Author/s:
Northfield, SE;Wielens, J;Headey, SJ;Williams-Noonan, BJ;Mulcair, M;Scanlon, MJ;Parker, MW;Thompson, PE;Chalmers, DK

Title:
Cyclic Hexapeptide Mimics of the LEDGF Integrase Recognition Loop in Complex with HIV-1 Integrase

Date:
2018-08-10

Citation:
Northfield, S. E., Wielens, J., Headey, S. J., Williams-Noonan, B. J., Mulcair, M., Scanlon, M. J., Parker, M. W., Thompson, P. E. & Chalmers, D. K. (2018). Cyclic Hexapeptide Mimics of the LEDGF Integrase Recognition Loop in Complex with HIV-1 Integrase. Chemmedchem, 13 (15), pp.1555-1565. https://doi.org/10.1002/cmdc.201800129.

Author Manuscript

Title: Cyclic Hexapeptide Mimics of the LEDGF Integrase Recognition Loop in Complex with HIV-1 Integrase

Authors: Susan Northfield, PhD; Jerome Wielens, Phd; Stephen Headey, PhD; Billy Williams-Noonan, BSc; Mark Mulcair, PhD; Martin Scanlon, PhD; Michael Parker, PhD; Philip Thompson, PhD; David Chalmers, PhD

To be cited as: 10.1002/cmdc.201800129

Link to VoR: https://doi.org/10.1002/cmdc.201800129

Cyclic Hexapeptide Mimics of the LEDGF Integrase Recognition Loop in Complex with HIV-1 Integrase.

Susan E. Northfield[a], Jerome Wielens[a,b], Stephen J. Headey[a], Billy J. Williams-Noonan[a], Mark Mulcair[a], Martin J. Scanlon[a], Michael W. Parker[b,c]*, Philip E. Thompson[a], David K. Chalmers[a]*

[a] Dr S. E. Northfield, Dr J. Wielens, Dr S. J. Headey, Mr B. Williams-Noonan, Dr M. Mulcair, Prof M. J. Scanlon, Prof P. E. Thompson, Dr D. K. Chalmers
Medicinal Chemistry
Monash Institute of Pharmaceutical Sciences
381 Royal Pde, Parkville, Victoria 3052, Australia
E-mail: david.chalmers@monash.edu

[b] Dr J. Wielens, Prof M. W. Parker
ACRF Rational Drug Discovery Centre
St. Vincent's Institute of Medical Research
41 Victoria Parade, Fitzroy, Victoria 3065, Australia
E-mail: mparker@svi.edu.au

[c] Prof M. W. Parker
Department of Biochemistry and Molecular Biology
Bio21 Molecular Science and Biotechnology Institute
University of Melbourne
30 Flemington Road, Parkville, Victoria 3010, Australia

Supporting information for this article is given via a link at the end of the document.

Abstract: The p75 splice variant of lens epithelium-derived growth factor (LEDGF) is a 75 kDa protein, which is recruited by the human immunodeficiency virus (HIV) to tether the pre-integration complex to the host chromatin and promote integration of proviral DNA into the host genome. We have designed a series of small cyclic peptides that are structural mimics of the LEDGF binding domain, which interact with integrase as potential binding inhibitors. Here we present the X-ray crystal structures, NMR studies, SPR analysis and conformational studies of four cyclic peptides bound to the HIV-1 integrase core domain. Although the X-ray studies show that the peptides closely mimic the LEDGF binding loop, the measured affinities of the peptides are in the low mM range. Computational analysis using conformational searching and free energy calculations suggest that the low affinity of the peptides is due to mismatch between the low-energy solution and bound conformations.

Introduction

The replication cycle of human immunodeficiency virus-1 (HIV-1) integrase is a complex, multistep process. One key event in replication is integration, where the proviral DNA is spliced into DNA of the host cell within the nucleus. Integration is mediated by the viral enzyme HIV integrase, which performs 3'-processing of the viral DNA to produce a sticky end and integration; insertion of the viral DNA into that of the host.[1] Integrase is also is a key factor in the transport of the HIV pre-integration complex (PIC) into the cell nucleus and localization of the PIC on the host DNA.[2] Recently, it has also become evident that integrase plays an important role in viral packaging, although the details of this are not yet well understood.[3]

Integrase is a validated anti-HIV drug target. Four compounds are currently approved to inhibit integrase strand transfer; raltegravir[4], elvitegravir[5],, dolutegravir[6] and bictegravir.[7] These compounds bind to the active site region of the integrase/DNA complex and inhibit 3'-processing and strand transfer[8]. The polyfunctional nature of integrase suggests that a variety of approaches might be used to target integrase function and there is, accordingly, considerable interest in developing new compounds that inhibit integration through binding to alternative sites on integrase and operating by allosteric inhibition or through inhibition of interactions with other cellular proteins. Several potentially druggable sites have been identified to date; these include the binding site for the cellular protein lens epithelium-derived growth factor (LEDGF/p75), which is discussed further below: a site which can accommodate a sucrose molecule that was identified by protein crystallography[9] (the 'sucrose pocket'), a binding pocket found by fragment screening[10] (the 'fragment pocket'), and a site adjacent to the flexible active site loop.[11]

LEDGF is a cellular protein that is recognised by integrase to assist importation of the PIC into the cell nucleus and targeting of the PIC to regions of DNA undergoing active transcription.[2b, 12] The nature of the integrase/LEDGF interaction has been characterised through X-ray crystallography and a variety of peptide structures have been reported of the LEDGF integrase-binding domain (IBD, residues 347–429) bound to the HIV-1 integrase core domain (residues 50–212),[13] the HIV-2 integrase N-terminal plus core domain[14] (residues 1–212) and also the homologous Visna virus integrase[14] (residues 1–219). In each case, the principal recognition of LEDGF occurs at a cleft present at the integrase dimer interface and a reverse-turn present in residues 364–367 (KIDN) located at the tip of a loop on the LEDGF IBD. Additional hydrophobic interactions are made between the LEDGF IBD and HIV-1 integrase residues W131, K127, T124 and A128. Further hydrophobic interactions are also present in the HIV-2 core+N-terminal domain crystal structure. Figure 1A shows the interaction of LEDGF with one of two symmetry-related sites in the crystal structure 2B4J[13] with the integrase and LEDGF residues coloured according the area of the contact interface. The contacting residues are shown in Figure 1B. Contact is made through two epitopes of LEDGF; residues 360–370, which have a buried area of 488 Å^2 and form a type I beta-turn and residues 402–408 with an area of 198 Å^2. LEDGF residues K364, I365, D366 and N367 make most of the

first epitope contact, totalling 393 Å^2. Within this stretch of four residues, LEDGF D366 makes bidentate hydrogen bonds to the backbone amide NH groups of integrase residues E170 and H171. The LEDGF binding pocket is also a target for small molecule inhibitors.[15]

The large contact area of LEDGF residues 364–367 raises the question, how accurately can the integrase/LEDGF interaction be replicated using small cyclic peptides? Structural data from such an investigation would provide valuable information about this PPI that could be subsequently utilised in the design and development of peptide and/or small molecule inhibitors. Peptides have previously been investigated as inhibitors of the of the integrase/LEDGF interaction. Linear peptides derived from the LEDGF binding loop [16] and cyclic[17] have been found to have modest affinities at the LEDGF site. A cyclic octapeptide based on the LEDGF sequence,[17c] cyclo(SLKIDNLD), has an activity of 85.2 µM in an AlphaScreen assay. Single residue modifications of the native sequence were only able to increase activity to 39.7 µM. While key binding interactions between integrase and I365 and D366 of the peptide residues were maintained, these cyclic peptides did not make any notable additional interactions with the binding pocket relative to the native protein.

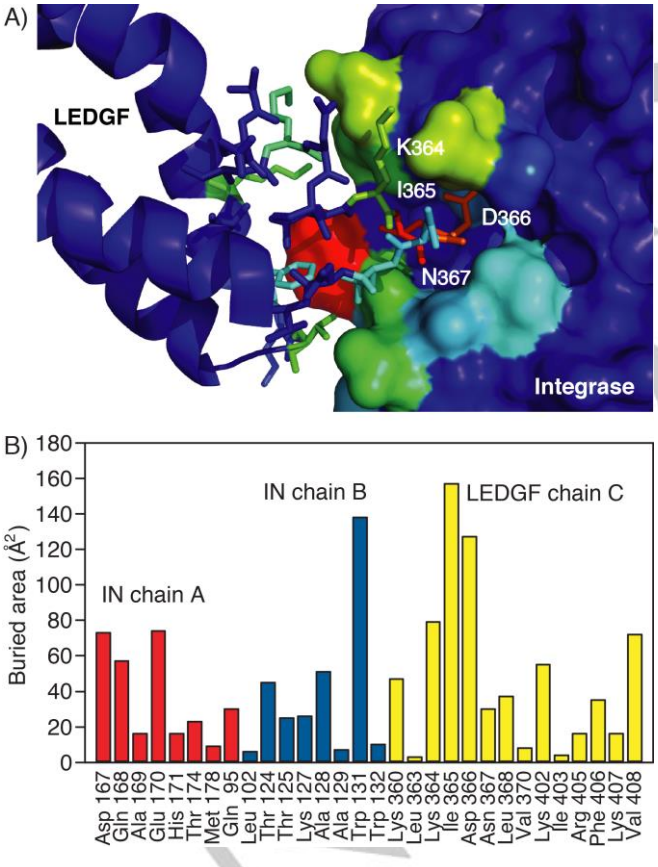


Figure 1. A) The interaction interface between LEDGF (ribbon with contacting residues shown as sticks) and HIV integrase (surface representation) in the crystal structure 2B4J. Residues are coloured in a spectrum according to contact area from red (most) to dark blue (zero). B) The amount of surface area buried in the 2B4J integrase/LEDGF interface by residue. Graph bars are coloured by protein chain.

Cyclic peptides offer us the ability to mimic the endogenous reverse-turn that LEDGF presents to the binding site on integrase (KIDN), and also to potentially improve potency and stability by increasing the structural rigidity of the peptides. However, structurally these cyclic octapeptides are interesting, as they do not fully maintain the backbone conformation of the native LEDGF.

In this work we set out to design reverse-turn mimetics of the key surface loop of LEDGF to better understand the structural requirements of this recognition process and to assist the development of small-molecule inhibitors of the interaction. Ultimately, our challenge was to make cyclic peptides that maintain the key interactions observed in the crystal structures of the integrase and the LEDGF IBD using a minimal number of residues. One simple way to induce a reverse-turn is the inclusion of a disulfide bond. Alternatively, reverse-turns can be induced in cyclic peptides by the incorporation of dipeptide scaffolds, placed external to the i→i+4 tetrapeptide (Figure 2).[18] Here we report the design, synthesis and structural analysis of a series of cyclic hexapeptides that utilise varied reverse-turn inducing dipeptide scaffolds to mimic the conformation of the LEDGF residues 364–367.

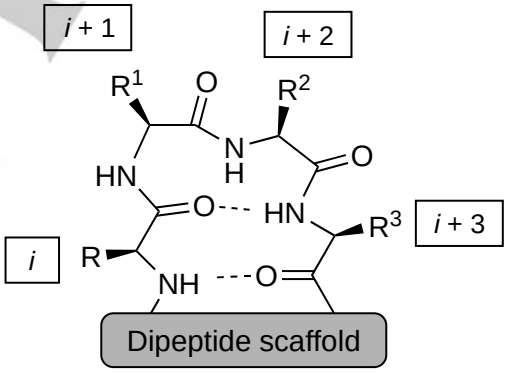


Figure 2. The approach used to design the reported peptides. The key residues in the tetrapeptide are constrained into a reverse-turn by the incorporation of a turn-promoting dipeptide scaffold. Intramolecular hydrogen bonds are shown as dashed lines.

Results and Discussion

As a first approach to mimicking the LEDGF binding loop, we cyclised the KIDN tetrapeptide by introducing terminal cysteine residues, which were oxidised to form a disulfide bridge giving ox(Ac-CKIDNC-NH_2) (**1**). The N-terminus was acylated and the and C-terminus was amidated to more closely approximate the backbone of the native protein. A similar cyclic peptide, with free N- and C-termini, has been previously reported to have a K_i of 180 nm for inhibition of LEDGF binding to integrase measured using a time-resolved fluorescence assay.[17a]

Many turn-inducing residues can be incorporated into cyclic peptides to manipulate the structure into a reverse-turn confirmation.[19] Using readily available amino acid building blocks compatible with automated peptide synthesis, we designed multiple turn-promoting dipeptides, comprised of a D-amino acid coupled with a proline.[20] Our design strategy utilised a dipeptide scaffold in the form of D-Xaa-L-Pro, where Xaa represents a naturally occurring amino acid. This approach has been previously reported, with particular emphasis on D-Pro-L-Pro as a scaffold that has been shown to induce a type-II' ß-hairpin turn.[21] The dipeptide scaffold was combined with the KIDN tetrapeptide to form cyclic hexapeptides (Figure 2). We wanted the ability to introduce a scaffold at the non-interface side of the cyclic peptide and leave the binding sequence unaltered. We accordingly designed the backbone cyclised peptides cyclo(PKIDNp) (**2**) using a D-Pro-L-Pro scaffold. The crystal structure of the integrase and LEDGF IBD shows that I366 of LEDGF binds in hydrophobic pocket. In order to exploit this, we incorporated norleucine (Z) into the Ile position with a D-Val-L-Pro scaffold to give cyclo(PKZDNv) (**3**). Further, we designed cyclo(PKIDNG) (**4**), which incorporates a glycine residue to increase the flexibility of the scaffold.

Peptide synthesis

The linear peptides required for cyclisation were prepared using standard Fmoc solid phase peptide synthesis. Linear Ac-CKIDNC-NH_2 was oxidised in ammonium bicarbonate to form the cyclic disulfide (**1**). The linear precursors to the head-to-tail cyclic peptides were synthesised on 2-chlorotrityl resin, allowing the peptides to be cleaved from the resin using mildly acidic conditions while retaining side-chain protecting groups. The cyclisation conditions reported here were found to be the most efficient for these peptide sequences after extensive exploration of alternative solvents, bases, cyclising reagents and reaction temperatures. All cyclic peptides generated were purified by RP-HPLC and obtained in > 95% purity.

X-ray crystallography

The cyclic peptides (**1**–**4**) were soaked into crystals of the HIV-1 integrase core domain and their X-ray structures were solved using molecular replacement. A summary of the data collection and refinement statistics is shown in Table 1. The four structures were refined to a resolution of 1.45–1.75 Å, with R_{factor} ranging from 17.9–20 % and R_{free} 22.2–23.5 %. Consistent with previously published data,[9] the integrase core domain is dimeric with a pseudo two-fold symmetry. In each structure, one copy of the cyclic peptide is bound at each of the two equivalent LEDGF IBD binding sites at the integrase dimer interface. Importantly, the structures show that the presentation of the native KIDN tetrapeptide sequence have mimicked the structure of the native binding interaction, achieving the primary objective for these truncated LEDGF mimetics. The peptide aspartate side-chain carboxylate forms critical hydrogen bonds with the backbone amide NH atoms of E170 and H171 and the side-chain hydroxyl of T174 of integrase. The peptide isoleucine side-chain is buried in the hydrophobic pocket of the enzyme, comprised of residues L102, A128, W132, A169, T174 and M178. The dipeptide scaffolds of compounds **2**, **3** and **4** do not make any binding interactions with the protein; rather they point toward the solvent and away from integrase. This is an interesting observation, as it appears that the reduction in size of the LEDGF mimetics to six residues serves to conserve only the key binding interactions of the KIDN tetrapeptide, in part by preventing potential conformational changes that could arise if the scaffolds were also attempting to bind to the enzyme.

Table 1. Summary of X-ray data processing and refinement statistics.

Data Collection	*$IN_{CORE4H123}$/ ox(Ac-CKIDNC-NH_2)*	*$N_{CORE4H123}$/ cyclo(PKIDNp)*	*$IN_{CORE4H123}$/ cyclo(PKZDNv)*	*$IN_{CORE4H123}$/ cyclo(PKIDNG)*
PDB ID	*3WNF*	*3WNG*	*3WNH*	*3WNE*
Spacegroup	*$P3_2$*	*$P3_2$*	*$P3_2$*	*$P3_2$*
Unit-cell (Å)	*a = b =50.1, c = 102.7, γ = 120°*	*a = b = 50.3, c =102.7, γ = 120°*	*a = b = 49.4, c =101.3, γ = 120°*	*a = b = 49.7, c = 102.9, γ = 120°*
Max resolution (Å)	*1.45*	*1.75*	*1.5*	*1.7*
No. measured reflections	*263341*	*153966*	*233224*	*166415*
No. unique reflections	*48158*	*27864*	*41718*	*29655*
Completeness (%)	*94.4 (86.1)*	*95.3 (90.4)*	*93.5 (87.8)*	*94.9 (86.6)*
R_{merge}† (%)	*4.9 (42.6)*	*4.9 (36.4)*	*6.7 (38.5)*	*4.3 (41.6)*
R_{meas} (%)	*4.3 (47.1)*	*5.4 (40.1)*	*4.5 (51.5)*	*4.7 (45.9)*
Mean I/σ(I)	*22.7 (3.7)*	*21.7 (4.4)*	*25.9 (4.0)*	*20.8 (4.2)*
Refinement				
Resolution range (Å)	*50.0 - 1.45*	*40.1 - 1.75*	*42.8-1.5*	*39.7 - 1.7*
No observations used	*47703*	*27548*	*41718*	*29394*
R_{factor} (%)	*20.0*	*17.9*	*19.8*	*19.5*
R_{free} (%)	*23.3*	*22.2*	*23.5*	*22.8*
Average B-factors				
Protein	20.8	24.1	19.7	29.2
Ligand	29.1	39.8	32.7	59.6
Water	28.4	29.5	31.5	36.4
Ramachandran distribution				
Most favoured (%)	98.9	98.8	98.8	98.8
Allowed (%)	1.1	1.2	1.2	1.2

†Values in parentheses are for the highest resolution shell.

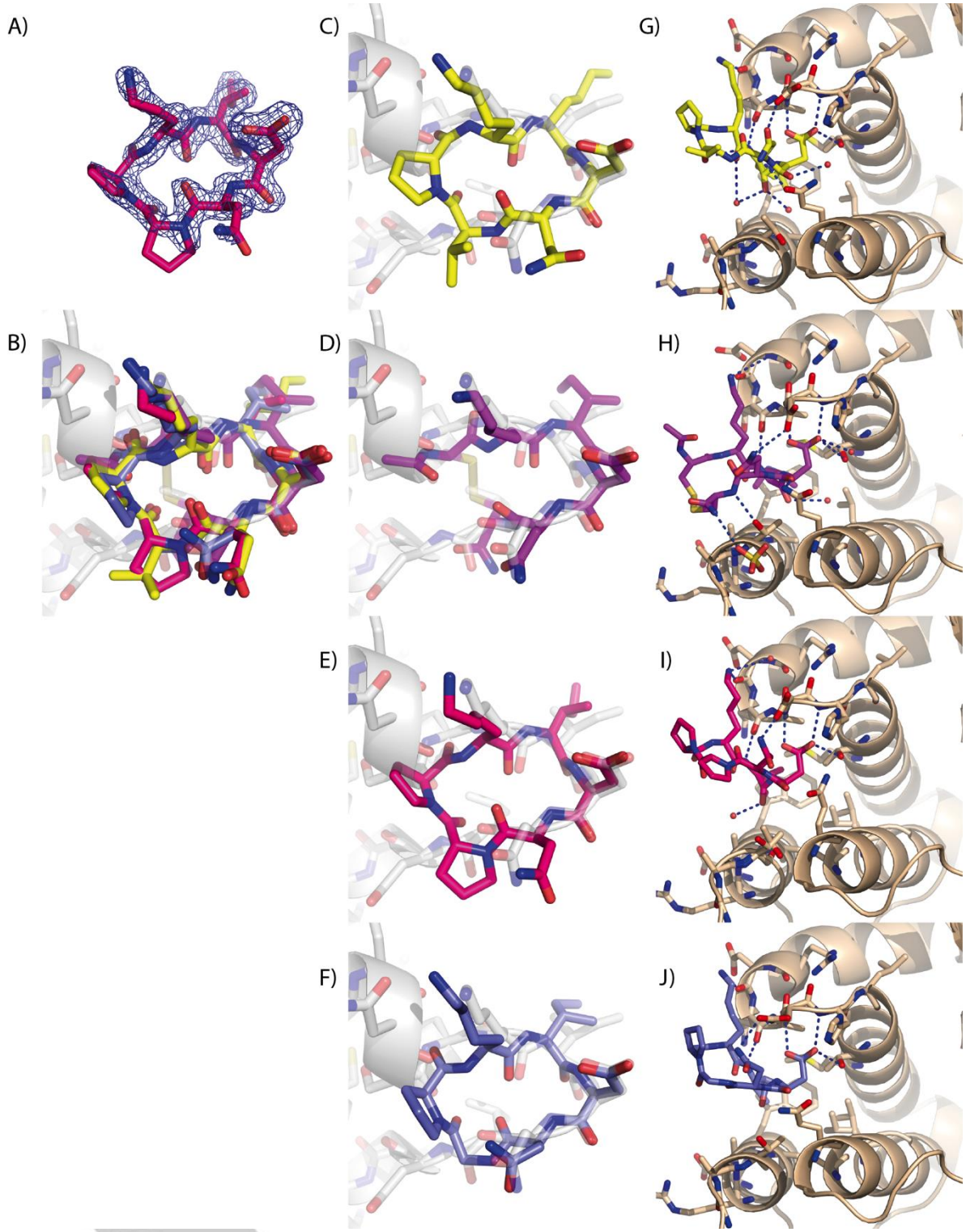


Figure 3. A) 2*Fo-Fc* electron density contoured at 1σ surrounding the cyclo(PKIDNp) peptide. B) Superimposition of peptides ox(Ac-CKIDNC-NH_2) (purple), cyclo(PKIDNp) (pink), cyclo(PKZDNv) (yellow) and cyclo(PKIDNG (blue), over the LEDGF IBD/integrase crystal structure 2B4J. C-F) Cyclic peptides cyclo(PKZDNv) (yellow), ox(Ac-CKIDNC-NH_2) (purple), cyclo(PKIDNp) (pink) and cyclo(PKIDNG) (blue) superimposed over the LEDGF IBD crystal structure 2B4J. G-J) Hydrogen bonding of the cyclic peptides with residues in the LEDGF binding pocket of integrase. Peptides and protein residues within 5 Å are shown as stick with the protein shown in cartoon. Crystallographic waters are shown in sphere and hydrogen bonds as dashed lines.

Changing the linker residue from Gly to D-Pro or D-Val did not have a great affect on the overall conformation of the cyclic peptide but did cause a subtle conformational change of the asparagine residue, which moved further away from integrase E170 (**2**) to protrude toward and form hydrogen bonds with integrase Q95 (**3**). In contrast, the disulfide linked cyclic peptide ox(Ac-CKIDNC-NH_2) (**1**) has a unique buckled shape that causes more distinct changes in the binding of the peptide to integrase. The positions of the peptide Asp and Asn side-chains were similar to those in the other complexes, however, the positions of the Lys and Ile residues were shifted, due to the different constraints introduced into the peptide ring by the disulfide link. This occurred because, unlike the dipeptide scaffolds in compounds **2**–**4**, the oxidised cystine residues that form the disulfide bond scaffold in **1** make hydrogen bonding interactions with integrase, specifically between the nitrogen of the C-terminal cysteine and T125 of integrase. This interaction causes the peptide to bend back on itself, and therefore pull the peptide Ile and Lys further from the protein than their head-to-tail cyclised counterparts. Despite this, the peptide Ile and Lys formed interactions with the integrase dimer which serves to demonstrate the importance of selecting an appropriate scaffold for these peptides in order to study peptide conformation, designing them such that they do not contribute to protein binding.

Each complex was also compared to the structure of the integrase/IBD complex (PDB ID 2B4J) (Figure 3B). In each case the KIDN motif in the cyclic peptides was successfully manipulated by the inclusion of our scaffold, and was presented in a manner that closely matched the equivalent residues in the parent structure (Figures 3C–J) where the Asp formed hydrogen bonds with the backbone nitrogen atoms of E170 and H171, and the T174 side-chain hydroxyl. The scaffolds themselves do not make binding contact with integrase. The compact size of the cyclic peptides in this study would prevent similar interactions with residues in the integrase N-terminal domain like those observed in the HIV-2 IN_{CORE+N}/LEDGF IBD crystal structure (PDB ID 3F9K).[14] Comparison with the longer cyclic peptides reported by Rhodes et al[17c], also shows a close structural match of the KIDN binding motif region, with the additional residues of the longer peptides extending into solution.

NMR and SPR: Peptide binding affinity

Analysis of peptide binding affinity to the core domain of integrase involved a combination of NMR and surface plasmon resonance (SPR) experiments (Table 2). Heteronuclear single quantum coherence (HSQC) spectroscopy was used to determine peptide affinity for integrase. We used 0.23 mM (4.4 mg/ml) ^{15}N-labelled integrase core domain with the mutations Q53E, C56S, W131E, F185K, Q209E (IN_{BAX})[22] and titrated in increasing concentrations of peptide (final concentrations 0.03–3.0 mM). Chemical shift perturbations were observed upon addition of increasing peptide for multiple residues at the LEDGF binding site (Figure 4). Peptides **1**–**3** were found to bind with millimolar affinity to the protein, while **4** bound too weakly to calculate more accurately than >10 mM.

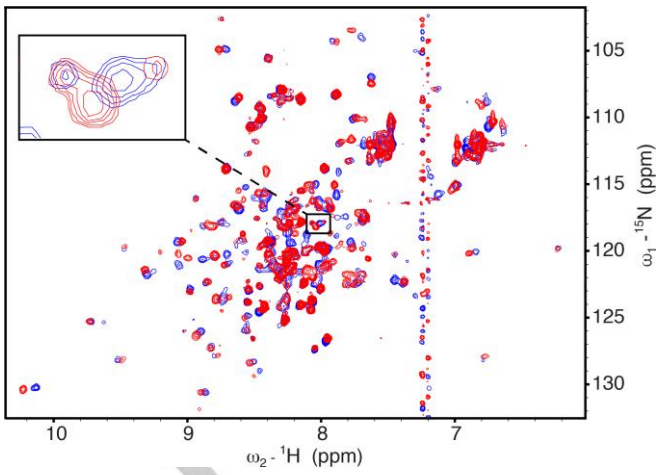


Figure 4: HSQC NMR spectra of ^{15}N-IN_{BAX} in the apo state (blue) and in the presence of 2 mM of cyclopeptide 2. Chemical shift perturbations of residues at the LEDGF binding site of integrase are observed upon peptide binding.

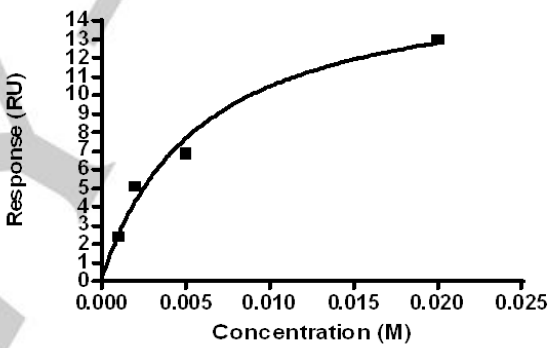


Figure 5: SPR binding curve for compound 3 binding to IN_{BAX}. A K_D of 5.8 mM was calculated from this data.

Table 2. Affinities of peptides for integrase measured by HSQC titration and SPR experiments.

Peptide	Sequence	K_D (mM) HSQC NMR	K_D (mM) SPR
1	ox(Ac-CKIDNC-NH_2)	5.8	ND[a]
2	cyclo(PKIDNp)	4.3	3.9
3	cyclo(PKZDNv)	3.1	5.8
4	cyclo(PKIDNG)	>10	>10

One letter code Z denotes norleucine. Lower case denotes D-amino acids. [a]The K_D could not be determined due to the presence of DTT in the SPR buffer which reduced the peptide disulfide bond. DTT was removed from NMR buffer to obtain a K_D (for that experiment only).

SPR was used as a corroborating measure of peptide affinity for the core domain of integrase. The IN_{BAX} construct was immobilised onto a BIACapture sensor chip, and the response to increasing concentrations of peptide solutions was measured to calculate the peptide K_D (Figure 5). Due to the presence of DTT in the SPR running buffer, attempts to analyse **1** by SPR would result in the reduced linear precursor, thus only compounds **2**–**4** were analysed by SPR.

Comparison of peptide bound and solution conformations

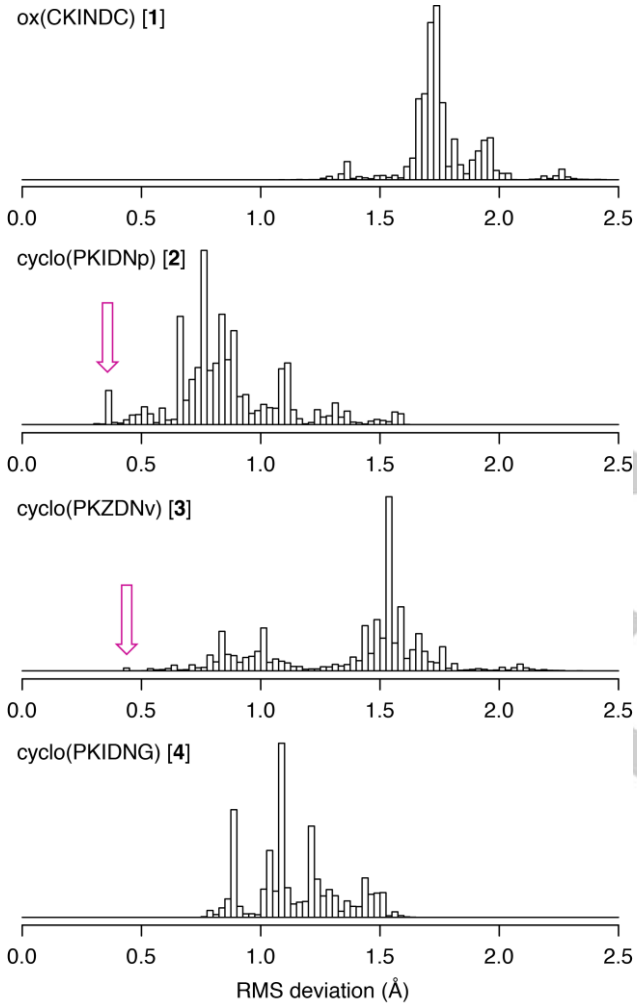


Figure 6: Comparison between modelled solution conformations of peptides **1**–**4** and the observed crystal structure conformations. Peptide conformers were generated using a conformational search using the GBSA continuum solvation model. The RMS deviation is calculated between the peptide backbone and side chain carbon atoms. Pink arrows denote that accessible solution conformations of peptides **2** and **3** most closely resemble the crystal structure conformations.

Given that the experimental affinities of the cyclic peptides for integrase were in the low mM range, we investigated whether this low affinity might be caused by a mismatch between the low energy conformations of the peptide in solution and the crystal structure-bound conformation. A Monte Carlo Multiple Minimum/low mode conformational search method[23] was used with the GBSA continuum water solvation model[24] to generate ~40,000 low energy conformations of each peptide. Each ensemble of low energy peptide conformations was superimposed on the corresponding crystal structure conformation and the RMS deviation of the backbone and side chain carbon atoms was calculated. Figure 6 shows the frequencies of RMS deviations from crystal structure for each peptide. In no case does a GBSA solution conformation exactly match the crystal structure conformation. The solution conformations of the higher affinity peptides **2** and **3** match the crystal structure more closely (pink arrows) than the lower affinity peptides **1** and **4**. This analysis suggests that the favoured solution conformations do indeed differ from the crystal structure conformations and that there is a significant free energy penalty associated with peptide binding.

Free energy of ligand binding: umbrella sampling and metadynamics

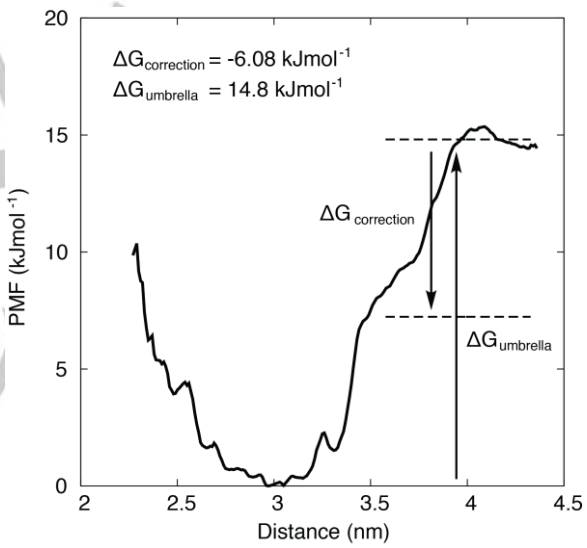


Figure 7: Potential of mean force curve (PMF) for the removal of cyclo(PKIDNG) (**4**) from the LEDGF binding site of integrase. The distance is defined as the Z-component of the distance between the ligand and the centre of monomer A.

To further investigate the poor binding of compound **4** (measured kD > 10 mM or ΔG > -11.5 $kJmol^{-1}$, Table 2) we used molecular dynamics simulations with an explicit water solvation model to calculate the free energy of binding. In the first instance, we used umbrella sampling simulations to extract peptide **4** progressively from the binding site. The ligand was constrained relative to the centre of mass of the relevant integrase monomer. Each umbrella window was sampled for 20 ns, with the first 10 ns was discarded as an equilibration phase. The potential of mean force (PMF) was calculated using the weighted histogram analysis method (WHAM)[25] and is shown in Figure 7. The constraining potential required to remove peptide **4** from the

LEDGF is 14.8 kJmol^{-1} (denoted $\Delta G_{umbrella}$), implying a K_D value of 2.49 mM at 300K.

Although the predicted K_D appears to be in reasonable agreement with the experimental lower bound of 10 mM, umbrella sampling simulations are likely to overestimate ΔG due to insufficient sampling of peptide conformations at each point along the pull trajectory. Thus the entropic cost of ligand binding is likely not adequately accounted for. To demonstrate poor sampling of peptide conformations, we performed a 200 ns simulation of peptide **4** in solution**,** commencing from the crystal structure pose. After this time, the RMSD of the Cα carbons of the free ligand differed by only 1.18 Å, showing (as expected) that sampling of peptide conformation is slow.

Seeking to estimate the difference in free energy[26] between the crystal structure pose and the lowest-energy solution conformations, we performed a 100 ns, well-tempered metadynamics[27] simulation using three internal, distances as the collective variables (CVs) (Supplementary Figure 1). The conformations from the simulation were clustered into 6 groups (Figure 8) and the free energies of each cluster were determined using Meta-GUI.[28] Cluster 2 has the lowest free energy, Figures 9A and B are representative structures accessible by unbiased simulations of the bound and free peptide. A depiction of each cluster which includes all elements of each cluster is can be found as Supplementary Figure 2. As shown in table 3, the crystal pose for peptide **4** differs from a representative structure from the lowest energy cluster with a backbone RMSD of 1.69 Å. Associated Gibbs free energies relative to the low energy conformations are also included to illustrate that there is a difference between the energy of the ligand conformation produced via steered MD, and that of the low energy conformation for that ligand. In fact, the lowest energy cluster of structures is 6.77 kJmol^{-1} lower than Cluster 4, which includes the conformer of the bound pose in solution within it (Table 3).

Table 3. Free energies for each cluster representative and the RMSD difference (Å) for each representative structure from the representative bound pose. Free energies were generated from the coordinates of the metadynamics derived 3D free energy surface, given the values relevant to each representative conformer, as illustrated above. Each free energy value is associated with the specified conformer in solution.

Conformer	Free Energy Relative to Lowest Energy Cluster	Backbone RMSD from Bound Pose (Å)
Cluster 1	0.25	1.24
Cluster 2	0.00	1.63
Cluster 3	11.56	1.17
Cluster 4	6.77	1.04
Cluster 5	11.36	1.59
Cluster 6	6.09	1.69

The Boltzmann weighted average free energy of each cluster was 0.69 kJmol^{-1}. However, when peptide **4** was first pulled into solution from the binding site, there was an energy change of 14.8 kJmol^{-1}.Further, from the exploration of conformational space with metadynamics we know that under equilibrium conditions, peptide **4** should have given a free energy value that was 6.08 kJmol^{-1} more positive (the absolute difference between 0.69 kJmol^{-1} and 6.77 kJmol^{-1}). Therefore, by addition of this 6.08 kJmol^{-1} corrective term to the binding free energy predicted by steered MD with umbrella sampling (ΔG = -14.80 kJmol^{-1}), a corrected binding free energy, ΔG = -8.72 kJmol^{-1} is obtained, or K_D = 30.1 mM. This is in better agreement with the experimental dissociation constant, K_D > 10 mM than without the metadynamics corrective term.

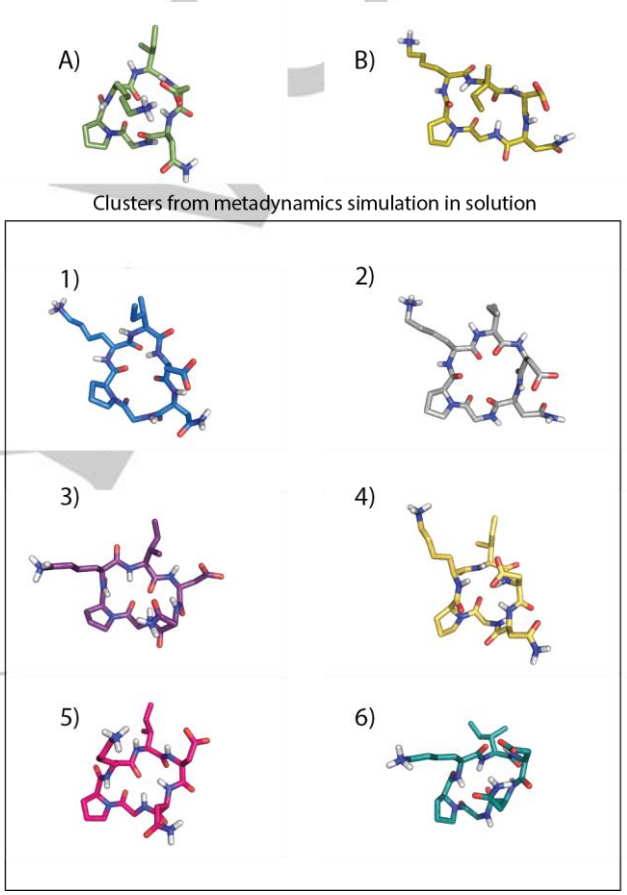


Figure 8: A) Representative structure taken from an unbiased 500 ns simulation of the bound ligand. B) Representative structure of the free ligand taken from an unbiased simulation in solution (200ns). (1-6) Representative members of clusters derived from metadynamics conformational search of the free ligand in solution.

In summary, it was found that peptides **1**–**4**, despite closely mimicking the KIDN binding sequence of LEDGF, bound poorly to the LEDGF site of HIV-IN. Using steered MD with umbrella sampling, the experimental binding affinity was predicted to be K_D = 2.49 mM when the experimental value for K_D was greater than 10 mM in this instance. By sampling the full free energy surface of the cyclic peptide with metadynamics using three trans-peptide distances as the collective variable, it was found that the lowest energy microstate of the peptide in solution was

actually 6.77 kJmol^{-1} lower than the value originally found by steered MD. Adding a corrective term of 6.08 kJmol^{-1}, which was found by a mixture of metadynamics and cluster analysis, the predicted K_D was corrected from 2.49 mM to K_D = 30.1 mM, which is in more literal agreement with the experimental value of K_D > 10mM.

The hypothesis is therefore that the lowest energy bound pose in solution is energetically dissimilar to the low-energy pose in complex with HIV-IN, thereby introducing an energetic barrier the ligand must cross before it can bind, and thus explaining the lower-than-expected experimental binding free energy for the cyclic hexapeptide mimetics.

Conclusions

In this work we report a set of four cyclic hexapeptides that were designed to mimic a key protein binding loop in the interaction of HIV-1 integrase with LEDGF. Three of the peptides contain dipeptide scaffolds (DPro-Pro, Gly-Pro and DVal-Pro) that are designed to restrict the conformational flexibility of the peptides. Crystal structures of the peptides bound to integrase confirm that the peptides adopt the designed conformations and closely replicate the conformation of the KIDN loop observed in the LEDGF/integrase complex. NMR and SPR binding studies reveal that, despite the close mimicry of the binding loop, the affinity of the more potent peptides is in the range of 3 – 4 mM. Computational analysis of the peptide conformations in solution using a Monte-Carlo simulation of the peptide in an MMGBSA solvent model suggests that while the solution conformations are compatible with the observed protein-bound peptide conformations, the lowest energy solution conformations deviate from the bound pose as per the crystal structure (RMSD 0.7 – 1.5 Å). Further, using a combination of MD, steered MD, metadynamics and cluster analysis of peptide **4** in explicit water, it was shown that the lowest energy conformation of the peptide (within Cluster 6 – see Table 3) differs from the aqueous bound pose in Cα positioning and free energy by approximately 2.15 Å RMSD and 14.8 kJmol^{-1}. The fact that the lowest energy free poses for each peptide are energetically and structurally different from the bound pose in solution most likely contributes to the low compound affinity by introduction of a large entropic penalty to binding. Finally, it is interesting to note that although the KIDN loop has been identified to be a hotspot in the binding of LEDGF to integrase, close mimicry of the crystal bound pose is not sufficient to generate high-affinity peptide ligands for the LEDGF site and that the additional interactions between LEDGF and integrase are required to produce viable peptide inhibitors of this important protein-protein interaction in the HIV-1 lifecycle.

Experimental Section

Peptide synthesis

Linear peptides: Peptide **1** was synthesised on Rink Amide resin (0.65 mmol/g, 100-200 mesh) and peptides **2**–**4** were synthesised on 2-chlorotrityl resin. Standard Fmoc protocols were used on a PS3 (Protein Technologies Inc.) automated solid-phase peptide synthesiser. Fmoc deprotections were performed using 14% piperidine in DMF (2 x 5 ml, 5 min). Couplings were performed using 3 molar equivalents of Fmoc-amino acids and HCTU, activating with 7% DIPEA in DMF. Coupling reactions were carried out for 50 minutes. Between successive steps, the peptide-resin was washed with DMF (6 x 5 ml).

*Ox(Ac-CKIDNC-NH$_2$) (**1**)*: To acetylate the amino terminus, the peptide-resin was treated with 100 µL acetic anhydride and 100 µL DIPEA in 5 ml DMF for 10 min. The peptide-resin was treated with cleavage solution (94% TFA, 2.5% H_2O, 2.5% EDT, 1% TIS) at room temperature for 90 min, stirring gently. The mixture was filtered and the filtrate was concentrated under a stream of N_2 gas. Diethyl ether (30 ml) was added and the resultant precipitate was collected by filtration. The crude peptide was purified by RP-HPLC (detection at 214 nm) and recovered as a white powder (80% yield). LCMS (*m/z*): 736.5 $[M + H]^+$. LCMS t_R: 10.42 min. The linear peptide was oxidised in 25 mM ammonium bicarbonate (1 mg ml^{-1}) at room temperature for 24 hours. The cyclic product was purified by RP-HPLC and recovered as a white powder (54% yield). LCMS (*m/z*): 734.5 $[M + H]^+$. LCMS t_R: 10.08 min. HRMS (*m/z*): $C_{28}H_{48}N_9O_{10}S_2^+$ requires $[M + H]^+$ 734.2960; found 734.2955.

*Cyclo(PKIDNp), cyclo(PKZDNv) and cyclo(PKIDNG) (**2**–**4**)*: The side-chain protected linear peptide was cleaved from the resin in 1% TFA in DCM (2 ml) for 5 x 2 min. The solution was filtered into a flask containing 10% pyridine in methanol (10 ml). The resin was washed with 3 ml DCM, 3 ml methanol, 2 ml DCM, and 2 ml methanol. The DCM was removed *in vacuo* and the peptide was lyophilised. Linear N(Trt)-DP-P-K(Boc)-I-D(OtBu): 95% yield. LCMS (*m/z*): 1081.75 $[M + H]^+$. LCMS t_R: 16.59 min. Linear N(Trt)-G-P-K(Boc)-I-D(OtBu): 93% yield. LCMS (*m/z*): 1041.75 $[M + H]^+$. LCMS t_R: 15.94 min. Linear N(Trt)- DV-P-K(Boc)-nL-D(OtBu): 97% yield. LCMS (*m/z*): 1083.8 $[M + H]^+$. LCMS t_R: 17.1 min.

To the protected linear peptide in DMF (4 mM final concentration of peptide) at 4˚C was added 3 molar equivalents of diphenylphosphoryl azide (DPPA) and 4 molar equivalents of DIPEA. The mixture was reacted at 4˚C for 2-5 days and reaction progression was monitored by LCMS. When complete, DMF was removed *in vacuo*. Side-chain protecting groups were subsequently removed in 95% TFA, 5% TIPS for 90 min. The mixture was concentrated under a stream of N_2 gas and 30 ml diethyl ether was added to precipitate the peptide. The cyclic peptide was filtered and purified by RP-HPLC. cyclo(pPKIDN): 20% yield. LCMS (*m/z*): 665.55 $[M + H]^+$. LCMS t_R: 10.05 min. HRMS (*m/z*): $C_{30}H_{49}N_8O_9^+$ requires $[M + H]^+$ 665.3617; found 665.3639. cyclo(GPKIDN): 10% yield. LCMS (*m/z*): 625.45 $[M + H]^+$. LCMS t_R: 9.55 min. HRMS (*m/z*): $C_{27}H_{45}N_8O_9^+$ requires $[M + H]^+$ 625.3308; found 625.3308. cyclo(vPK-Nle-DN): 66% yield. LCMS (*m/z*): 667.5 $[M + H]^+$. LCMS t_R: 10.57 min. HRMS (*m/z*): $C_{30}H_{51}N_8O_9^+$ requires $[M + H]^+$ 667.3774; found 667.3763.

Protein expression and purification

Mutagenesis: Two constructs of the core domain of integrase were expressed for analysis of peptide binding. For crystallography, a truncated HIV-1 IN_{50-212} construct similar the NL43 strain containing an N-terminal 6-histidine (His_6) tag followed by a thrombin cleavage site (Stratagene) and solubilising mutations C56S, W131D, F139D and F185H was used ($IN_{CORE4H123}$). This construct has been previously reported[10] and includes changes at residues 123-127 from STTVK to GATVR. For NMR studies, ^{15}N-labelled His_6-tagged core domain IN_{50-212} was expressed with the solubilising mutations Q53E, C56S, W131, F185K, and Q209E. This construct[22] is referred to as IN_{BAX}.

Protein expression and purification for crystallography: $IN_{CORE4H123}$ was expressed and purified from *Eschericia coli* (*E. coli*) using methods previously reported by Wielens et al.[10]

Protein expression and purification for NMR: IN_{BAX} was expressed and purified from *E. coli* as described by Fitzkee et al.[22] The protein was purified by Ni-NTA affinity chromatography using a HisTrap™ column (GE Healthcare) and eluted using an imidazole gradient. The eluted His_6-tag-IN_{BAX} was cleaved by thrombin and the protein subsequently purified by Ni-NTA, followed immediately by size-exclusion chromatography using a HiLoad Superdex 75 column, pre-equilibrated with the NMR buffer (20 mM HEPES, pH 6.8, 150 mM NaCl, 40 mM $MgCl_2$). Prior to NMR analysis, the protein was concentrated to 4 mg/ml using Vivaspin centrifugal concentrators (Sartorius) with a 10,000 MWCO polyether sulfone membrane at 4000 g. 10% D_2O was added for locking purposes.

X-Ray crystallography

Briefly, crystals of $IN_{CORE4H123}$ were grown by mixing 2 µl of 8 mg ml^{-1} $IN_{CORE4H123}$ in 25 mM HEPES pH 6.5, 0.5 M NaCl and 5 mM DTT with an equal volume of reservoir solution (1.8 M ammonium sulfate, 0.15 M sodium citrate pH 4.6 and 5 mM cadmium chloride) using the vapour diffusion method at 293 K. $IN_{CORE4H123}$ crystals were then soaked overnight in a solution containing reservoir buffer and 5 mM cyclic peptide. Crystals were transferred to a solution comprising 5 mM cyclic peptide, 25 % (*w/v*) sucrose and reservoir buffer for 1-2 min before crystals were frozen by plunging into liquid nitrogen. Diffraction data for the $IN_{CORE4H123}$/peptide crystals were collected at 100 K on the MX2 beamline at the Australian Synchrotron (Melbourne, Australia) using the Blu-Ice interface[29]. Data were indexed and scaled using XDS[30]. Initial phases were determined using AMoRe[31] with the unliganded HIV-1 core domain (PDB ID: 3L3U[9, 32]) as the search model. Model building and refinement were completed using Coot[33], Refmac5[34] and Phenix[35]. The final statistics of refinement are summarised in *Table 1*. The models were analysed with the program PROCHECK[36] and MOLPROBITY[37], which showed that the stereochemical quality was similar or better than expected for structures refined at similar resolution. The coordinates of the four $IN_{CORE4H123}$/cyclic LEDGF peptide complexes have been deposited in the PDB[38] with the accession numbers 3WNF, 3WNG, 3WNH and 3WNE.

HSQC NMR studies

Single point experiments: Peptide binding was evaluated by $^{15}N,^{1}H$-HSQC experiments. Peptides were dissolved in H_2O to give 32 mM stock solutions. 10 µL of stock solution was used for each peptide analysis. Uniformly ^{15}N-labeled IN_{BAX} (0.2 mM; 4 mg/ml) was incubated with 2 mM of peptide for 30 min. $^{15}N,^{1}H$-HSQC spectra were recorded on a Varian INOVA 600 MHz spectrometer fitted with a cryogenically cooled probe at 298 K. Typical acquisition parameters were sweep widths of 14 ppm (^{1}H) x 32 ppm (^{15}N) with 2048 x 128 increments respectively, and 16 scans per ^{15}N increment. The recycle delay was 1.2 sec. Spectra were processed and analysed using Topspin 3.0 (Bruker Biospin).

Titration experiments: Stock solutions of 250 mM peptide were made for titration experiments by dissolving the peptide powder in H_2O. Serial dilutions were performed to give five peptide concentrations. Increasing concentrations of peptide (up to 3 mM) were titrated into ^{15}N-labeled IN_{BAX} (0.23mM; 4.4 mg/ml). The standardised chemical shift perturbations (σ^s) of individual amide resonances in $^{15}N,^{1}H$-HSQC spectra observed upon addition of increasing concentrations of peptide were obtained using the formula below which is derived from Pythagoras' theorem with the ^{15}N perturbation standardised to ^{1}H.

$$\sigma^s = \sqrt{(\sigma^1\mathrm{Hholo} - \sigma^1\mathrm{Hapo})^2 + 0.04\,(\sigma^{15}\mathrm{Nholo} - \sigma^{15}\mathrm{Napo})^2}$$

To derive the equilibrium binding constant (K_D) the standardised chemical shifts were fitted to a single-state model with ligand depletion using GraphPad Prism 5.0 (GraphPad Software, San Diego California USA, www.graphpad.com) with the K_D constrained to a shared value for all resonances.

Surface plasmon resonance

Peptide binding to IN_{BAX} was measured on a Biacore T200 instrument and data was analysed using GraphPad Prism (GraphPad Software, San Diego California USA, www.graphpad.com).

Minimal biotinylation of integrase proteins: IN_{BAX} was biotinylated by incubating with equimolar amounts of NHS-LC-LC-biotin on ice for 2 h. Removal of excess biotin and buffer exchange into integrase Biacore buffer (50 mM HEPES, 150 mM NaCl, 5 mM DTT, 10 mM $MgCl_2$, 0.05% P-20 surfactant, 5% DMSO, pH 7.4) was achieved by passage through a pD-10 gel filtration column.

Immobilisation of biotinylated integrase proteins: Biotinylated IN_{BAX} was immobilised on the surface of a BiaCAPture sensor chip (Biacore) according to manufacturers' instructions. Following 3 injections of 8 M guanidine hydrochloride/1 M NaOH, the chip surface was activated with a 300s injection of BiaCAPture reagent, followed by a 300s injection of IN_{BAX}. Final immobilisation level was typically 500-600RU. A reference flow cell was formed by injection of BIACapture reagent followed by an injection of 10 mM biotin.

Binding of peptides to immobilised integrase proteins: Binding assays were carried out at 25 °C using a flow-rate of 40 uL/min in integrase Biacore buffer. Assays were carried out in a manual run format, with an association phase of 60 s followed by a dissociation phase of 30 s. Following each binding, the surface was regenerated and re-activated as described above. K_D values were determined from the raw data using GraphPad Prism, with data fitted to a non-linear, one site binding model.

Computational analysis

Conformational search: MacroModel Batchmin version 10.3 (Schrödinger) was used to perform a 100000 step mixed Monte Carlo Multiple Minimum (MCMM)/Low-mode search using the OPLS 2005 force field and the MM-GBSA water solvation model. Each search produced approximately 40000 unique conformations. The low-energy conformers of each peptide were superimposed onto the corresponding crystal structure-bound conformation backbone N, CA and C atoms and side chain carbon atoms. The RMS deviation between conformer and the crystal structure was determined and Boltzmann weighted histograms were plotted using R (version 3.4).

Molecular Dynamics Simulations: All simulations were performed using GROMACS 5.1.2, using a 2 fs time-step and a cubic periodic cell at 300 K. The GROMOS 54a7 force-field was used with the SPC/E water model.[39] NPT conditions were maintained with the Parrinello-Rahman barostat[40] (1 bar) and a modified Berendsen thermostat (V-Rescale, 300 K).[41] A HIV integrase core domain dimer was used to model the protein, with peptide **4** bound to the LEDGF site of only one of the monomeric units. The binding mode for the protein-ligand complex was taken from PDB structure 3WNE. The HIV-IN dimer was constructed using the crystal structure 4CJU from the protein data bank (PDB), and missing loops were filled with Prime-X.

Steered MD and Umbrella Sampling: Using the previously defined MD protocol, the GROMACS pull code was used to harmonically constrain all ligands mentioned herein at varying distances the COM of the arbitrarily defined binding site. Following the pulling of the ligand to each point and harmonically constraining it in place with high force constants, the harmonic force-constants (K) constraining the ligand in place were lowered to the following values for umbrella windows 1 to 16, each window a small increment further away from the binding site with the first umbrella window sampling the initial distance: K = [50, 4000, 4000, 4000, 4000, 2000, 2000, 2000, 2000, 750, 750, 750, 750, 750, 750, 750] kJmol^{-1}nm^{-1}. The WHAM analysis was performed using the GROMACS package, *gmx wham*[42]. Sixteen umbrella windows used for umbrella sampling at uniformly incremented distances, away from the binding site and each umbrella window was run for 20 ns. Each increment was 1 nm further from the binding site, and the ligand was constrained at these increments via the Z axis alone. In the umbrella sampling simulations, the Cα atoms of the monomer without a bound ligand were position restrained in place using flat-bottomed restraints, using a force-constant of 1000 kJmol^{-1} nm^{-1}. The ligand was pulled away from the integrase monomer that did not have any biasing force acting to restrain it in place, and the rhombic dodecahedron-shaped periodic box was constructed such that the ligand could not be pulled across more than half the box length, and that no relevant biomolecule was less than 20 Å from its periodic neighbour at any point.

Metadynamics: Well-tempered metadynamics simulations were performed using Plumed version 2.2.[43] Three trans-peptide distances were used as collective variables (CVs) (Supplementary Figure 1) to force the cyclic peptide to adopt different conformations in a periodic box of solvent. The Gaussian height of the metadynamics biasing potential was set to 0.5 kJmol^{-1}, while the σ-values for the Gaussians acting to bias the three trans-peptide distances were 0.05 nm, 0.025 nm and 0.01 nm respectively. The bias factor for this well-tempered simulation was 15 and Gaussians were added to the free energy surface every 100 steps. Reconstruction of the free energy surface depicting the Gibbs Free Energy of peptide **4** as a function of the three distances used as CVs was also performed using Plumed. Clustering was performed based on the three CVs using Meta-GUI which additionally gave average free energies for each cluster. Representative structures for clusters or trajectories were defined as the lowest RMSD structure from the average (median) value for the CVs. A depiction of each cluster which shows every element of each cluster is provided in Supplementary Figure 2.

Acknowledgements

The integrase core domain construct used for NMR was a kind gift of Ad Bax (NIH, Maryland). We also thank the MX beamline staff for their help at the Australian synchrotron. This work was supported by the Australian Research Council Linkage project (LP0775192), the Australian Synchrotron Research Program and the Australian Cancer Research Foundation to M.W.P. M.W.P. is an ARC Federation Fellow and National Health Medical Research Council Honorary Fellow. We acknowledge computational resources and technical support provided by the National Computational Infrastructure (NCI), and Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE) through grant y96. We also acknowledge the provision of an Australian Post-graduate Award to SN.

Keywords: HIV-1 integrase • LEDGF • peptide mimetics • crystallography • NMR

References:

[1] M. K. Lewinski, F. D. Bushman, *Adv Genet* **2005**, *55*, 147-181.

[2] aJ. Luban, *Current Biol* **2008**; bL. Krishnan, A. Engelman, *J Biol Chem* **2012**, *287*, 40858-40866.

[3] K. A. Jurado, H. Wang, A. Slaughter, L. Feng, J. J. Kessl, Y. Koh, W. Wang, A. Ballandras-Colas, P. A. Patel, J. R. Fuchs, M. Kvaratskhelia, A. Engelman, *Proc Natl Acad Sci U S A* **2013**, *110*, 8690-8695.

[4] V. Summa, A. Petrocchi, F. Bonelli, B. Crescenzi, M. Donghi, M. Ferrara, F. Fiore, C. Gardelli, O. Gonzalez Paz, D. J. Hazuda, P. Jones, O. Kinzel, R. Laufer, E. Monteagudo, E. Muraglia, E. Nizi, F. Orvieto, P. Pace, G. Pescatore, R. Scarpelli, K. Stillmock, M. V. Witmer, M. Rowley, *J Med Chem* **2008**, *51*, 5843-5855.

[5] M. Sato, T. Motomura, H. Aramaki, T. Matsuda, M. Yamashita, Y. Ito, H. Kawakami, Y. Matsuzaki, W. Watanabe, K. Yamataka, S. Ikeda, E. Kodama, M. Matsuoka, H. Shinkai, *J Med Chem* **2006**, *49*, 1506-1508.

[6] S. Min, I. Song, J. Borland, S. Chen, Y. Lou, T. Fujiwara, S. C. Piscitelli, *Antimicrob Agents Chemother* **2010**, *54*, 254-258.

[7] A. Markham, *Drugs* **2018**, *78*, 601-606.

[8] P. Cherepanov, G. N. Maertens, S. Hare, *Curr Opin Struct Biol* **2011**, *21*, 249-256.

[9] J. Wielens, S. J. Headey, D. Jeevarajah, D. I. Rhodes, J. Deadman, D. K. Chalmers, M. J. Scanlon, M. W. Parker, *FEBS Lett* **2010**, *584*, 1455-1462.

[10] J. Wielens, S. J. Headey, J. J. Deadman, D. I. Rhodes, G. T. Le, M. W. Parker, D. K. Chalmers, M. J. Scanlon, *ChemMedChem* **2011**, *6*, 258-261.

[11] D. I. Rhodes, T. Peat, N. Vandegraaff, D. Jeevarajah, Le, Jones, J. Smith, Coates, Winfield, Thienthong, J. Newman, D. Lucent, J. Ryan, G. Savage, C. Francis, Deadman, *Antiviral Chem Chemother* **2011**, *21*, 155-168.

[12] aA. Engelman, P. Cherepanov, *PLoS Pathog* **2008**, *4*, e1000046; bE. M. Poeschla, *Cell Mol Life Sci* **2008**, *65*, 1403-1424.

[13] P. Cherepanov, A. L. Ambrosio, S. Rahman, T. Ellenberger, A. Engelman, *Proc Natl Acad Sci U S A* **2005**, *102*, 17308-17313.

[14] S. Hare, F. Di Nunzio, A. Labeja, J. Wang, A. Engelman, P. Cherepanov, *PLoS Pathog* **2009**, *5*, e1000515.

[15] aK. Gupta, T. Brady, B. M. Dyer, N. Malani, Y. Hwang, F. Male, R. T. Nolte, L. Wang, E. Velthuisen, J. Jeffrey, G. D. Van Duyne, F. D. Bushman, *J Biol Chem* **2014**, *289*, 20477-20488; bL. D. Fader, E. Malenfant, M. Parisien, R. Carson, F. Bilodeau, S. Landry, M. Pesant, C. Brochu, S. Morin, C. Chabot, T. Halmos, Y. Bousquet, M. D. Bailey, S. H. Kawai, R. Coulombe, S. LaPlante, A. Jakalian, P. K. Bhardwaj, D. Wernic, P. Schroeder, M. Amad, P. Edwards, M. Garneau, J. Duan, M. Cordingley, R. Bethell, S. W. Mason, M. Bos, P. Bonneau, M. A. Poupart, A. M. Faucher, B. Simoneau, C. Fenwick, C. Yoakim, Y. Tsantrizos, *ACS Med Chem Lett* **2014**, *5*, 422-427; cA. Sharma, A. Slaughter, N. Jena, L. Feng, J. J. Kessl, H. J. Fadel, N. Malani, F. Male, L. Wu, E. Poeschla, F. D. Bushman, J. R. Fuchs, M. Kvaratskhelia, *PLoS Pathog* **2014**, *10*, e1004171; dP. A. Patel, N. Kvaratskhelia, Y. Mansour, J. Antwi, L. Feng, P. Koneru, M. J. Kobe, N. Jena, G. Shi, M. S. Mohamed, C. Li, J. J. Kessl, J. R. Fuchs, *Bioorg Med Chem Lett* **2016**, *26*, 4748-4752; eK. Gupta, V. Turkki, S. Sherrill-Mix, Y. Hwang, G. Eilers, L. Taylor, C. McDanal, P. Wang, D. Temelkoff, R. T. Nolte, E. Velthuisen, J. Jeffrey, G. D. Van Duyne, F. D. Bushman, *PLoS Biol* **2016**, *14*, e1002584.

[16] aZ. Hayouka, J. Rosenbluh, A. Levin, S. Loya, M. Lebendiker, D. Veprintsev, M. Kotler, A. Hizi, A. Loyter, A. Friedler, *Proc Natl Acad Sci U S A* **2007**, *104*, 8316-8321; bZ. Hayouka, A. Levin, M. Maes, E. Hadas, D. E. Shalev, D. J. Volsky, A. Loyter, A. Friedler, *Biochem Biophys Res Commun* **2010**, *394*, 260-265; cM. Maes, A. Loyter, A. Friedler, *FEBS J* **2012**, *279*, 2795-2809.

[17] aM. Tsiang, G. S. Jones, M. Hung, S. Mukund, B. Han, X. Liu, K. Babaoglu, E. Lansdon, X. Chen, J. Todd, T. Cai, N. Pagratis, R. Sakowicz, R. Geleziunas, *J Biol Chem* **2009**, *284*, 33580-33599; bZ. Hayouka, M. Hurevich, A. Levin, H. Benyamini, A. Iosub, M. Maes, D. E. Shalev, A. Loyter, C. Gilon, A. Friedler, *Bioorg Med Chem* **2010**, *18*, 8388-8395; cD. I. Rhodes, T. S. Peat, N. Vandegraaff, D. Jeevarajah, J. Newman, J. Martyn, J. A. Coates, N. J. Ede, P. Rea, J. J. Deadman, *ChemBioChem* **2011**, *12*, 2311-2315; dB. A. Desimmie, M. Humbert, E. Lescrinier, J. Hendrix, S. Vets, R. Gijsbers, R. M. Ruprecht, U. Dietrich, Z. Debyser, F. Christ, *Mol Ther* **2012**, *20*, 2064-2075.

[18] D. Obrecht, E. Chevalier, K. Moehle, J. A. Robinson, *Drug Discov Today Technol* **2012**, *1*, e63-e69.

[19] M. A. Camerino, D. K. Chalmers, P. E. Thompson, *Chim Oggi* **2008**, *26*, 46-51.

[20] aJ. W. Bean, K. D. Kopple, C. E. Peishoff, *J Amer Chem Soc* **1992**, *114*, 5328-5334; bD. K. Chalmers, G. R. Marshall, *J Amer Chem Soc* **1995**, *117*, 5927-5937; cI. N. Rao, A. Boruah, S. K. Kumar, A. C. Kunwar, A. S. Devi, K. Vyas, K. Ravikumar, J. Iqbal, *J Org Chem* **2004**, *69*, 2181-2184; dE. F. Healy, J. Sanders, P. J. King, W. E. Robinson, *J Mol Graph Model* **2009**, *27*, 584-589.

[21] B. Chatterjee, I. Saha, S. Raghothama, S. Aravinda, R. Rai, N. Shamala, P. Balaram, *Chemistry* **2008**, *14*, 6192-6204.

[22] N. C. Fitzkee, J. E. Masse, Y. Shen, D. R. Davies, A. Bax, *J Biol Chem* **2010**, *285*, 18072-18084.

[23] I. Kolossavary, W. C. Guida, *J Comput Chem* **1999**, *20*, 1671-1684.

[24] I. Massova, P. A. Kollman, *Perspect Drug Discov* **2000**, *18*, 113-115.

[25] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J Comput Chem* **1992**, *13*, 1011-1021.

[26] B. J. Williams-Noonan, E. Yuriev, D. K. Chalmers, *J Med Chem* **2018**, *61*, 638-649.

[27] A. Barducci, G. Bussi, M. Parrinello, *Phys Rev Lett* **2008**, *100*.

[28] X. Biarnes, F. Pietrucci, F. Marinelli, A. Laio, *Comput Phys Commun* **2012**, *183*, 203-211.

[29] T. M. McPhillips, S. E. McPhillips, H. J. Chiu, A. E. Cohen, A. M. Deacon, P. J. Ellis, E. Garman, A. Gonzalez, N. K. Sauter, R. P. Phizackerley, S. M. Soltis, P. Kuhn, *J Synchrotron Radiat* **2002**, *9*, 401-406.

[30] W. Kabsch, *J Appl Crystallogr* **1993**, *26*, 795-800.

[31] J. Navaza, *Acta Crystallogr A* **1994**, *50*, 157-163.

[32] J. C. Chen, J. Krucinski, L. J. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt, R. M. Stroud, *Proc Natl Acad Sci U S A* **2000**, *97*, 8233-8238.

[33] P. Emsley, K. Cowtan, *Acta Crystallogr D* **2004**, *60*, 2126-2132.

[34] A. A. Vagin, R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, G. N. Murshudov, *Acta Crystallogr D* **2004**, *60*, 2184-2195.

[35] P. D. Adams, R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter, T. C. Terwilliger, *Acta Crystallogr D* **2002**, *58*, 1948-1954.

[36] R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, *J Appl Cryst* **1993**, *26*, 283-291.

[37] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res* **2007**, *35*, W375-W383.

[38] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res* **2000**, *28*, 235–242.

[39] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *The Journal of Physical Chemistry* **1987**, *91*, 6269-6271.

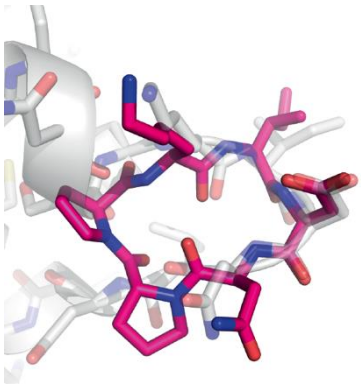
[40] M. Parrinello, A. Rahman, *Physical Review Letters* **1980**, *45*, 1196-1199.

[41] G. Bussi, D. Donadio, M. Parrinello, *The Journal of Chemical Physics* **2007**, *126*, 014101.

[42] J. S. Hub, B. L. de Groot, D. van der Spoel, *J Chem Theor Comput* **2010**, *6*, 3713-3720.

[43] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput Phys Commun* **2014**, *185*, 604-613.

Entry for the Table of Contents



Cyclic peptides are shown by X-ray crystallography to closely mimic the binding of the LEDGF loop to HIV integrase. However, these compounds have affinities in the mM range. Computational modelling suggests that although the peptides can conform to their protein binding site, their conformation in solution differs substantially from the bound conformation, imparting an entropic cost to binding.