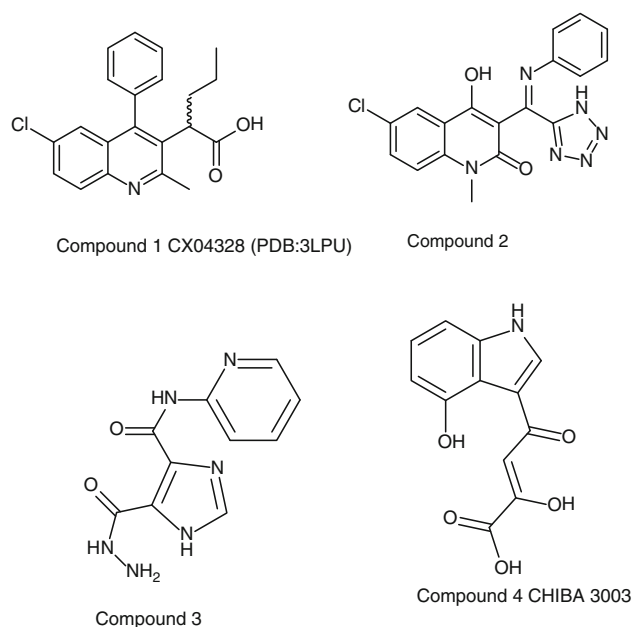
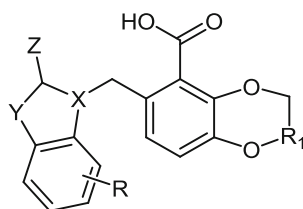


Fig. 2 Compounds described as binding at the LEDGF IBD binding site of HIV integrase, compound 1 CX04328 from Christ 2010 [12] crystal complex PDB 3LPU, compound 2 from De Luca 2011 [33], compound 3 from Serrao 2013 [35], and compound 4 CHIBA3003 from De Luca 2009 [34]

Table 1 Compounds of formula 1, SPR values as determined, compounds defined as binders or non-binders (that is if a crystal complex was obtained)

X=N; Y=CH₂;
 Z = NHCO(CH₂)₂NH₂
 or X=CH₂; Y=NH
 Or X=Y=CH₂

R1=CH₂, R2 absent or
 R1=CH₂, R2=(CH₂)₃CO₂H

Compound #	X	Y	Z	R1	R	Binder	SPR K _D (μM)
101178	N	C=O	C=O	CH ₂	5-CH=CH-CO ₂ H	No	
40872/40869	N	CH ₂	C=O	CH ₂	5-NH ₂	No	
40868	N	CH ₂	C=O	CH ₂	5-Cl	No	
40871/40883	N	CH ₂	C=O	CH ₂	6-Cl	No	
17558	CH ₂	CH ₂	NHCO(CH ₂) ₃ NH ₂	—	—	Yes	
17556	CH ₂	CH ₂	NHCO(CH ₂) ₄ NH ₂	—	—	No	
17557 -17587-101124	CH ₂	CH ₂	NHCO(CH ₂) ₅ NH ₂	—	—	Yes	752 ± 158 (17557) 1460 ± 89 (17587)
101124	CH ₂	CH ₂	NHCO(CH ₂) ₅ NH ₂ ; X=CH ₂ olefin	—	—	Yes	
38747	CH ₂	CH ₂	NHCO(CH ₂) ₅ NH ₂	CHCH ₂ CH ₂ CO ₂ H	—	Yes	
38749	CH ₂	CH ₂	NHCO(CH ₂) ₃ NH ₂	CHCH ₂ CH ₂ CO ₂ H	—	Yes	
38748	CH ₂	CH ₂	NHCO(CH ₂) ₄ NH ₂	CHCH ₂ CH ₂ CO ₂ H	—	Yes	
17375	CH ₂	CH ₂	CH-OH	—	—	Yes	Racemic
40868	N	CH ₂	C=O	—	6-Cl	No	
40883	N	CH ₂	C=O	—	7-Cl	No	
17377	N	CH ₂	C=O	—	H	Yes	
17260	CH ₂	N	C=O	—	H	Yes	
17257	N	CH ₂	C=O	—	H	Yes	No dioxolane ring
101140	N	C=O	C=O	CH ₂	6-Br	Yes	
17286	=CH	NH	C=O	—	H	Yes	E and Z olefin
17285	=CH	NH	C=O	—	H	Yes	No dioxolane ring
38743	=CH	CH ₂	CH-OH	CHCH ₂ CH ₂ CO ₂ H	H	Yes	Racemic at OH
38742	=CH	CH ₂	C=NH-OH	CHCH ₂ CH ₂ CO ₂ H	H	Yes	
101133	C	CH ₂	NHCO-(CH ₂) ₅ NH ₂	CH ₂	H	No	

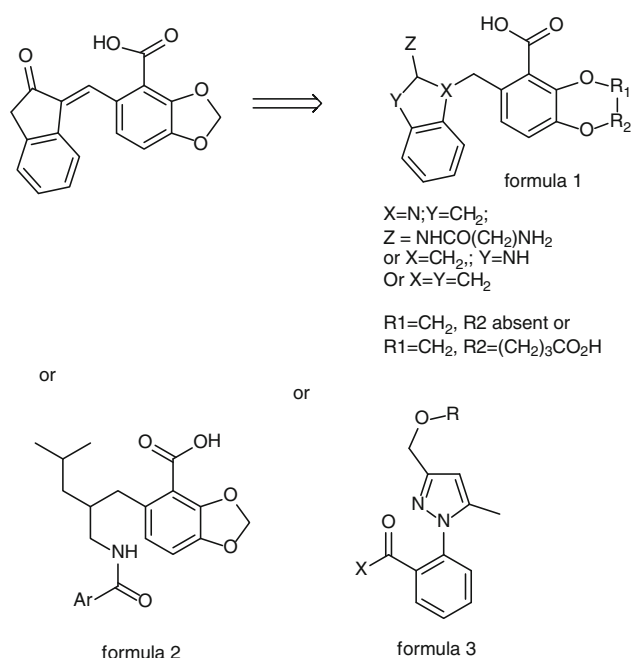


Fig. 3 Evolution of the SAMPL4 project compounds. Identification of dioxolanes from the fragment screen, led to selection of the first compound by shape matching. A series of analogues of this compound of general formula 1 and 2 were synthesized and tested. Compounds of formula 3 were designed as analogues of compounds seen to bind in the fragment pocket

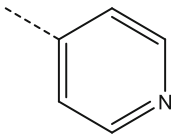
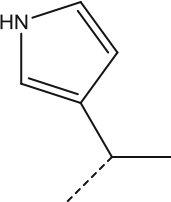
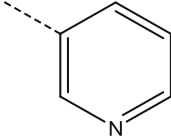
Table 2 Compounds defined as binders or non-binders (binders if a crystal complex was obtained)

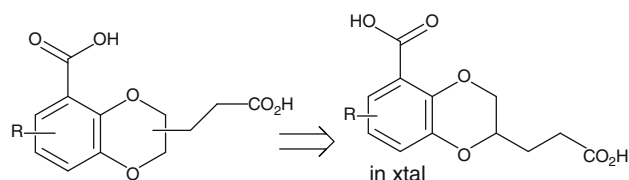
Compound #	R	Binder
40909/40910	R=H; R1=ptBu	No
40988	R=H; R1=3-Cl	No
40911	H	Yes
62772	CH2-phthalimide	No
62777	CH2-NHCO-(4-Pyr)	No
62778	CH2-NHCO-CH2-imidazole	No
101123	R = NHCO-(CH ₂) ₄ -NH ₂	No
101125	R=H; R1=CH ₂ -NH-CO-(4'-pyridyl)	No

Table 3 Compounds of formula 2, compounds defined as binders or non-binders (binders if a crystal complex was obtained)

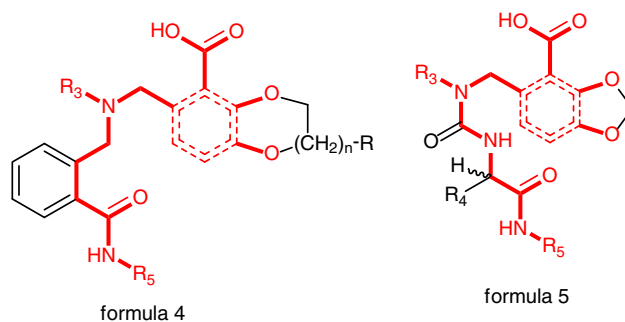
Compound #	R	Binder	SPR K _D (μM)
40920 (GL5243-102)		Yes	
40919 (GL5243-100)		Yes	
40812		Yes	256 ± 12
17542	(CH ₂) ₄ -NH ₂	Yes	
17543	(CH ₂) ₅ -NH ₂	Yes	
17560	(CH ₂) ₃ -NH ₂	Yes	268 ± 28

Table 3 continued

Compound #	R	Binder	SPR K_D (μM)
GL5243104		Yes	432 ± 17
GL5243106	4-(PhNH ₂)	Yes	200 ± 20
GL5243100		Yes	
40811		Yes	481 ± 83

**Fig. 4** The 2,3-dihydrobenzo[b][1,4]dioxine series were synthesized as mixtures of regio-isomers with the propanoic acid chain in the 1 or 2 positions. These mixtures were tested in the bio-assay and soaked into the crystal, however the resulting crystal complexes only contained the 2-substituted compounds (*right figure*), which must preferentially bind in the crystal

IC_{50} of $1.37 \mu\text{M}$ and an EC_{50} of 2.73 or $3.45 \mu\text{M}$ for HIV-1 infection of MT-4 or PBMC cells, respectively. This compound was crystallized with the F185K mutant of the CCD domain of HIV integrase and shown to form hydrogen bonds with residues Glu170, His171 and Thr174 of HIV integrase (PDB: 3LPU) [20] [12]. It was observed that the side chain of A128 packs against the molecule. It is notable that these interactions are similar to those proposed by modelling for CHIBA3003 (compound 4, Fig. 2). It is possible that this compound also binds to the active site of

**Fig. 5** Two further series of compounds were designed and synthesized of formula 4, based on the crystal complex found for compound 17715 (Table 4). The phenyl linker of formula 4 was replaced to try to improve binding giving a series of compounds of formula 5 (Table 5). The two series share a common linker length, as can be seen for the atoms highlighted in red

integrase since it was shown to have an IC_{50} of $19.5 \mu\text{M}$ for inhibition of the strand transfer activity of HIV integrase.

A thieno [2, 3] pyridine series was evolved from this work, with a lead that had an IC_{50} $0.58 \mu\text{M}$ in MT-4 cells, an EC_{50} $0.76 \mu\text{M}$ and a CC_{50} of $72.1 \mu\text{M}$. Further assay of these inhibitors with known mutants of HIV integrase, A128T in HIV-1 and A128M in HIV-2, showed more than 100 fold resistance to these compounds, suggesting that the binding of these involves interactions in addition to those of amino acids 170, 171 and 174 [20]. The compound CHI-1043 inhibited at $0.14 \mu\text{M}$ in the active site directed strand transfer assay, but only $36 \mu\text{M}$ in the AlphaScreen assay for binding at the LEDGF site [31].

High throughput screening identified a series of 5-carbonyl-1H-imidazole-4-carboxamide compounds, and follow-up led to compound 3 (Fig. 2) with an IC_{50} of 400 nM in an AlphaScreen assay, which had no effect on an active site (strand transfer) assay [35], although these compounds were inactive in whole cell assays up to $20 \mu\text{M}$. From docking studies, it was proposed that these compounds formed hydrogen bonding interactions with the backbone NH of E170 and the imidazole nitrogen of H171 and the backbone oxygen of Q168, as well hydrophobic interactions with a T125, A128 and W131 and W132 (Table 1).

History of the Avexa compounds

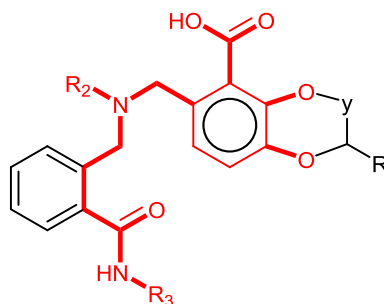
Based on the hits obtained from the initial fragment screen, several analogues were chosen from the CSIRO compound library and these were tested via SPR and crystallography for binding affinity and the location of binding, respectively (see Experimental methods). Our first compound (Fig. 3) was found by SPR to have affinity $750 \mu\text{M}$ for the core3H construct of the CCD. Based on this first compound

we decided to synthesize a series of 1*H*-inden-2(3*H*)-one compounds (of general formula 1, Fig. 3), and observed from the crystallography that in fact the aromatic ring does occupy the same hydrophobic pocket [PDB 3ZT3, Table 2] as residue I365 in the LEDGF/p75 IBD loop [11]. To mimic the interaction of I365 even more closely we designed a series of compounds of formula 2 (Fig. 3).

SAMPL4 compounds

Most of the compounds classified as binders put forward for the SAMPL challenge had both good electron density and consistent binding, as determined by multiple SPR measurements. For the challenge we have used data derived from using the core3H integrase construct, because

Table 4 Compounds of formula 4, compounds defined as binders or non-binders (binders if a crystal complex was obtained)



	R2	R	y	R3		SPR K_D (μ M)
38753	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -Ph	Yes	
38705	Me	H	–	Me	No	
101118	Me	H	CH ₂		Yes	
	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH(Me) ₂ -Ph	Yes	
38781	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -Chx	Yes	
38787	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -Chx	Yes	
	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -Chx		
	Me	–	–	CH ₂ -Chx	3ZSY	>2000 (core3H +/4H)
38780	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ Ph(4-OMe)	Yes	1080 \pm 123
	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ Ph(4-OMe)	3ZSV	>2000 (core3H +/4H)
38786	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ Ph(4-OMe)	3ZSR	?
17715	H	–	–	CH ₂ Ph(4-OMe)	Yes	
	H	–	–	CH ₂ Ph(4-OMe)	3ZT1	>2000 (core3H +/4H)
	Me	–	–	CH ₂ Ph(4-OMe)	3ZTO	595;1180 (core3H/4H)
	Allyl	–	–	CH ₂ Ph(4-OMe)	3ZCM	519;880 (core3H/4H)
40859	Allyl	–	–	CH(Pr)Ph(4-OMe)	3ZSQ	11;74 (core3H/4H)
40860	Allyl	–	–	CH[Ph(4-OMe)] ₂	3ZSO	8;76 (core3H/4H)
38782	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -cyclo-Bu	Yes	
38788	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -cyclo-Bu	Yes	
38783	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -(-2furyl)	Yes	
38789	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ -(-2furyl)	Yes	
38785	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH ₂ Ph	Yes	
38784	Me	(CH ₂) ₂ CO ₂ H	CH ₂	CH(Me)CH ₂ CH ₃	Yes	
	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH(Me)CH ₂ CH ₃	3ZSW	>2000 (core3H/4H)
	Me	–	–	CH(Me)CH ₂ CH ₃	3ZSZ	>2000 (core3H/4H)
38779	Allyl	(CH ₂) ₂ CO ₂ H	CH ₂	CH(Me)CH ₂ CH ₃	Yes	
38708	Me	–	–	Ch ₂ Ph	Yes	
	Allyl	–	–	CH ₂ Ph	3ZSX	1755;1065 (core3H/4H)

Table 4 continued

	R2	R	y	R3	SPR K_D (μ M)
101121				Yes	214 \pm 35
101122				Yes	
101119				Yes	793 \pm 50

Red PLOS one paper; blue means in PLOS one and SAMPL

the alternative core4H construct has an additional Y131D mutation in the wall of the LEDGF/p75 pocket, which we had previously shown can modulate binding of compounds at this site [24]. Those compounds that did not show binding at better than 2 mM in the SPR assay were determined to be ‘non-binding’ and used as the non-binding controls for the challenge. SPR binding was not perfectly correlated with crystallographic binding. The experiments are run under different conditions and at different pHs, so we would not expect perfect correlation (see Experimental methods). We found that as the affinities improved (low micromolar instead of high micromolar or low millimolar affinities) that the correlation between the different methods improved. Some compounds had limited solubility, which can have different effects on the SPR and crystallographic experiments.

There were cases, particularly where an aliphatic moiety was present (e.g. AVX17557), when the density for part of the compound was not as strong as for the rest of the compound. Enantiomeric compounds were generally not separated prior to either SPR or crystallographic experiments. This has potential implications for the determined binding affinity (with a single stereocenter, it can be up to 2 \times better than stated if one enantiomer of a pair binds and the other has no or very low affinity), but it generally does not for the crystallographic experiments as the resolution was high enough to determine which enantiomer bound or if there was binding of both (e.g. AVX38743).

We started off with about 70 compounds that had reasonable crystallographic density and that had been run through SPR and bioassays and hadn’t been previously published. This list was cut down to those that had the best

Fig. 6 AVX38741 was observed in the crystal complex with integrase to be the compound on the *right hand side* (Table 4)

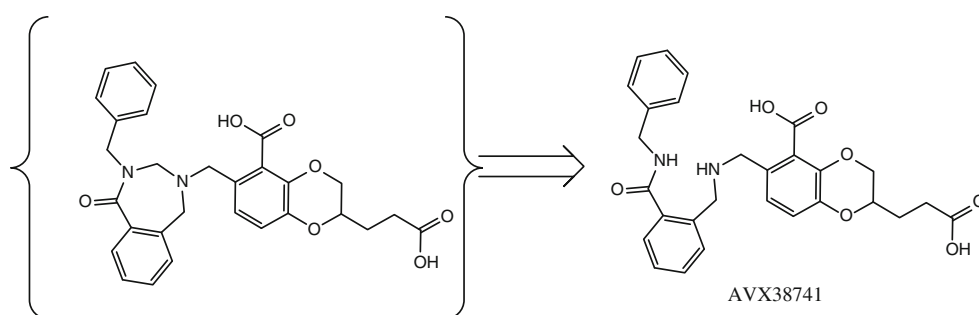
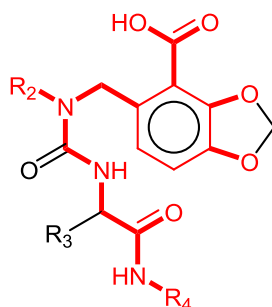


Table 5 Compounds of formula 5, compounds defined as binders or non-binders (binders if a crystal complex was obtained)



Compound #	R2	R3	R4	Binder
38669	Me	Ph	R ₄ = H	No
38673	Me	Ph	nBu	Yes
17628	H	Ph	CH ₂ Ph	No
17629 (Et ester)	H	Ph	CH ₂ Ph	No
38669	Me	Ph	H	No
38670	Me	Ph	CH ₂ Ph(pOMe)	No
38671	Me	Ph	Me	No
38672	Me	Ph	tBu	Yes
38674	Me	Ph	CH ₂ Ph	Yes
17627	H	CHPh ₂	–	No

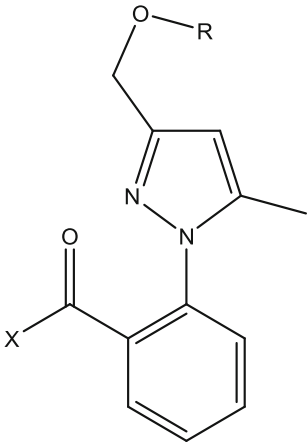
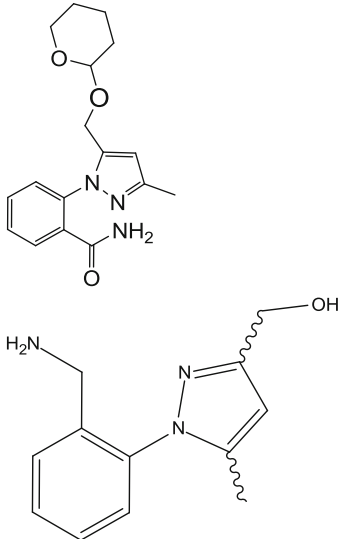
data (multiple SPR experiments, the best electron density, and therefore the most confident data) leaving 58 compounds. Some of the compounds had different names but were in fact the same (e.g. compounds that were made as different salts were given different names, such as AVX17557 and AVX17587) giving 57 for the SAMPL4 challenge.

As generally observed in our work (and that from other groups), almost all small molecules found to date (see Figs. 2, 3, 4, 5 and 6) have a carboxylic acid that makes a virtually identical interaction to E170 and H171 as is seen with D366 of the IBD [11]. This charge interaction is key to the series developed here, to the series of peptides that have been shown to interact with the LEDGF binding site on HIV integrase, and key for the binding of other small

molecules that have been developed by other groups. This was the basis for our first analogue compound #3 in Ref. [24], which is of formula 1, Fig. 3 (this paper, X is C=, Y=CH₂, Z=CH–OH), which demonstrated similar affinity to compound 1 (compound #2 in Ref. [24]) in the AlphaScreen assay (AS) (270 μM and 200 μM respectively, Table 1 of Ref. [24]), and the SPR core3H assay (1,435 μM vs. 1,375 μM, Table 1 of Ref. [24]) and clear density in the LEDGF/p75 site.

Modelling suggested that compounds of formula 1 (Fig. 3) with a methylene linker would tend to favor conformations with the plane of the two bicyclic rings at an angle to each other. To allow greater flexibility, a longer alkyl linker was introduced into the series of formula 2 (Fig. 3; Table 3). All of the compounds of this series tested

Table 6 Compounds of formula 3. Compounds defined as binders or non-binders (binders if a crystal complex was obtained)

Compound #	R	Binder
17384	X = NH ₂ ; R = tetrahydropyran	No
17285	X = NH ₂ ; R = H	Yes
17384	X = NH ₂ , R = tetrahydropyran	No
17288	X = OH, R = tetrahydropyran	No
17287		Yes
		
17379		Yes

were found to give a crystal complex with integrase (Table 3) and some had measurable activity in the AS assay (e.g. <0.3 mM for AVX17560, Table 3).

Based on fragments seen to bind in the fragment pocket of integrase [23], a series of compounds of formula 3 (Fig. 3; Table 6) were synthesized and tested. Some compounds of this series, such as AVX17287, had partial occupancy of both the fragment and LEDGF binding sites, whereas AVX17379 was observed in the crystals only in the LEDGF site. In comparison, AVX17285, which as an amide differs only from the amine compound AVX17379

by having the side chain carbonyl, was only observed with weak density at either site.

For those compounds with the 2,3-dihydrobenzo[b][1,4]dioxine ring (Fig. 4), the synthesis from the catechol gave a mixture of regioisomers, and these were tested as mixtures in the SPR assay. However in most cases where a crystal complex was formed, the 2-alkyl series was seen to preferentially bind (see Fig. 4; Tables 1, 4). The density was clear in several cases for the alkyl chain with carboxylic acid (e.g. AVX38747) but there were also cases where the density was unclear (e.g. AVX38743). We had hoped that a carboxylate in this position would interact with K173 but this was not observed.

As described previously [24] synthesis unexpectedly gave a series of compounds, including AVX17715 (of formula 4, Fig. 5; Table 4) that were observed to bind to the LEDGF/p75 site. It was observed that the R5 group (formula 4, Fig. 5) was now in proximity of the hydrophobic site formed by W131 in the LEDGF/p75 site. Also the amide nitrogen of these molecules typically took part in a hydrogen bond with the backbone carbonyl of Q168 of the HIV integrase. The SAR for this series indicated that higher affinity compounds had an alkyl R3 group and so we considered that modulating the pKa and/or hydrogen donating capability of this nitrogen might influence binding, so a series of urea linked molecules (formula 5, Fig. 5 and Table 4) was prepared. Also the 1,2-disubstituted phenyl ring of the compounds of formula 4 sat above the protein backbone, not making any clear interactions and so we replaced this by an R4 group in the series of formula 5.

It is worth noting that our original hypothesis was that a 7 membered ring could be modelled to occupy the hydrophobic site occupied by I366 of the IBD and subsequently compound AVX38741 was found to bind, however as observed before, it was the ring open form [24] (Fig. 6, right hand side) that was found in the crystal structure.

These data were given to the challenge organisers and were subsequently both manually and computationally vetted prior to inclusion in the challenge.

One of the more intriguing aspects to this particular challenge is that there were multiple binding sites for molecules to bind. This feature alone distinguishes this test from many others. Another distinguishing feature is that the non-binders looked very similar to the binders, many being in the same series and others using very similar scaffolds. We speculate that the fragment pocket was basically too small to bind much except fragments and a few promiscuous compounds. The Y3 pocket is a long shallow groove that picked up some promiscuous binders, but we couldn't determine features that would give specificity to this site alone. The majority of the compounds bind in the LEDGF pocket as that is what we decided to

Table 7 Miscellaneous compounds. Compounds defined as binders or non-binders (binders if a crystal complex was obtained)

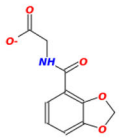

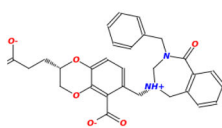
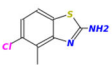

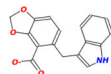
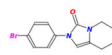
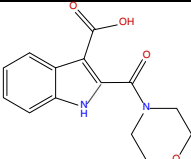
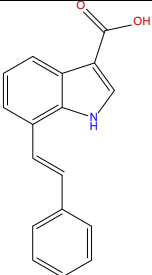
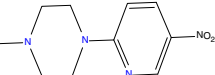
17389		yes
17679		yes
38741		yes
15988		yes
17631		yes
17287		yes
pC2A03		yes
17386		no
17387		no
17228		no

Table 7 continued

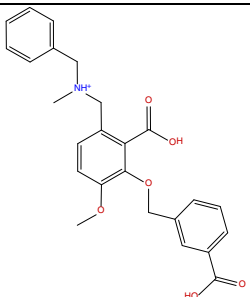
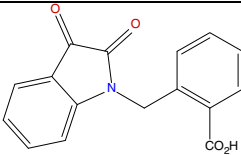
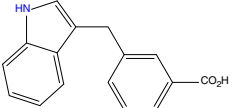
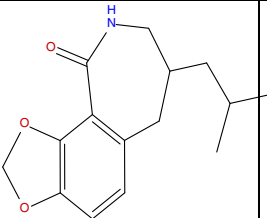
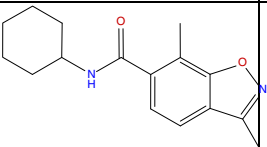
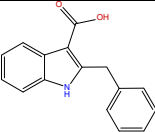
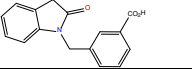
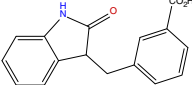
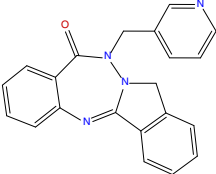
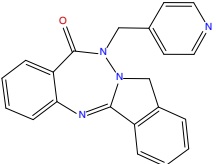
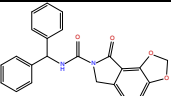
40944		no
16984		no
16981		no
17544		no
38710		no
17385		no
17258		no
17259		no
40816		no
40813		no
17552		no

Table 7 continued

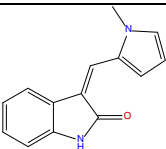
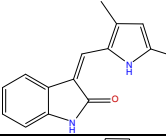
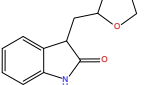
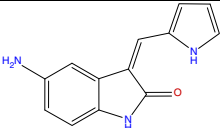
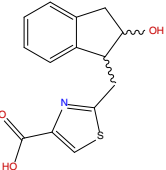
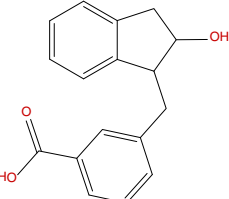
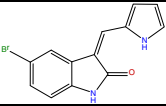
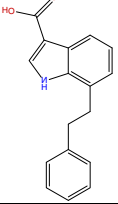
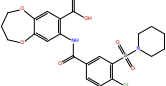
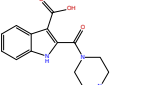
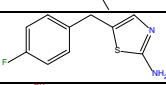
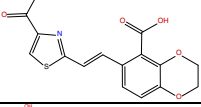
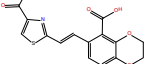
40801		no
62524		no
62525		no
62526		no
17482		no
17369		no
62528		no
17476		no
16757		no
17478		no
17479		no
62730		no
17378		no

Table 7 continued

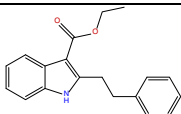
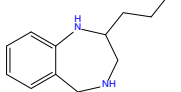
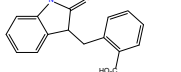
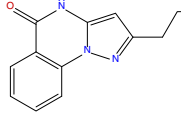
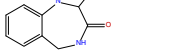
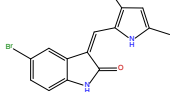
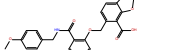
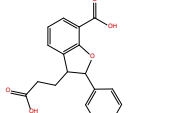
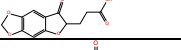
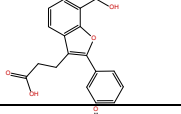
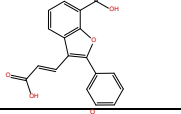
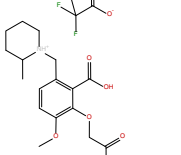
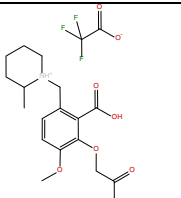
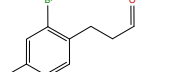
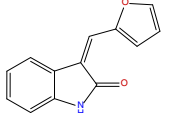
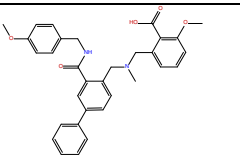
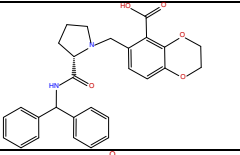
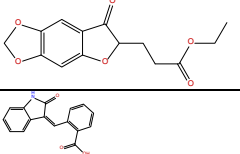
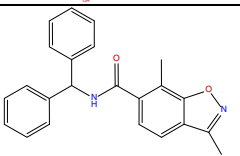
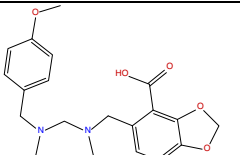
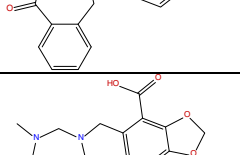
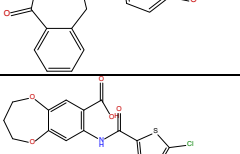
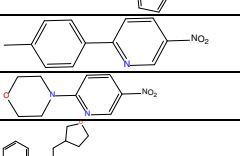
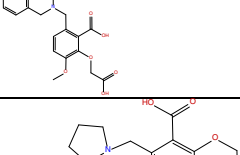
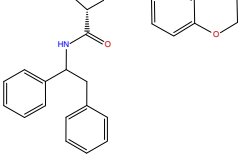
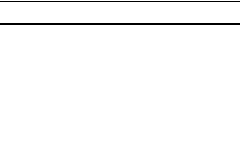

17370		no
17264		no
17267		no
17266		no
17261		no
62530		no
40861		no
38711		no
17269		no
38719		no
38791		no
40916		no
40915		no
17734		no
16688		no

Table 7 continued

101120		no
62808		no
17280		no
17284		no
38709		no
17687		no
17685		no
16756		no
17233		no
17231		no
101139		no
62807		no

target after the fragment screening and it was both large enough and had enough potential interactions that we could get specific binding of compounds.

Acknowledgments DLM acknowledges the financial support of the National Institutes of Health (1R15GM096257-01A1), and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513. The work was supported by a commercial ready grant COMO4229 from the Commonwealth of Australia, Department of Innovation, Industry, Science and Research to Avexa Pty Ltd. We thank the Australian Synchrotron and the beamline scientists at MX1 and MX2 for their help in data collection; and the Collaborative Crystallisation Centre for producing the protein crystals used in this study. Compounds in this study were prepared for Avexa Pty Ltd by Dr G. Le (now at Alchemia Pty Ltd), Dr J.A. Smith (now at F.B. Rice Pty Ltd), SYNthesis med chem Pty Ltd, Dr J. Spencer (now at the University of Sussex, UK) and Dr H. Patel at the time at the University of Greenwich, UK. We thank OpenEye Scientific Software for a license to their software.

References

- Arribas JR, Eron J (2013) Advances in antiretroviral therapy. *Curr Opin HIV AIDS* 8(4):341–349
- Metifiot M, Marchand C, Pommier Y (2013) HIV integrase inhibitors: 20-year landmark and challenges. *Adv Pharmacol* 67: 75–105
- Messiaen P et al (2013) Clinical use of HIV integrase inhibitors: a systematic review and meta-analysis. *PLoS ONE* 8(1):e52562
- Flexner C, Saag M (2013) The antiretroviral drug pipeline: prospects and implications for future treatment research. *Curr Opin HIV AIDS* 8(6):572–578
- Geretti AM, Armenia D, Ceccherini-Silberstein F (2012) Emerging patterns and implications of HIV-1 integrase inhibitor resistance. *Curr Opin Infect Dis* 25(6):677–686
- Quashie PK, Mesplede T, Wainberg MA (2013) Evolution of HIV integrase resistance mutations. *Curr Opin Infect Dis* 26(1):43–49
- Jenkins TM et al (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci USA* 92(13):6057–6061
- Goldgur Y et al (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci USA* 95(16):9150–9154
- Chen JC et al (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* 97(15):8233–8238
- Cherepanov P et al (2005) Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nat Struct Mol Biol* 12(6):526–532
- Cherepanov P et al (2005) Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc Natl Acad Sci USA* 102(48):17308–17313
- Christ F et al (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat Chem Biol* 6(6):442–448
- Rhodes DI et al (2011) Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design. *Antivir Chem Chemother* 21(4):155–168
- Bukrinsky MI et al (1992) Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci USA* 89(14):6580–6584
- Tsiang M et al (2009) Affinities between the binding partners of the HIV-1 integrase dimer-lens epithelium-derived growth factor (IN dimer-LEDGF) complex. *J Biol Chem* 284(48):33580–33599
- McNeely M et al (2011) In vitro DNA tethering of HIV-1 integrase by the transcriptional coactivator LEDGF/p75. *J Mol Biol* 410(5):811–830
- Gijsbers R et al (2011) Role of the PWWP domain of lens epithelium-derived growth factor (LEDGF)/p75 cofactor in lentiviral integration targeting. *J Biol Chem* 286(48):41812–41825
- Rhodes DI et al (2011) Crystal structures of novel allosteric peptide inhibitors of HIV integrase identify new interactions at the LEDGF binding site. *ChemBioChem* 12(15):2311–2315
- Maertens G et al (2003) LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* 278(35):33528–33539
- Hare S et al (2009) A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog* 5(1):e1000259
- Balakrishnan M et al (2013) Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. *PLoS ONE* 8(9):e74163
- Madlala P et al (2011) Association of polymorphisms in the LEDGF/p75 gene (PSIP1) with susceptibility to HIV-1 infection and disease progression. *AIDS* 25(14):1711–1719
- Wielens J et al (2013) Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification? *J Biomol Screen* 18(2):147–159
- Peat TS et al (2012) Small molecule inhibitors of the LEDGF site of human immunodeficiency virus integrase identified by fragment screening and structure based design. *PLoS ONE* 7(7): e40147
- Giannetti AM, Koch BD, Browner MF (2008) Surface plasmon resonance based assay for the detection and characterization of promiscuous inhibitors. *J Med Chem* 51(3):574–580
- Leslie AW, Powell H (2007) Processing diffraction data with mosflm. In: Read R, Sussman J (eds) *Evolving methods for macromolecular crystallography*. Springer, Netherlands, pp 41–51
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72–82
- McCoy AJ et al (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674
- Emsley P et al (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255
- De Luca L et al (2011) 4-[1-(4-Fluorobenzyl)-4-hydroxy-1H-indol-3-yl]-2-hydroxy-4-oxobut-2-enoic acid as a prototype to develop dual inhibitors of HIV-1 integration process. *Antiviral Res* 92(1):102–107
- Fan X et al (2011) Design of HIV-1 integrase inhibitors targeting the catalytic domain as well as its interaction with LEDGF/p75: a scaffold hopping approach using salicylate and catechol groups. *Bioorg Med Chem* 19(16):4935–4952
- De Luca L et al (2011) Inhibitors of the interactions between HIV-1 IN and the cofactor LEDGF/p75. *ChemMedChem* 6(7): 1184–1191
- De Luca L et al (2009) Pharmacophore-based discovery of small-molecule inhibitors of protein-protein interactions between HIV-1 integrase and cellular cofactor LEDGF/p75. *ChemMedChem* 4(8):1311–1316
- Serrao E et al (2013) Discovery of a novel 5-carbonyl-1H-imidazole-4-carboxamide class of inhibitors of the HIV-1 integrase-LEDGF/p75 interaction. *Bioorg Med Chem* 21(19):5963–5972