Targeting uracil-DNA glycosylases for therapeutic outcomes using insights from virus evolution.

Renos Savva

Institute of Structural and Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK. +44 (0)20 7631 6805 r.savva@mail.cryst.bbk.ac.uk

Ung-type uracil-DNA glycosylases are frontline defenders of DNA sequence fidelity in bacteria, plants, and animals; Ungs also directly assist both innate and humoral immunity. Critically important in viral pathogenesis, whether acting for or against viral DNA persistence, Ungs also have therapeutic relevance to cancer, microbial, and parasitic diseases. Ung catalytic specificity is uniquely conserved, yet selective antiviral drugging of the Ung catalytic pocket is tractable. However, more promising precision therapy approaches present themselves via insights from viral strategies, including sequestration or adaptation of Ung for non-canonical roles. A universal Ung inhibition mechanism, converged upon by unrelated viruses, could also inform design of compounds to inhibit specific distinct Ungs. Extrapolating current developments, the character of such novel chemical entities is proposed.

Executive Summary

- *Introduction*

  Ungs are essential enzymes at the forefront of pathogenic states, both defending cells and being co-opted by pathogens.

- *Ung structure and mechanism*

  Ung is an exquisitely specific catalytic domain, but pre-catalytic variations promise specificity between targeting the host enzyme and pathogenic variants.

- *Origins and significance of uracil-substituted DNA in pathogenesis*

  Viruses have evolved to silence, or co-opt Ung to facilitate pathogenic states.

  Understanding these origins allows search for novel Ung-interacting proteins.

  Knowledge of Ung-interacting protein interfaces can facilitate drug discovery.

- *Motivations and contexts for therapeutic interventions targeting Ung*

  Understanding the significance of Ung in diverse pathogenic states permits suitable intervention to be designed and implemented.

- *Convergence on a universal Ung-inhibitory mechanism by unrelated virus proteins*

  Natural protein-protein interactions that irreversibly inhibit Ung activity, by convergence on a common mechanism have independently evolved at least 3 times from different protein architectures.

  The naturally evolved inhibitor proteins do not target uracil specificity of Ung.

- *Synthetic selective Ung inhibition and future trends*

  Drug design centred upon uracil-analogues has shown specificity is achievable.

  Compounds featuring uracil and hydrophobic tails are not ideal drug candidate molecules.

  Compounds targeting Ung protein-protein interactions have shown selective potency against poxviruses.

- *Conclusions*

  Eschewing uracil-specificity to emulate features of natural protein-protein interactions, promises novelty in selective drug discovery to target Ungs in pathogenic states.

Introduction

Ung is on the frontline of DNA repair

Ung, is the archetypal uracil-DNA glycosylase and the first enzyme to be described with a role in DNA damage repair [1,2]. Ung is a DNA binding protein with its primary and ostensible role in limitation of DNA damage, via rapid and exquisite deoxyuridine-triggered initiation of the base excision repair (BER) response [3,4]. One of the most frequent DNA mutation events is the conversion of cytosine residues in DNA, to uracil, by spontaneous hydrolytic deamination [5]. Unless this mutation is corrected prior to replication, a cytosine to thymine transition mutation will be recorded upon replication. All cells must therefore contain an enzyme capable of recognising uracil in DNA, and efficiently removing it, to ensure the longevity of DNA fidelity.

For Ung-processed DNA, repair is just one possible fate

Ung is also known to play a role in the cellular response to pathogens. Ung is described as a frontline defence molecule providing innate immunity in both eukaryotes and prokaryotes, in which it may act as a restriction enzyme on uracil-substituted viral DNA [6,7]. Ung is furthermore recruited to an essential role in humoral immune system development and maturation processes in mammals: Somatic hypermutation (SHM), and, Class Switch Recombination (CSR) [8,9]. It is therefore unsurprising that during active pathogenesis, host cellular Ungs are found to be variously repurposed or targeted by viruses, or else specialised Ung variants are encoded by viruses: These themes will be explored in this article. Given the varied role of Ungs in pathogenesis, endeavours to develop inhibitors directed towards the uracil binding pocket in the catalytic centre of the enzyme [10-15], would appear obvious.

Targeting Ungs need not involve uracil analogues

Uracil or its analogues however, as will be discussed in this article, may not be the only route to developing selectivity against Ung. Not least because the catalytic pocket is highly conserved in described structures of the Ung catalytic domain. Therefore, development of selective novel chemical entities targeting other identified characteristics of Ungs, such as the uniquely shaped DNA binding cleft or catalytically important features distal to the catalytic pocket, may provide a lower barrier and a wider range of possibilities for drug development. Consideration of other Ung contexts, including their known protein-protein interactions (PPIs), provides additional perspectives from which to inform development routes for drugs to target any particular Ung.

Naturally informed routes to selective Ung-inhibitor design

Considerations for targeting Ungs in drug discovery include the fact that some viruses encode Ung variants of their own. In spite of a high degree of conservation between all Ungs, quite reasonable initial selectivity with novel chemical entities based upon uracil-analogues has been shown [10-15]. However, developing these to the point where they show no off-target effects will be challenging, given that uracil-specificity is not greatly variant across known Ungs. Furthermore, achieving favourable toxicity profiles may prove just as challenging, given the potential of uracil analogues to intercalate in DNA. Finally, in pursuit of selectivity, the targeting of variant hydrophobic crevices interior to the enzyme via the uracil-specificity pocket [12-14] presents drug design challenges of its own, in terms of solubility profiles. Therefore, looking instead at features not involving uracil-specificity could prove more favourable, as will be explored in this article.

One of the conserved motifs essential for Ung catalysis is located outside the uracil-specificity pocket, and is more sequence diverse between evolutionarily distinct Ungs. This sequence variation has been observed to underlie structural dissimilarity and mechanistic differences, with relevance to pathogenesis. This opens the door to selectivity, therefore Ung could be targeted in cancers, microbial pathogens, and parasites, to potentiate the action of other drugs. Also, virally encoded Ungs present non-enzymatic features involved in PPIs that are important to the virus survival strategies. Viruses that do not encode Ung, but that manipulate the host enzyme to their advantage, similarly rely on PPIs to achieve those ends.

Another rationale relevant to drug targeting of Ungs, is to note that naturally evolved strategies to silence Ung activity do not target the uracil-specificity of these enzymes. Instead, sequestration and inhibition of Ung, is via Ung-specific DNA mimetic contacts from viral proteins [16-21]. Viral antagonism to greatly limit or silence Ung activity, is mediated via depletion or physiologically irreversible sequestration of Ung. The genes for Ung inhibitory proteins have their origins in evolutionarily distinct types of viruses, whether infecting prokaryotes or eukaryotes [22-25]. These inhibitory proteins populate architecturally discrete classes, and even within a single class there is significant sequence variation: To the point that it is not straightforward to search for potential structural relatives in other genomes. It is of note that the poxviruses encode an Ung that has a key structural variation in the essential pre-catalytic loop, rendering it immune to these virally evolved inhibitors (reviewed in [26]). Any of the presently known viral inhibitor proteins will nevertheless inhibit any compliant Ung, albeit with individual variation in the finer chemical details of the association.

The potential is that novel chemical entities developed to exploit aspects of this universal mechanism of Ung inhibition, could be diverse in design and in the details of their action. These could be made highly specific to any individual Ung of interest, and there is no reason why this could not include any outliers, such as Ung encoded by poxvirus, or β-herpesviruses. Of relevance to the wider scope of this article, poxvirus Ung forms a druggable PPI with the viral DNA polymerase (the unique properties of poxvirus Ung, and its association with the viral polymerase are reviewed in [26,27]) [28].

The remainder of this article will therefore consider in more detail: (1) Ung structure and mechanism, (2) origins and significance of uracil-substituted DNA in pathogenesis (3) motivations and contexts for therapeutic interventions targeting Ung, (4) a universal mechanism for Ung inhibition by architecturally unrelated proteins encoded by viruses, and (5) the known Ung inhibitory synthetic compounds reported to date. To conclude, the broader context and future synthetic possibilities for drugging Ung selectively in light of this information will be considered.

1 Ung structure and mechanism, a slight comparison

For the purposes of this overview, Ung structure and mechanism will be considered essentially universal, with limited examples selected to highlight differences in certain pathogens versus the human enzyme. Also, for the purposes of this article, unless specifically referring to the human nuclear isoform UNG2, the Ung catalytic domain will be referred to universally as 'Ung'.

Ungs (KEGG orthology: K03648) were the first described, and are the most ubiquitous, type of the uracil-DNA glycosylase (UDG) superfamily of DNA N-glycosidic hydrolases (EC: 3.2.2.27). Thus, Ungs comprise family branch 1 in the UDG superfamily. The UDG superfamily has at least 6 major branches, with general conservation of motifs important to ligand binding, substrate selectivity and catalysis, as well as overall architectural similarities especially at the catalytic centre (comprehensively reviewed in [29]). New variants of almost all UDG families are regularly reported in the literature [30,31], including variants of family 1 enzymes (i.e. Ungs) [32,33]. Ung variants from pathogens may have relevance as targets for therapeutics development to combat pathogenesis [26-28,34-37], and in some cases novel chemical entities of interest have been reported [10,11,14,28,38] (Table 1).

Table 1 Variant Ungs encoded by pathogen genomes, with properties of relevance to drug discovery.

Pathogen Ung structural variant	**Identified difference with respect to UNG2**
Herpesvirus (α- / HSV-1)	Shorter pre-catalytic loop; uracil-analogue inhibitors described [10,11,14].
Herpesvirus (β- / HCMV, Roseolavirus)	Significantly variant pre-catalytic loop; no structural information or inhibitors described to date.
Herpesviruses (γ- / EBV, KSHV)	Elongated/structured pre-catalytic loop is essential to replication [34,35].
Poxviruses [vaccinia]	Divergent features/ variant pre-catalytic loop, resistant to phage-mediated inhibition; potent PPI inhibitors identified [26-28].
Mycobacterium tuberculosis	Atypical catalytic structure, weakened phage-mediated inhibition [36,37].
Plasmodium falciparum	Divergent sequence; uracil-analogue inhibitors identified [38].

In pathogen encoded Ungs, structures are sufficiently diverged in important catalytic respects, that their selective inhibition might present more options for drug design. For example, the variants encoded by γ-herpesviruses are sufficiently different in important respects with regard to catalysis as to present unique possibilities for their selective inhibition (Table 1, Figure 1, Figure 2) [33,34]. Another example of a significantly variant Ung, is that encoded by vaccinia virus. Importantly the vaccinia Ung is indispensable to viral replication [39] (vaccinia Ung is reviewed in [26,27]). In the description that follows, the *Ung catalytic domain* is the inferred canonical structural form of UDG family 1.

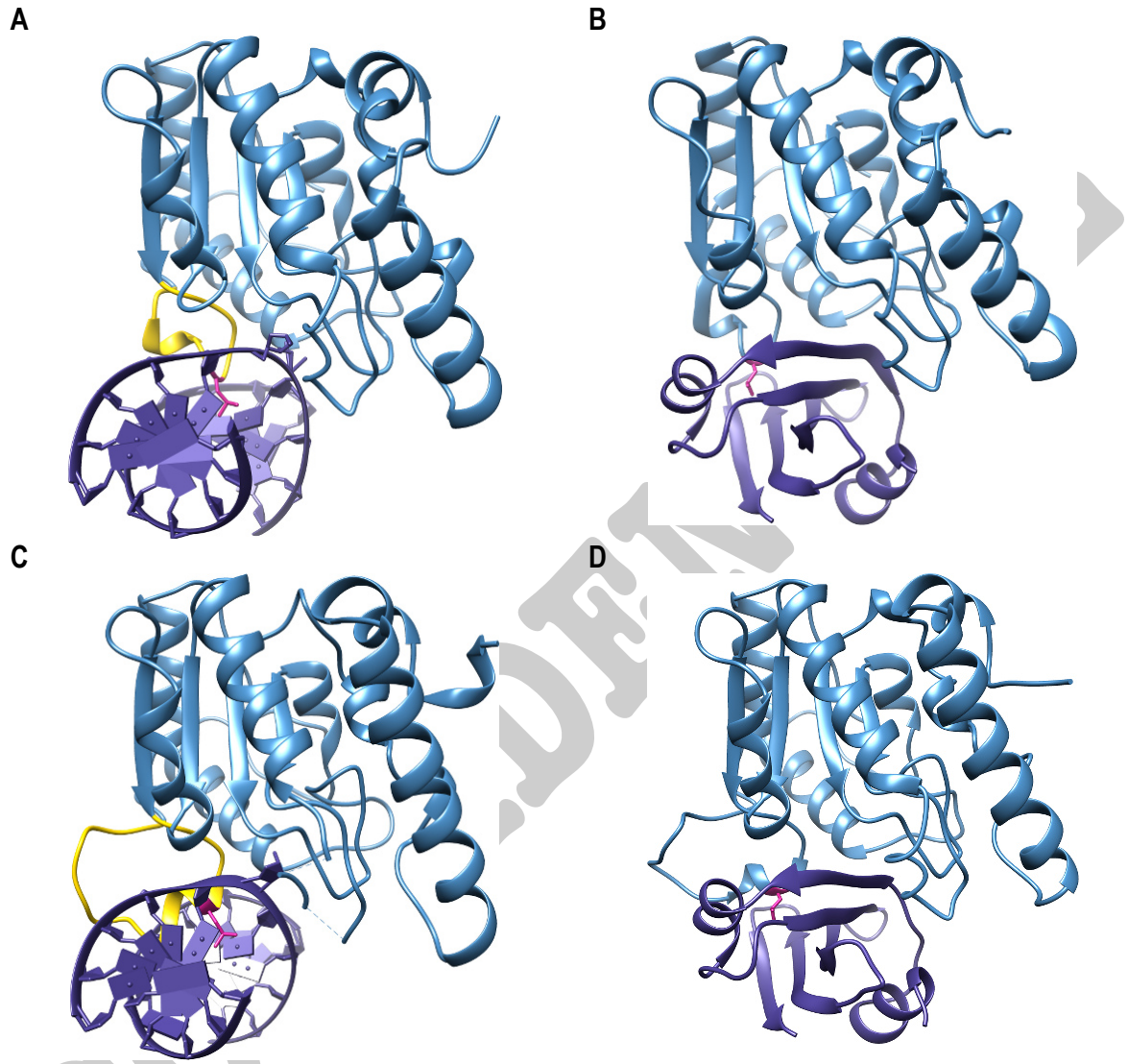

Figure 1 The Ung domain makes specific contacts to dsDNA, some of which are mimicked by viral Ung-inhibitory proteins. Structure cartoons, for illustration, rendered in Chimera*:
A - 1SSP, **B** - 1UGH (UNG2 complexes with respectively, dsDNA, and Ugi from *Bacillus* phage PBS1).

C – γ-herpesvirus Ung complexes: 5NNU (KSHV [HHV-8] with dsDNA), chains A, S and T only; **D** – 2J8X (EBV [HHV-4] with Ugi from *Bacillus* phage PBS1), chains A and B only.

Poses in panel B, and C are relative to panel A, via structure match to chain E from 1SSP. In panel B, pose is relative to panel D, via structure match to chain B from 2J8X.

The Ung minor groove DNA binding loop is shown in gold: The extended loop of the viral Ung in 5NNU (panel C) interacts with the DNA more intricately, compared to UNG2 (panel A). The apical leucine (also referred to as the pre-catalytic leucine residue) is shown in pink.

* Pettersen EF, Goddard TD, Huang CC et al. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13), 1605-12 (2004).

The Ung catalytic domain

Ung polypeptide backbone structures from the pdb may be readily aligned via structural superposition. The architecturally well conserved Ung catalytic domain is observed to be generally superimposable between representatives of evolutionarily disparate sources [29]. However, variant Ungs nevertheless differ in several respects, including alterations in even the most critical signature motifs involved in substrate engagement and catalysis, located in the C-terminal 200 ±c.20 residues (Figure 2).

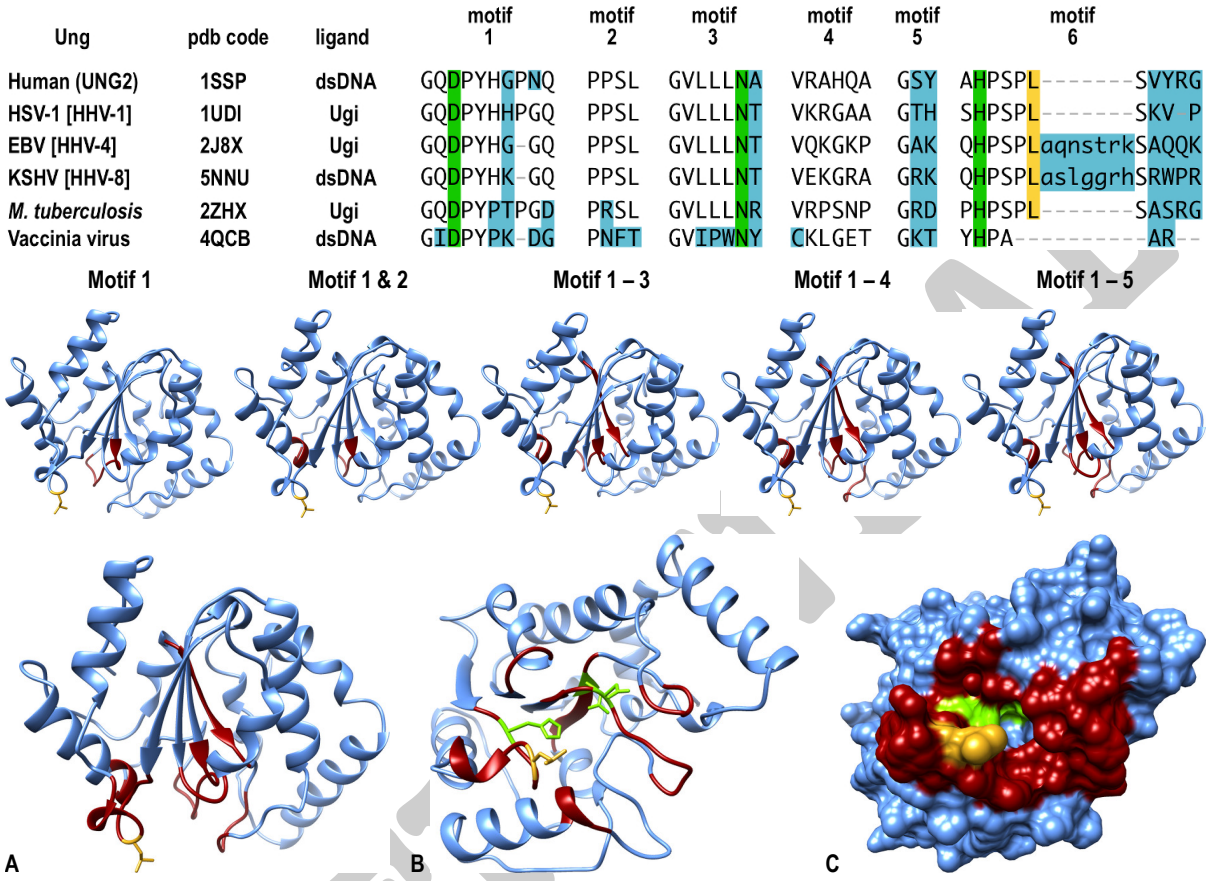

Figure 2 Conservation and variation in the key motifs involved in Ung substrate engagement and catalysis.

Sequence alignment: The structure of the canonical Ung domain is highly conserved across large evolutionary distances but tolerates structural variation (e.g. HHV-4, HHV-8, and vaccinia) and key motifs show spatial and chemical variation (sky-blue background) that could be exploited for drug discovery. Motif 6 comprises the minor groove DNA binding loop, also referred to as the pre-catalytic loop (apical leucine residue on gold background, with vaccinia being the exception; note the vaccinia loop is structurally variant and also misses residues in deposited structures leading to uncertainty in alignment), which is sequestered by viral proteins that inhibit Ung. Note that the Ung catalytic triad (green background) is invariant in all examples shown here. The Ung domains aligned are in complex with either dsDNA (1SSP, 5NNU, 4QCB) or with the viral Ung inhibitory protein Ugi (1UDI, 2ZHX). Structural alignment of each structure with 1SSP as the reference, was initially assisted by global superposition of Ung backbones using the software Chimera, and default settings.

Structure depictions: The structure 1SSP (UNG2 in complex with dsDNA) with the DNA removed for clarity, is depicted as a cartoon in light blue, and motifs 1 through 5 from the sequence alignment are coloured in red sequentially from left to right. In the lower row, cartoon A depicts all 6 motifs from the above sequence alignment, shown in red. In all cases, the pre-catalytic loop leucine residue is shown in gold for orientation purposes. Cartoon B in the centre of the lower row is rotated so that the DNA binding surface of UNG2 is in the foreground, and the catalytic triad residues have been coloured bright green. The same view is depicted as a surface on the right of the lower row, labelled C.

Ung substrate engagement, and uracil specificity

Ungs catalyse the hydrolytic cleavage of the N-glycosidic linkage between the pyrimidine base uracil, and the deoxyribose; catalysis can proceed in ssDNA or dsDNA substrates. In ssDNA the reaction is highly efficient, except in very short oligonucleotides, but even a dinucleotide can be utilised as a substrate [40]. In dsDNA the reaction is observed to be relatively slower than in ssDNA. As described below, this is likely due to necessary enzymatic motions and induced structural changes in the double helix. Hydrolysis of the N-glycosidic bond takes place in a concave pocket, exquisitely specific to uracil and some analogues, buried inside the Ung DNA binding cleft [3,41]. In order for uracil to enter the catalytic pocket, the target nucleotide must be extruded from the helical base stack [4]. This motion is facilitated by the dynamic structural nature of the search mode of Ung [42,43].

The structures of Ung catalytic domains in complex with dsDNA, demonstrate a subtle global closure of the domain upon occupation of the Ung DNA binding cleft (Figure 3). Detailed analysis of similarities and differences in domain closure under different circumstances implicates positions of possible interest in drug development [13,36,44]. The closure due to dsDNA has also shed light upon the mode of interrogation of DNA bases by Ung [42,43].

During substrate search mode, Ung engages dsDNA whereupon it introduces a local distortion to the double helix due to the shape of the Ung DNA binding cleft. Since Ung enzymatic activity does not require an external source of energy, the impact (via electrostatic attraction) and strain forces alone prove sufficient to drive the subtle and concerted motions that interrogate and catalyse substrate DNA. Due to the local dsDNA distortion induced by Ung binding, breathing motions of the DNA bases are exaggerated in the vicinity of dsDNA bound Ung: During breathing, the base pairs spontaneously break (due to distance and geometry fluctuations induced by natural compression/torsional motions of the DNA at large) then rapidly re-form. Base pairs involving adenine, namely thymine (under normal circumstances) or more rarely uracil (via misincorporation of deoxyuridine, except in specific biological contexts referred to in other parts of this article), are especially susceptible to exaggerated breathing during Ung binding. Also susceptible is guanine when mismatched with uracil.

The exaggerated breathing of DNA bases permits the minor groove DNA binding loop of Ung to swing in and occupy the position vacated by a displaced base (as will be discussed, viral inhibitory proteins use DNA mimetic stealth to irreversibly trap this loop). In an Ung complexed with dsDNA, this loop motion also serves to boost progress of the displaced base towards the Ung catalytic pocket. The resulting arrangement results in a fleeting base pair mimetic contact to the remaining intrahelical base from the residue sidechain at the apex of the Ung minor groove DNA binding loop (typically, but not exclusively, leucine. N.B. This same apical leucine is irreversibly hydrophobically sequestered by viral Ung inhibitory proteins). With the lone intrahelical purine base thus stabilised, residency of the extruded pyrimidine in the Ung pre-catalytic complex, is prolonged. For brevity, this Ung leucine is alternatively referred to as the ‘pre-catalytic leucine’, and the minor groove DNA binding loop upon which this leucine resides, is elsewhere referred to as the ‘Ung pre-catalytic loop’.

At the mouth of the Ung catalytic pocket, the presence of any pyrimidine 5-methyl group would meet with steric hindrance from a phenylalanine side chain. This allows a conserved Ung tyrosine residue to favourably stack with extruded thymine bases at a site external to the catalytic pocket. This blockade prevents the entry of thymine into the catalytic pocket under normal circumstances [3,13,43]. The result is that thymine will return rapidly to the interior of the DNA double helix. In contrast, the lack of a bulky methyl group at the 5-position of the pyrimidine ring means that uracil will pass unhindered into the catalytic pocket, where immediate water-mediated nucleophilic attack will efficiently remove it from the deoxyribose. Due to the structural arrangement of the catalytic pocket and its immediate exterior, cytosine will not be presented for catalysis due to unfavourable chemical properties, and purines would be sterically excluded from interrogation, altogether [3,4,42,43].

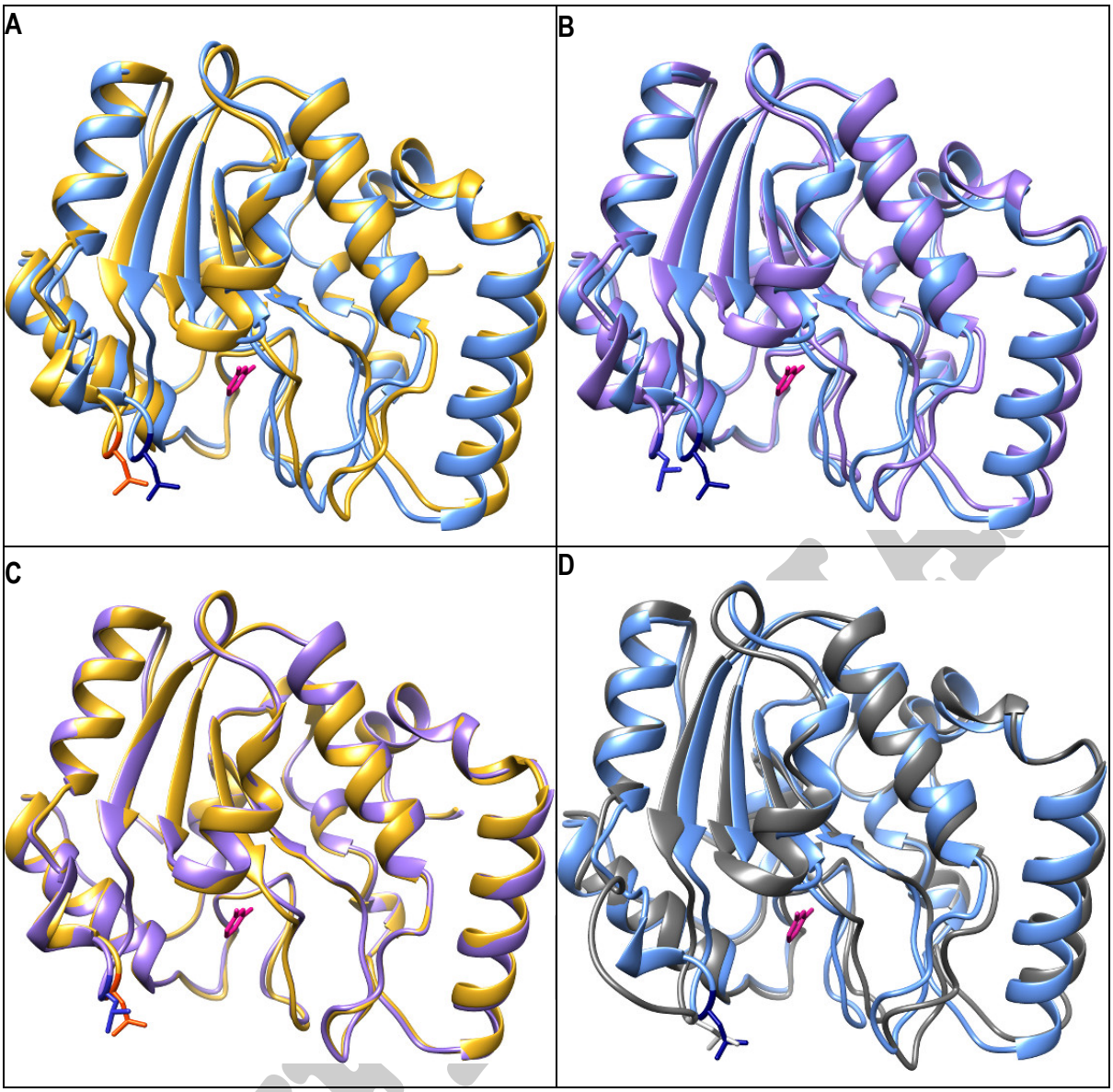

Figure 3 The Ung domain structure closes around dsDNA upon complexation (panel A), but not when viral Ung-inhibitory proteins complex with Ung (panels B – D). 1AKZ (gold) = apo form of UNG2; 1SSP (chain E, blue) = UNG2 complexed with dsDNA; 1UGH (chain E, purple) = UNG2 complexed with Ugi from *Bacillus* phage PBS1; 5JK7 (chain D, dark grey) = UNG2 complexed with Vpr from HIV-1.

For clarity, dsDNA, and viral inhibitor proteins, respectively, have been removed. N.B. Uracil base has been included (bright pink) in the uracil-specificity (catalytic) pocket to appreciate its distance from the apical leucine (shown as stick) of the Ung pre-catalytic minor groove DNA binding loop, which is targeted by viral inhibitor proteins of Ung.

Cartoon structures, rendered in Chimera, show UNG2 only: **A** – 1AKZ/1SSP, **B** - 1SSP/1UGH, **C** – 1AKZ/1UGH, **D** – 1SSP/5JK7

2 Origins and significance of uracil-substituted DNA in pathogenesis

Interestingly, uracil in a DNA context is important to pathogens both in terms of its promutagenicity, its effects on efficiency of viral DNA manipulation, and in avoidance of host surveillance. Uracil in a canonical DNA context is normally targeted for excision and repair. Uracil in canonical DNA occurs via either the spontaneous deamination of cytosine, or occasional misincorporation of deoxyuridine (levels of dUTP in normal cells are strictly managed to minimise the likelihood of its utilisation by DNA polymerases). However, the cellular immune repertoire may also introduce uracil, by enzymatic deamination of cytosine, for diverse fates. Furthermore, viruses of prokaryotes are also known to antagonise Ung, and even to deliberately bias or re-program the nucleotide biosynthesis pathways to preferentially utilise uracil in DNA as a replacement for thymine.

Cytosine is enzymatically deaminated in the innate cellular response to pathogen DNA

Enzymatic deamination of cytosine to uracil in human cells, can be promoted by AID [Activation-Induced (Cytidine) Deaminase] in SHM/CSR [8,9], or as considered elsewhere in this article, by APOBEC proteins [the Apolipoprotein B mRNA Editing Catalytic polypeptide-like family] in innate immunity contexts [45, 46]. This effect of cytidine deaminases in the innate immune context (e.g. APOBECs), is to severely compromise the sequence fidelity of pathogen DNA, also rendering it a substrate for enzymes such as UNG2. Problematically however, under particular circumstances this response appreciably amplifies the rate of cytosine deamination in general. This is thought to contribute to hypermutation, leading to carcinogenesis: This is a factor of relevance to any potential therapeutic interventions targeting UNG2 [47].

Viruses can control uracil-substitution for survival and persistence strategies

In viruses, uracil-DNA appears to be of some significance. For example, in HIV-1, integration of reverse transcribed viral DNA is sensitive to the degree of uracil substitution [7,48-50]. Furthermore, the level of uracil-DNA and its rate of removal are finely controlled by HIV-1 encoded proteins [51-57], possibly to promote a tolerable level of hypermutation to tune viral fitness in the host [46,58]. This important potential Achilles heel in the HIV-1 persistence strategy is elaborated elsewhere in this article.

The α-, β-, and γ-herpesviruses all carry a copy of a variant Ung, and in the latter two classes this appears essential to efficient viral lytic phase replication [59,60]. In viral genomic latency of HSV-1 (an α-herpesvirus), long periods in the form of an episomal DNA are typical. Latency occurs in terminally differentiated neuronal cells of the trigeminal ganglion, which do not provide maintenance via DNA repair. Therefore, it is conjectured that the role of Ung in reactivation from latency, is to make repairs to the viral genome during or in emergence from latency [61]. In γ-herpesviruses, the virus encoded latency associated nuclear antigen (LANA) has been demonstrated to co-opt UNG2, presumably for maintenance during latency or in pre-lytic repair [62].

Bacteriophages exhibit an ambivalent relationship to uracil-DNA, which nevertheless in all cases requires inhibition of Ung activity. Phages maintaining an entirely uracil-substituted viral genome [63-65] or those sensitive to Ung-promoted nicking of spontaneous occurrences of uracil [66-70], must encode a gene product preventing uracil-DNA repair. Known examples stoichiometrically inhibit Ung directly (Table 2). There is biological analogy, therefore, between the retroviral requirement for accessory proteins manipulating Ung, and requirements of phage for an Ung-targeting protein. That is, the uracil-substitution state of the viral DNA is an important factor in successful viral propagation.

Bacillus phage PBS1 comprises a genomic DNA of 0.25Mb wherein base pairing to adenine is entirely serviced via deoxyuridine instead of thymidine [63]. Ung has predictably devastating effects on such uracil-substituted naked DNA in the absence of other factors, reducing dsDNA to short oligonucleotides [6,71]. This degradative effect of Ung on highly uracil-substituted DNA is due to proximal removal of uracil on both duplex strands resulting in appreciable rates of spontaneous backbone nicking via beta elimination throughout the numerous abasic sites thus generated. In a cell, this catastrophic disintegration of the genome would be accelerated by the action of the BER response in the wake of

Ung. BER first creates endonucleolytic breaks in the DNA phosphodiester backbone 5' to the abasic site due predominantly to the action of host enzyme APE1/endonuclease IV. However, curiously in *Bacilli*, PBS1 phage replicate seemingly without any delays or untoward effects [66,72]. No Ung activity could be detected in viral lysates and this was ultimately traced to a single open reading frame in the phage genome encoding a protein, Ugi, which stoichiometrically and irreversibly inhibits Ung [22,73].

Table 2 Viruses and their utilisation of Ung.

Virus type	Relevance to Ung	Interaction
Bacteriophages: Phages with Uracil-DNA genome [*Bacillus* phages AR9, PBS1, and probably *Yersinia* phage ϕR1-37, and *Staphylococcus* phage S6], or with terminal-protein primed genome [picovirinae including *Bacillus* phage ϕ29]; phages with ssDNA nicked genomes [i.e. PBS1 and other phages, e.g. coliphage T5]. [63,66-70]	Ung inactivated by virally encoded protein. This prevents Ung from acting as an antiviral restriction factor. [22,55,67]	DNA mimetic interactions from viral protein (Ugi in AR9 and PBS1; p56 in picovirinae) with residue-specific hydrophobic sequestration and inactivation of Ung. [16-21]
Retroviruses (HIV-1) Uracil-substitution of the HIV-1 reverse-transcribed genome demonstrated as an important factor for integration compliance [7,48-50], and in controlled mutagenesis for fitness in adaption to host [46,58].	Human Ung nuclear isoform UNG2 inactivated by protein-protein complexation to virally encoded Ung-targeting protein Vpr; also UNG2 is kidnapped by Vpr, and potentially also integrase, for diverse fates in virus maturation and cannot act as an antiviral restriction factor. [55]	1) DNA mimetic interactions from viral protein Vpr, with residue-specific hydrophobic sequestration and inactivation of UNG2. [21] 2) Interactions that employ UNG2 in an active form within the virion, possibly via integrase. [47,56]
Poxviruses (vaccinia) [37,39]	Specialised viral Ung. [26,27]	Viral Ung essential for cytosolic viral replication. [39]
Herpesviruses (α- e.g. HSV-1, β- e.g. hCMV, and γ- e.g. EBV,KSHV) [59-62]	Specialised viral Ung in lytic phase, and utilisation of UNG2 in the latent phase maintenance/reactivation. [3, 33,34]	Viral Ung is implicated in latent maintenance in HSV-1 [61], while in hCMV and γ-herpesviruses viral Ung is an essential part of the lytic phase replication machinery [34,60,61]. In the latter class, UNG2 is recruited to assist in latency-associated processes [62].

Control of DNA nicking in viral replication strategies requires viral command of Ung

It makes some evolutionary sense to consider that Ung inhibition by viral proteins might have first arisen to protect another apparently more common mechanism of phage replication, not necessarily associated with utilisation of uracil-DNA: Specifically, the use of orchestrated single strand nicks [67-70]. Such programmed single strand breaks appear to be a feature of unrelated bacteriophage genomes, relevant to specialised mechanisms of replication or/and encapsidation of nucleic acids. Stochastic deamination of cytosine leads, via Ung activity, to BER initiation and the appearance of random nicks. These random interlopers apparently poison finely-tuned, programmed-nick-dependent processes. This has a sufficiently negative effect on population size and fitness of viral progeny, that Ung-inhibiting proteins are deployed by such viruses [23,25,73].

Uracil-substituted DNA in viruses as an offshoot of extant Ung silencing strategies

Compromise of orchestrated nicking processes however, would appear to be a weakening phenotype, rather than a lethal one. Therefore, this lends logic to how or why viruses came to employ the curious and seemingly suicidal strategy of utilising entirely uracil-substituted DNA. If Ung-inhibition is deployed for nick-protection purposes, then global uracil-substitution of DNA can be adopted without danger of catastrophic genome disintegration.

Uracil-substituted DNA imbues a different set of survival advantages, such as immunity to some restriction enzymes [74] and, by implication, firewalling from the cellular transcription program [65,75]. The latter tactic is analogous to well-studied transcription program isolation mechanisms employed by

phages such as coliphage T7 [76]. The disadvantage is the metabolic trade-off, that viruses relying on uracil-DNA are genomic giants, encoding entire nucleotide biosynthesis modifying pathways, uracil-DNA specific transcription apparatus, and their own uracil-biased DNA polymerase [65,77].

Ung inhibition has also been reported for the coliphage T5, which does not utilise uracil-DNA, but no reading frame encoding an Ung-inhibitory protein has been identified in the T5 genome to date [25]. The Ung inhibitory requirement in T5, is instead presumably linked to its programmed genomic single-stranded nicks [70]. Interestingly when considering this line of reasoning, the giant phage PBS1 features programmed genomic nicks, as well as comprising fully uracil-substituted DNA [63,69].

Base substitution and Ung silencing by viruses compromises genomic encoding potential

Precariously, in ultimate survival terms, inhibition of Ung activity effectively leads to abrogation of the most efficient pathway to uracil-DNA repair thereby steering viruses toward potential genomic extinction. Preventing the removal of deaminated cytosine applies inevitable attrition to the %GC content of genomes, due to fixing of C to T transition mutations through serial rounds of replication. The known uracil-DNA phages are all at the lower end of the genomic %GC spectrum (<30% globally, and significantly lower in some loci). Lowering %GC tends toward a series of survival boundaries: below each of which, the encoding of certain natural amino acids can no longer be achieved via the Universal Genetic Code. It is of note that many deposited phage genome sequences exhibit dramatically low global %GC, but published information about programmed nicks or genomic uracil-substitution is limited to very few phages. Known Ung-inhibitory encoding sequences are presently limited to those discussed in this article. Additionally, a far more complex parasite, *Plasmodium falciparum*, has a global genome %GC of only around 20%, yet counter intuitively with respect to the above reasoning, encodes its own Ung variant (Table 1). Plasmodium Ung has been a focus for drug discovery [38].

However, the tendency to lose cytosine positions to thymine, can in phages at least, also be explained by adoption of other base modifications to escape restriction enzymes. For example, adoption of 5-methylcytosine as a substitute for the canonical base [78] (note that cytosine, and other pyrimidine derivatives, including hydroxymethylated, and glycosylated forms, have also been observed [79,80]), turns out to be a comparably dangerous strategy to the adoption of uracil-substituted DNA. This is because the rate of deamination of 5-methylcytosine is greater than that of cytosine, and the product is thymine rather than uracil [81]. To avoid fixing the mutation for posterity requires the pre-replicative intervention of mismatch-DNA repair. The upshot of this, is the sidestepping of a requirement for Ung-mediated DNA repair. Cytosine derivatisation may therefore present a risky but workable general solution to the genomic %GC attrition problem. It may nonetheless be conjectured that the accelerated rate of deamination could potentially outstrip the capacity for mismatch repair (a slower process than Ung-BER) during an active phage infection. The likely picture is that Ung inhibition as a survival strategy will lead to inevitable transition mutations, and in turn, to pressures that adaptively evolve viral genomic sequences to retain, or substitute for, essential functions.

Horizontal transfer assists survival and mediates pathogenesis

Of further relevance, is the possibility that, whether or not a phage genome is sensitive to Ung, the potential to inhibit Ungs may nevertheless be acquired horizontally. PBS1 is a transducing phage capable of infecting diverse *Bacilli* [82,83], as is another giant uracil-DNA phage, S6, with broad *Staphylococcal* specificity of infection [84]. The giant uracil-DNA phages: AR9 (also a transducing phage [85]), and PBS1, encode the Ung-inhibiting Ugi protein. A structural relative of the Ugi protein is Sa-Ugi, encoded by a phage-derived DNA in MRSA (Methicillin-resistant Staphylococcus aureus) *Staphylococcal* genomes. The origins of Sa-Ugi, which is mentioned in the following section, may therefore be via a phage like S6.

The S6 genome has not been sequenced, but it is feasible that Ung-inhibitory genes may be horizontally transferred to bacterial genomes, and subsequently acquired by unrelated phages. Indeed, it is the case that strains of bacteria are known with unusual pyrimidine nucleotide biosynthesis and utilisation properties, related to lysogenic phage [86]. Thus, development of uracil-DNA or other non-

canonical DNA pyrimidine genome properties, in viruses, may be viewed as evolutionary survival strategies with origins in the inhibition of Ungs. Of interest to drug discovery directed at Ung activity, this has relevance to possible potentiation of drugs in tumours and also in bacteria.

Extrapolating further, the suggestion is that uracil-DNA or/and Ung-inhibition mediated by phages could also be more widespread than is currently known [84,87], perhaps even including novel permutations of architecture and modes of interaction: This is important in the context of this review article, because any newly discovered natural Ung-inhibitory strategies might further inform new avenues to design novel chemical entities against Ung activity. Ugi, and Sa-Ugi, vary significantly when their sequences and structures are compared. Yet these natural differences, as well as differences that can be engineered (and may even exist in as yet unknown natural variants) could point to unique ways to target pathogen Ungs, or even the host enzyme [88]. Indeed, the sequenced p56 type Ung inhibitors vary in sequence within the conserved region (Figure 4).

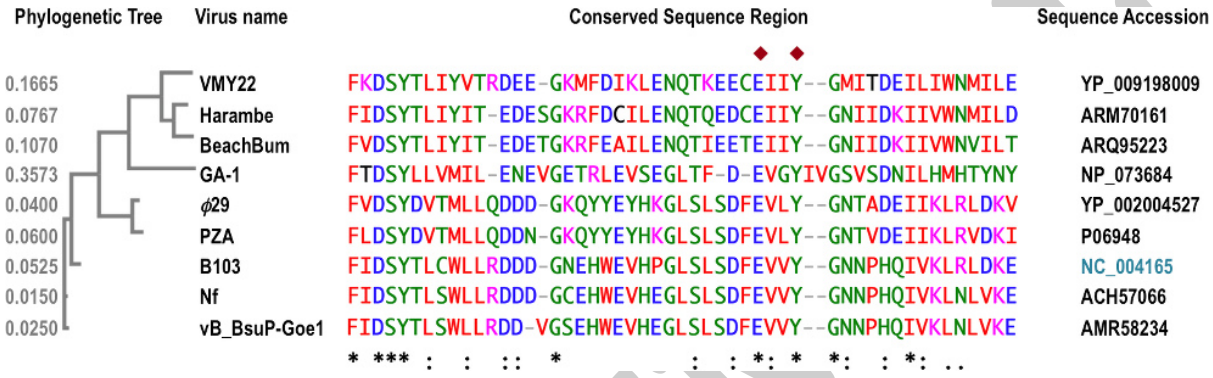

Figure 4 Variation within the conserved region of virus-encoded p56-type Ung-inhibitory proteins. In the rightmost column, the protein sequence accession number is given: note, that in the case of the B103 protein, this is a nucleotide sequence accession (cyan letters) because the p56 ORF does not appear in the genome annotation. The B013 p56 is the translation product of the reverse complement between nucleotides 488 and 658.

Sequence alignment was performed with default settings at the Clustal Omega web server, then manual curation was performed using the p56 protein structure as a guide in Chimera (inspection of chain D from 4L5N). Residue colours and scoring of matches in the form of symbols below the alignment are Clustal Omega defaults (Gonnet PAM 250 matrix). The pair of ◆ symbols above the alignment show the hydrophobic residues that, in the context of the dimerised protein, trap the Ung pre-catalytic loop leucine residue. The phylogenetic tree at the left of the figure was generated with Clustal Omega*, and is a default Neighbour-joining tree without distance corrections.

* Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.*, 27(1), 135-145 (2018).

3 Motivations for specific targeting of Ung variants via synthetic novel chemical entities

Considering differences in key Ung motifs, those in motif 5 (Figure 2) are likely sufficient to enable the development of specific novel chemical entities to target them.

Bacterial Pathogens

In the case of pathogenic states of commensal bacteria, inhibition of microbial Ung could in principle be used to potentiate the action of compounds that cause lethal DNA damage in the absence of Ung. The key residue type or geometry nuances between UNG2 and those of, for example, pathogenic Mycobacteria (Figure 2, Table 1) could permit the possibility of development of selective Ung inhibitors [35,36,44]. These differences include markedly different interaction profiles for Mycobacterial Ung, versus UNG2, when either is complexed with the virus-encoded inhibitor protein, Ugi [35]. Ugi, and mechanistically analogous proteins (Table 2) like p56, and Vpr, as well as Sa-Ugi (also described immediately below), are proposed in this article as mechanistic templates upon which to base new forms of Ung inhibiting novel chemical entities.

In one pathogen of acute healthcare interest, the picture is intriguing: MRSA Staphylococci harbour fragmented phage remnants within the pathogenicity island [24], including a gene that in recombinant experiments encodes a functional Ung inhibitor [20]. Could this Sa-Ugi gene be active? It would in all likelihood cause genomic sequence fidelity maintenance to be compromised in the host if its product Sa-Ugi were expressed, because it would suppress Ung base-excision repair. Such an effect might in part contribute to the extreme plasticity of SCCmec (staphylococcal cassette chromosome mec) pathogenicity islands [24]. It could be conjectured that Sa-Ugi expression is conditional on some metabolic status or stress. If any such triggers exist, then proto-therapeutic induction of Sa-Ugi might potentiate the action of DNA-damaging drugs that are potent in the absence of Ung activity. This type of strategy would result in selective toxicity to MRSA pathogens, which is an intriguing thought indeed.

Cancer

Related to the above conjecture, there is precedent, with notable success in potentiation of DNA-damaging drugs as a strategy in cancer chemotherapies, via poly(ADP-ribose) polymerase (PARP) inhibitors (reviewed in [89]). Considering the above, similar targeted deployment of UNG2 inhibitory compounds [7,8] to tumour sites is therefore also an attractive proposition.

Viral pathogens

Viral pathogens (Table 1, Table 2), such as all (i.e. α-, β-, and γ-) herpesvirus families [59-61] and the poxviruses [39], have genomically acquired and adapted Ung as a central and even indispensable part of the viral replicative or latency processes. Ung is therefore of importance to these viruses [90], and exploiting the catalytically important structural differences between host and viral Ungs presents a bona fide target for synthetic antiviral development.

a) Vaccinia

The poxvirus vaccinia Ung is radically variant, in that it has adapted to an essential role in viral replication processivity and fidelity, via PPI with the viral polymerase. Vaccinia Ung also contains a structurally altered pre-catalytic loop, which lacks an apical leucine [26]. Both of those features differ from almost all other canonical Ung domains, which feature a hydrophobic (leucine, or in some *Bacilli*, a phenylalanine) side-chain apical on the Ung pre-catalytic loop. Vaccinia Ung (also known as, viral protein D4) is also required as an essential subunit by the viral (protein E9) DNA polymerase: Without the involvement of D4 in viral replication, vaccinia is not viable [26-28,39], which presumably applies to all poxviruses given their overall similarity in these encoded subunits. Furthermore, UNG2 cannot rescue the situation for the virus, which in any case replicates in the cytoplasm (note again that UNG2 is an Ung isoform localised in the nucleus). The implications for antiviral potency are thus obvious, and indeed small molecule targeting of the D4-E9 PPI interface disrupts viral DNA synthesis in vitro [28] (and reviewed in [26,27]). The suggestion is poxviruses are highly mutable and able to cross species

barriers, therefore future therapeutic need is not out of the question, even if smallpox outbreaks may essentially be ruled out under normal conditions [91].

b) Herpesviruses

The three types of herpesviruses, α- (e.g. HSV-1, HSV-2, VZV), β- (e.g. hCMV), and γ- (e.g. EBV, and KSHV), regardless of host cell, all contain viral genome copies of Ung [59-61]. These viral Ungs differ in sequence from the host enzyme in all cases, but essentially retain the hallmark architectural conservation in structure observed across all known homologous canonical Ungs (Figure 2). The most overt variants are Ungs from the β-, and γ-herpesviruses. Structural information is lacking for β-herpesvirus Ungs, which precludes detailed analysis. In the γ-herpesviruses, Epstein-Barr (EBV) [HHV-4] and Kaposi's sarcoma-associated herpesvirus (KSHV) [HHV-8], the pre-catalytic loop of Ung is seen to be appended by a sequence motif of conserved length but variable composition (Figure 2). Structural studies reveal that upon DNA docking this elaborated motif adopts a conserved structure [33,34] (Figure 1). In engineered mutant forms of γ-herpesvirus Ung, lacking the loop motif appendage, infecting viruses are highly compromised in lytic phase replication [60,92] and equivalent mutants are rendered biophysically unstable [34]. Targeting the structural underpinning of the extended loop conformation, or its dynamics, with novel chemical entities, would be expected to result in a similar disabling phenotype: The viral lytic phase of γ-herpesviruses can thus be considered druggable via Ung.

The picture is somewhat complicated by the presence of dual OriLyt foci in γ-herpesviruses, each associated with support of viral replication in different cell types [93,94], thus the strategy may not be applicable to all infected cell types. However, given the variety of health complications (e.g. Infectious mononucleosis, Burkitt's Lymphoma, and Nasopharyngeal Carcinoma, in EBV; Primary Effusion Lymphoma, Multicentric Castleman's Disease, and Kaposi Sarcoma, in KSHV) caused by these viruses however, being able to target at least some of these debilitating, pathological, or carcinogenic conditions would nevertheless be worthwhile. In a final twist, γ-herpesviruses, co-opt UNG2 during latency for latency-associated maintenance processes and to preserve DNA sequence fidelity in the viral episome [62]. Therefore, mapping the UNG2 protein interactions with viral factors, opens the door to drugging the host enzyme as a therapeutic option. This would be expected to impact latent episome integrity to reduce viral load [7,8,15].

Ung inhibitors to counter virus-associated neurological decline?

Finally, both α- (e.g. HSV-1, VZV), and β- (e.g. HHV-6A, HHV-7) human herpesviruses have been implicated as aetiologic agents in dementias and other neurological conditions (reviewed in [95]). An increased array of options in small molecule therapeutics that prevent formation of mature virions would therefore be useful. Could small molecules efficiently targeting viral encoded Ung in neurons and beyond the blood brain barrier be conceived? In fact, the similarities with HIV-D, a viral dementia associated with HIV-1 infection are noted (reviewed in [95]), and fall into the same therapeutic category. As will next be discussed, drugging UNG2, or virally encoded factors associated with it, is an attractive option.

c) HIV-1

HIV-1 is a major human viral pathogen of relevance when considering potential therapeutic targeting of Ung (Table 1). In contrast to the viral examples discussed so far, HIV-1 does not encode an Ung, but rather it encodes proteins modulating Ung localisation, availability and function [96]. The HIV-1 encoded accessory protein Vpr is able to associate by formation of complexes with various human cellular protein targets, including UNG2 [51,97].

Interestingly, Vpr apparently serves multiple diverse ends with its UNG2 hostage: Predominantly, it targets UNG2 for destruction by associating with DDB1–DCAF-1, thus delivering UNG2 to the CRL4 proteasome degradation complex [21,53,55]. Vpr additionally transcriptionally downregulates Ung (Vpr is comprehensively reviewed in [98]). It is known that the degree of uracil-substitution of DNA is critical

for reverse-transcribed viral genome integration into host genomic material [7,48-50]. UNG2 could clearly interfere, which makes sense of its Vpr-mediated downregulation (including by specific transcriptional suppression [54]).

Intriguingly then, UNG2 also appears to be inserted into HIV-1 virions along with Vpr [52] or/and potentially via its independent association with HIV-1 integrase [99] This is suggested to make use of a non-enzymatic function of UNG2, along with the p32 subunit of the host RPA complex, important for viral fitness during reverse transcription and for the optimisation of cDNA integration [49,56,57]. In terms of the importance of UNG2-specific PPIs, this is somewhat analogous to non-enzymatic roles for both viral Ung, and UNG2, in herpesvirus replication strategies.

A chemical or peptide, designed to inhibit UNG2 association with Vpr, ought to be detrimental to HIV-1 fitness [100]. It is also noted here, that natural plant extracts acting as general inhibitors of Vpr have been reported [101]. Any general Vpr-inhibiting compounds making it into the clinic would have the potential to compromise HIV-1 viability fairly comprehensively [102]. It should be noted that Vpr, in common with other accessory proteins, is under selective pressure following initial infection and [presumably active] variants are easy to find in the databases (Figure 5) [103]. As will be explored next, UNG2 and its essential PPIs in HIV-1 maturation, are of parallel interest as potential targets for antiretroviral therapy.

Reasons to target UNG2 in the context of HIV-1

Curiously, it would appear that HIV-1 may also mediate its own genetic drift via virion located UNG2. However, UNG2 kidnapped by Vpr is likely to be enzymatically inactivated via stoichiometric inhibition [21,49]. It is therefore presumed that this active virion-associated form of UNG2 is a fraction reported to be recruited via HIV-1 integrase, and essential to viral genome integration [99]. Limited mutagenesis inside the virion involving UNG2 is enacted in concert with host APE1, and is likely at least partially reliant on the prior activity of another host restriction factor APOBEC3 (specifically, isoforms F or G, when recruited to the virion, in the absence of the viral antagonist HIV-1 accessory protein Vif) [46].

APE1 is part of the ER-associated SET complex (APE1, NM23-H1, TREX1) reported to associate with the HIV-1 preintegration complex, and important in HIV-1 defence against autointegration (a suicidal complication of retroviral DNA integration, wherein the viral DNA fruitlessly integrates with itself) [48, 104]. APOBEC3 isoforms F and G, are molecules of innate cellular immunity against viruses that enter the cell. APOBECs cause deamination of cytosine bases in ssDNA at replication or transcription complexes [45,47], and the resulting uracil would be a substrate for UNG2, normally leading to viral genome disintegration, as described earlier. However, HIV-1 retains its RNA genome within the capsid core particle, thus enacting reverse transcription in situ, which must aid in shielding the viral nucleic acids from DNA sensing (a trigger of innate immune response) and thus innate immune compromise [105,106].

HIV-1 could therefore be described, analogously to DNA phages using non-canonical pyrimidines such as uracil, as living on the edge: Using its multi-tasking accessory proteins, Vpr and Vif, to deplete cellular resistance factors that would normally destroy it. HIV-1 re-contextualises UNG2 and APOBEC proteins, as well as the SET complex, to optimise its survival [107,108]. Fractions of these host restriction factor populations remain active, apparently deliberately recruited within the capsid. The virus is thereby pushed even closer to the brink of extinction and yet is able to profit from selection of sub lethal hypermutated genome states [46,58]. As an antiretroviral strategy with novel chemical entities, targeting UNG2 PPIs with the HIV-encoded proteome, would therefore be expected to promote a twofold effect: First, to negatively impact the probability of viral genomic integration due to the requirement for active captive UNG2 by viral integrase, and second, to radically skew genetic drift towards non-viability in viral progeny due to consequent unrepaired cytosine deamination in viral cDNA produced by APOBEC3F/G.

4 Convergence on a universal Ung-inhibitory mechanism by unrelated virus proteins

Both prokaryote-specific and eukaryote-specific viruses, inhibit, antagonise, or co-opt Ung, as would be expected of a ubiquitous molecule of innate cellular immunity (Table 2). Stoichiometric and irreversible inhibition of Ung enzymatic activity has, to date, been observed in bacteriophage encoded DNA mimetic proteins. These are Ugi, and p56 [15-19], as well as Sa-Ugi (a horizontally transferred phage-derived structural relative of Ugi, residing in a transposed pathogenicity island of a bacterial genome) [20]. The structure of Vpr in complex with UNG2 would strongly suggest inhibition of nuclear-localised UNG2 activity is also a strategy of the HIV-1 virus [21].

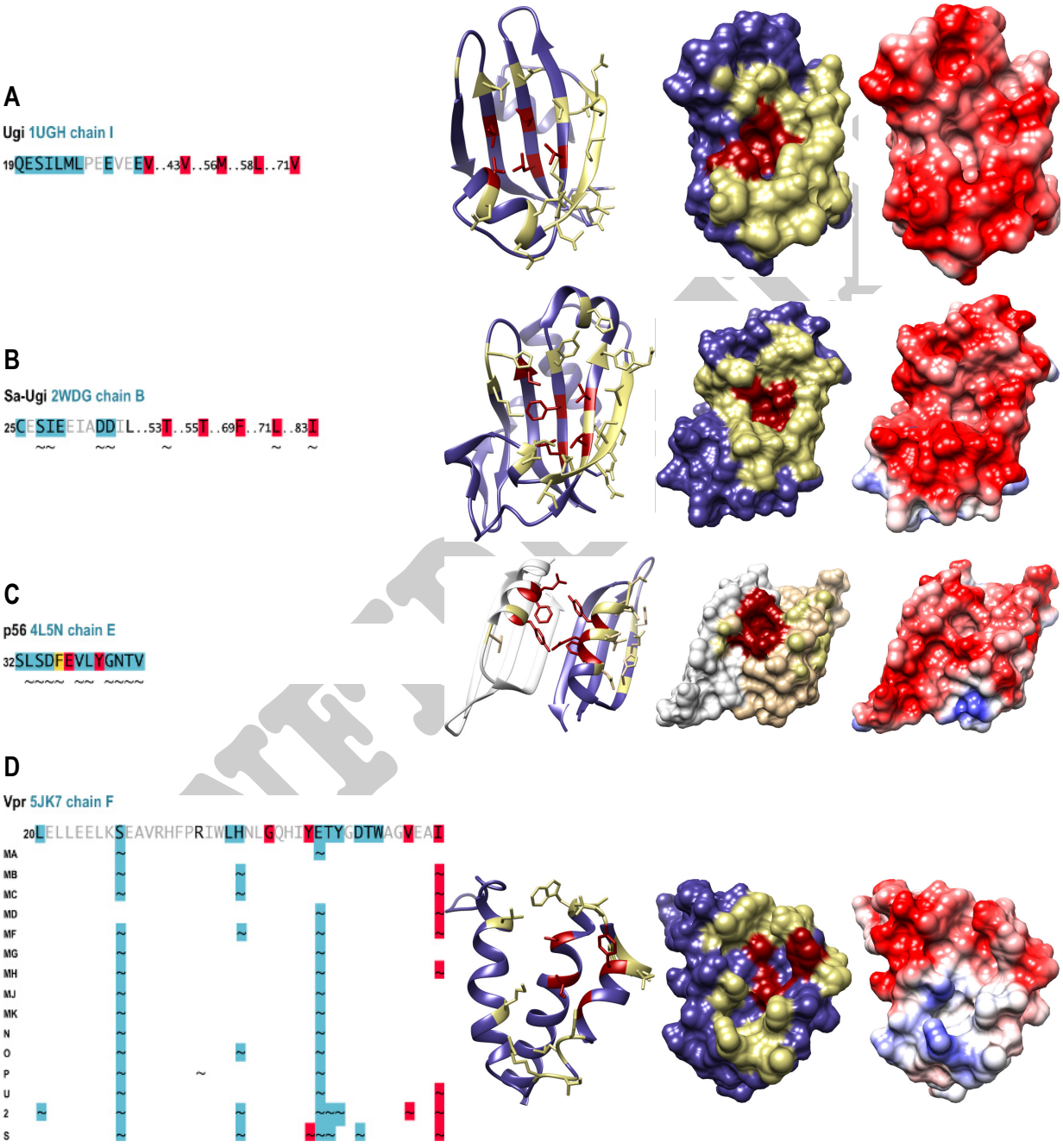

Figure 5 Evolutionarily disparate protein structures utilise an analogous structural mechanism to sequester and inactivate Ung, through convergence upon spatially precise residue contacts. Within each structural class of Ung inhibitor, interacting residues are subject to substitution in different strains.

Sequence alignment: Left column. DNA mimetic residue contacts are made by the motifs shown on a sky-blue background; grey font indicates those positions not in contact with Ung in deposited structures. Residues indicated "~" can be substituted in related sequences. The Ung minor groove DNA binding loop leucine [or its equivalent] is sequestered by residues at positions with a red background.

A For Ugi, in the top row, there are no variant sequences yet known. **B** In the next row, there are several Sa-Ugi sequence variants (indicated "~") from sequenced Staphylococcus genomes: all of which, it is presumed, are able to inhibit Ung. **C** In the p56 row, there is a residue shown on a gold background, forming the base of the leucine sequestration pocket, which can vary in related sequences; tolerable variation is probably explained by the fact this residue does not make contact with the leucine. Interestingly, there are p56 sequences in which every residue other than those directly interacting with the Ung leucine, vary (indicated "~") with respect to the known structures; there is probable structural equivalence in spite of such sequence plasticity, which will be of relevance in drug design targeting the Ung DNA binding cleft. **D** In the Vpr row, it is clear that key contacts may tolerate residue changes (indicated "~") which are observed in sequences throughout the major HIV-1 groups (M – with subtypes indicated by the second letter, N, O, P, and U) and close relatives HIV-2 (labeled "2") and SIV (labeled "S"). Grey font indicates residues that are not in contact with Ung in 5JK7; it is noted that such residues are relatively more prone to substitution in Vpr sequences from included groups, for the purposes of clarity this variability is not indicated by symbols in this view. Sequence data for Vpr were accessed from: HIV sequence database, at https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html

Structure depictions, rendered in Chimera, illustrate analogy in residue arrangements for Ung-complexation in spite of architectural diversity. The keynote is versatility in generating effective Ung inhibition, which alludes to the possibility of development of specific novel chemical entities to target specific Ungs in diverse relevant pathologies. Cartoons coloured as follows: Gold for residues involved in DNA mimetic contacts with Ung, red for hydrophobic residues that sequester the apical (predominantly, leucine) residue on the Ung pre-catalytic loop. In p56 the DNA mimetic monomer subunit in this dimeric structure is coloured in dark blue as per the monomeric proteins in the other rows, while the partner monomer, primarily involved in completing the leucine sequestering hydrophobic cavity makes relatively few contacts to Ung, and is coloured light grey. In the leftmost column of structure triptychs in each row, is a cartoon to show side chains involved in contacting Ung. The same view in the middle column of each triptych shows a view of Ung-contacting surfaces. In the rightmost column of the structure triptych is the same surface view, with coulombic potential indicating the charge alignment propensity of these markedly acidic proteins, to the appreciably basic Ung DNA binding cleft.

Those viral genomic protein inhibitors of Ung are essential to viral replication. Vpr variants are known in the major groups and sub-types of HIV-1 and also related retroviruses (Figure 5). The NMR structure of the apo form of Vpr [109], in common with initially reported NMR structures of Ugi and p56 [110,111] did not permit detailed modelling of the PPI interface with Ung. The interaction of Vpr with UNG2 has been variously explored (Vpr is comprehensively reviewed in [98]), implicating the essential contribution to complexation with UNG2, of the Vpr tryptophan 54 residue [100]. The crystal structure of the Vpr complex with UNG2 and DDB1–DCAF-1, revealed striking mechanistic analogy in Ung sequestration to the previously described phage-derived proteins [21], and Vpr tryptophan 54 indeed makes a key contact to UNG2.

The converged Ung inhibitory mechanism deployed by these evolutionarily disparate virus-encoded proteins, has the following common features: The known Ung inhibitor proteins are typically less than 16kD and markedly acidic in character [16-21,108]. Ung inhibiting proteins utilise (1) a charge-based alignment to orient on the Ung DNA binding cleft, then (2) Ung-specific DNA mimicry, docking to the Ung DNA binding cleft, and finally (3) hydrophobic sequestration of an Ung residue essential for pre-catalytic dsDNA substrate stabilisation [16-21]. This final capture step effects stoichiometric inhibition of Ung, irreversible under normal cellular conditions [19]. Despite a lack of sequence and structural homology between the Ung inhibitory protein classes (Figure 5), the mechanism operates equivalently and affects a wide range of known Ungs when tested in vitro [16-22,36,66,67,72,73,88]. Reassuringly for the purposes of emulating this mechanism for drug discovery, the various viral Ung inhibitory proteins interact differently in precise terms with their biological targets and furthermore in an Ung-dependent manner. An array of deposited biological structures are known, involving different viral Ung inhibitor proteins in complex with both UNG2 or a variety of pathogen Ungs (Figure 1, Figure 6). Thus, detailed insight for design, and increased scope for selectivity, should be achievable should it prove possible to emulate aspects of this Ung-inhibitory mechanism in the guise of novel chemical entities.

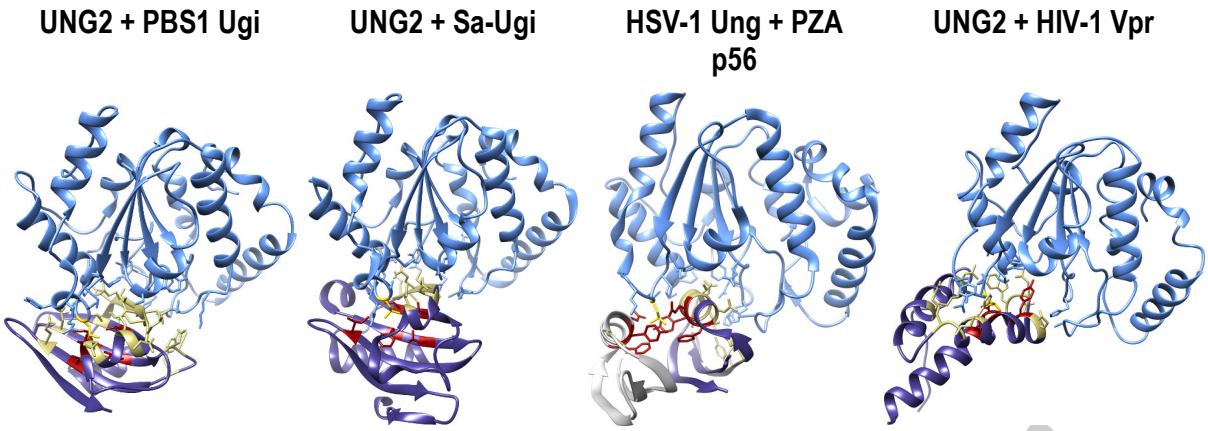

Figure 6 Analogy in mechanism to sequester and inactivate Ung. Convergence upon equivalent functional protein chemistry in the unrelated protein architectures evolved in discrete viral genomes. Within each class of Ung inhibitor, specific contacts to Ung differ, but retain the functional DNA-mimetic mechanism for Ung sequestration and nullification of activity.

Structure cartoons (1UGH chains E-I, 1AYR chains A-B, 4L5N chains B-E-F, 5JK7 chains D-F), rendered in Chimera, illustrate occupation of the Ung (light blue) DNA binding cleft (light blue sticks), and sequestration of the pre-catalytic Ung leucine (gold sticks) in spite of architectural diversity between the inhibitor proteins (dark blue). Effective Ung inhibition via inhibitor hydrophobic residues (red sticks), involves DNA mimicry (pale yellow sticks). One structurally identical monomer subunit of p56 (light grey), although essential for Ung inhibition, plays little part in DNA mimicry. That role is served by the Ung-docked DNA mimetic (dark blue) monomer subunit; the dimeric interface effects Ung inhibition by p56.

DNA mimicry involves contacts to the Ung DNA binding cleft using inhibitor protein amino acid sidechain chemistry to emulate parts of the bonding pattern encountered by Ung when resident upon dsDNA [16-21,88]. Ugi and Sa-Ugi utilise the edge of a beta strand to convey DNA mimicry, p56 utilises the alpha-helix of one of its monomer subunits, and Vpr employs a variable loop between two helices. All proteins contain a localised depression on the surface, lined with hydrophobic residues to trap the Ung pre-catalytic leucine (a phenyalanine in *Bacilli*) residue (Figure 5, Figure 6). The modified loop in poxvirus Ung cannot be sequestered, and it is untested as to whether the positively charged arginine or lysine residues found as substitutes for leucine only in β-herpesvirus Ungs could be captured. Certain similarities in charge distribution, and protrusion of contacting residues, are observable when the Ung-facing surfaces of the structurally described proteins are compared (in Figure 5, see the rightmost columns depicting surface views). This represents a remarkable level of analogous, spatially precise protein chemistry by functional convergence originating from radically different protein architectures [16-21].

Herpesvirus	Motif 6 Sequence Region
α HHV-1	SHPSPL-------SKV-P
β HHV-5	CHPSPR-----ntTR--A
β HHV-6A	AHPSPK----vksAR-MP
β HHV-7	AHPSPR----tkgSKT-P
γ HHV-4	QHPSPLaqnstrkSAQQK
γ HHV-8	QHPSPLaslggrhSRWPR
	****

Figure 7 Variant Ungs of herpesviruses feature differences in the pre-catalytic loop region, of potential importance to drug design.

Sequence alignment: The structure-based sequence alignment spacings (against 1SSP as the reference) originally shown in Figure 2, are retained for the alpha- [HHV-1], and gamma- [HHV-4, and HHV-8] herpesvirus sequences in this alignment of the pre-catalytic loop motif (i.e. referred to in Figure 2, as Motif 6). To compare with the beta-herpesviruses [HHV-5, HHV-6A, and HHV 7], for which there is currently no representative structure deposited in a public database, a manual alignment of the same pre-catalytic loop region (i.e. Motif 6 in Figure 2) is added. Notably, aside from some elaboration involving the loop length and residue composition (cyan coloured lower case text), there is a charged residue in the place where a hydrophobic residue is normally found in canonical Ung sequences (this residue is shown in Clustal Omega default colouring). This substituted position has obvious implications for building specificity into inhibitor design.

How then, might new inhibitors of Ung be designed? Targeting Ung for chemical inhibition is not a new endeavour [10-15,28,38], and atomic resolution insights for Ungs are plentiful (the PDB at the time of writing returns 99 structures with the search term “3.2.2.27”, and 199 structures with the search term “ung”). These deposited structures include evolutionarily diverse Ung domains in the apo form, and in

complex with natural and synthetic ligands as well as with the known viral inhibitor proteins: Ugi, p56, Sa-Ugi, and Vpr. Perusing just those structures alluded to in this article, reveals differences between Ungs in the details of ligand-bound forms that could be exploited in drug discovery. To summarise these very broadly: (1) A tightly bound blockade of any suitable site in an Ung DNA binding cleft could prevent association with DNA, and thus would have a silencing effect on Ung activity. (2) Alternatively, blockade of PPI interfaces on Ung, or its viral proteome ligands, would lead to weakened viral processes related to replicative fitness.

5 Synthetic selective Ung inhibition and future trends

It is worth recalling at this point, that classic drug design of the early 1990s developed a uracil-analogue compound with a hydrophobic tail that occupied the uracil-binding pocket in the Ung catalytic site. This was particularly aimed at herpesvirus encoded Ung, specifically from the α-herpesvirus HSV-1 [HHV-1]. These efforts have included multiple attempts at development of series of synthetic inhibitor compounds [10-15]. The approach has been one of targeting the exquisite affinity of Ung for uracil, and building in hydrophobic character, which according to modelling studies, may provide an analogous effect to the natural virally encoded polypeptide inhibitors of Ung [14]. The issue with the inhibitors described is not in their relative selectivity for the viral Ung over UNG2, which was found to be suitably distinct (reported to be at 8μM vs 300μM in the original OctAU inhibitor molecule) [10,11]; the inherent hydrophobic character of the molecules is more the problem to overcome as a key issue in compound development.

Subsequently developed compounds have similar properties [12-15], with those latter studies indicating the hydrophobic portion of the compounds enters a site interior to the enzyme through the catalytic pocket, rather than akin to the viral proteins' targeting of the pre-catalytic leucine residue. Such compound series therefore contrast with the natural inhibitors encoded by viruses, which do not occupy the uracil-binding pocket. Avoiding uracil analogues would be advantageous, since these would likely cause toxicity issues associated with DNA intercalation propensity; additionally, hydrophobic tails do not suggest the most favourable drug like properties.

Thus, the Ung pre-catalytic loop, and the DNA binding cleft area exterior to the uracil-specific catalytic site, would appear to be the key inhibitory weak point of Ung from an evolutionary antagonistic vantage point. It is natural then to suppose that variant Ungs, differing as they do in this key respect of Ung structure and chemistry conservation (i.e. Motif 6, in Figure 2), would represent start points for inhibitor design. To reiterate, the structurally and chemically variant analogous loop in vaccinia Ung is one reason why Ugi has no inhibitory action against that particular Ung variant. However, from a specific drug design perspective that feature alone would represent a unique targeting opportunity. In this case it would represent an alternate approach from targeting the PPI with vaccinia polymerase E9 [28]. The different pre-catalytic loop chemistry (arginine or lysine, rather than leucine) found on β-herpesvirus Ungs would suggest yet another unique feature to target on a pathogen variant Ung, relative to UNG2.

Transposing natural Ung-targeting via modern drug design for key therapeutic areas

It is hoped that three things will now remain in mind when considering design and synthesis of healthcare-transforming Ung inhibitors:

1) The first is that Ung is an actual or potential Achilles heel to be targeted in at least several pathogens (Table 1) responsible for current worldwide critical healthcare burdens; including those mentioned in this review, namely: Tuberculosis, MRSA, malaria, herpesviruses, and HIV-1, noting also that viral Ung would make an excellent addition to compounds targeting poxviruses, and furthermore that UNG2 in certain pathological contexts (its PPIs in viral diseases, and its activity in cancer) is also a therapeutic target.

2) The second is that Ung selectivity has been demonstrated via classic drug discovery approaches from the 1990s, and with the wealth of structural information now in our hands also via structure-led fragment-based drug design methods. The emphasis to date has always been on starting with uracil or its analogues [10-15].

3) The third consideration is the importance of Ung in pathogenically relevant PPIs, and the convergent DNA-mimetic mechanism of natural virally encoded Ung inhibitor proteins, thus uracil is not the only way forward. Concepts of drug design that target PPIs, have been shown to be viable (reviewed and discussed in e.g. [112]). Furthermore, new approaches to the targeting of DNA-binding proteins specifically, could in principle (1) close the Ung catalytic domain or (2) reprise elements of the DNA mimetic protein mechanism or interfering with the pre-catalytic loop, which is not the case in drug

designs explored to date [10-15]. In fact, the Ung DNA binding cleft in general could offer selective druggable sites that could be highly specific and tight binding. A summary and consideration of the components that would support this approach is presented here to conclude this article.

From protein mimics of Ung-bound DNA, to peptide mimics, to therapeutics?

Three unrelated protein architectures: Ugi/Sa-Ugi, p56, and Vpr, have zeroed in on a single universal mechanism for inhibition of Ung activity [16-21]. Even within each class of Ung inhibitor, sequences vary within the Ung-contacting motifs and yet retain the functional mechanism for Ung sequestration and nullification of activity (Figure 4, Figure 5, Figure 6). Elements of this evolutionarily road-tested sweet spot for Ung silencing remain to be exploited in the search for Ung-specific novel chemical entities, yet the potential therapeutic vistas to be thus opened appear tantalising.

Considering also, PPIs: disrupting them with novel chemical entities may progress from mapping out druggable patches at the contact interfaces using peptides [113]. This strategy in itself may result in a drug-like molecule, since peptidic drugs are successful in their own right (as reviewed in [114]). It is possible that a suitable peptide docking in the Ung DNA binding cleft may elicit closure of the Ung catalytic domain. Thus, a peptide therapeutic to act on Ung, can quickly be sketched out; but, the question is, how specific or general can such an imagined Ung inhibitor be? The answer is, probably more inter-species specific than an inhibitor based upon uracil. DNA binding clefts in diverse Ungs display ample sequence variation, observed as attributable conformation changes. Indeed, residue conformation in proteins is known to underlie sequence variation that nevertheless supports structure conservation [115], thus it stands to reason that the landscape for drug binding would be appreciably different in such homologues if the cleft were targeted rather than the pocket. To reiterate: uracil analogues need not be part of the equation at all, in principle.

The Ung DNA binding cleft need not be fully occupied to trigger domain closure, since closed conformations of Ung have been observed under different circumstances [44]. Thus, the drug we can envisage in development could be based upon a molecule arising from a library that samples the cleft area of Ung, and for which domain closure, or just tight binding, is the readout of interest. In fact, similar strategies in compounds based upon uracil binding have been shown to be useful approaches [13]. This could yet be a peptidic molecule, however new approaches in effecting DNA mimicry for protein interaction using aromatic oligoamides, might also offer an alternative [116].

The apical residue on the Ung pre-catalytic loop is almost always hydrophobic (almost always leucine; see Figure 2) but, in an imagined novel chemical entity emulating that residue's sequestration by virally encoded inhibitors, binding could instead target another type of chemistry relevant to that loop. In principle, such inhibitors could emanate from a scaffold that takes advantage of cleft differences between Ungs, heralding specificity for any selected Ung. Interestingly, although it has already been effectively targeted via its PPI with viral polymerase E9, the poxvirus Ung has a structurally diverged loop, among other gross differences relating to its essential role in viral replication [26]. The β-herpesvirus Ungs feature a charged apical pre-catalytic loop residue; γ-herpesvirus Ungs have a separate character to take advantage of in this pre-catalytic loop area, by virtue of a structured extension to that loop, necessary for effective lytic replication [33,34]. In terms of designing a drug, this has a slight complication in that the loop is disordered prior to DNA docking; nevertheless, the structural and mutagenesis data available provide important design clues [34].

Looking at the HIV-1 accessory protein Vpr, its interactions with UNG2 show a different bonding pattern in the Ung DNA binding cleft, supported by extra interactions on the surface around it [21]. Again, this could provide more clues as to how to build a novel chemical entity which effects uracil-free inhibition of Ung, as in effect this structural insight provides some pre-prepared peptide mapping [113]. There are PPI interfaces to define, other than Vpr, which could be sites for important new UNG2 targeting drug approaches against HIV. In particular, the details of a reported association with integrase [99] have not been described to date.

Conclusion

Ungs are key molecules in a number of important pathological states. Therefore, the time could be right to open potentially very useful avenues for novel therapies targeting Ungs, with new advances in drug design and delivery. It is suggested these could now enable NCE emulation of independently evolved viral Ung-targeting strategies via PPIs, including Ung inhibition.

Importantly this Ung-inhibitory mechanism does not involve the uracil-selectivity of Ung. There are currently no novel chemical entities that do not target uracil-binding, to exploit key features of the natural Ung-inhibitory mechanism, but the potential for increased selectivity by exploring this path seems highly attractive. There is relative variability in the cleft region of Ung involved in interaction with viral inhibitors, and variation too in the contact features of those inhibitors.

Furthermore, PPIs involving either UNG2 or virally encoded Ung, appear central to pathogenic virus viability and present another therapeutic horizon. The utility of targeting Ung PPIs has been demonstrated with poxvirus Ung: This approach could therefore be applied to more immediately relevant pathogenic viral states, such as those caused by human herpesviruses, or HIV-1.

Future Perspectives

It is envisaged that developments in AI-enhanced bioinformatics will provide greater power in uncovering currently hidden sequence and structure signatures. Such tools could discover proteins interacting with Ungs in ways that promote or support pathogenesis. The evolutionary insights provided by an expanded known universe of pathogen interactions via Ung, will assist development of potent novel chemical entities capable of disabling specific or multiple pathogen Ungs, or even UNG2 where more appropriate. The chemical state of the art will enable targeting of Ungs into specific cells (e.g. via antibody-drug conjugates) and intracellular locations, or to promote their uptake into pathogenic organisms.

Acknowledgements

RS is grateful to Dr Tracey Barrett for critical pre-reading of the manuscript.

References

[1] Lindahl T. An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl. Acad. Sci. U.S.A.*, 71(9), 3649-3653 (1974).

[2] Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens, B. DNA N-Glycosidases: Properties of uracil-DNA glycosidase from Escherichia coli. *J. Biol. Chem.*, 252(10), 3236-3294 (1977).

[3] Savva R, McAuley-Hecht K, Brown T, Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, 373(6514), 487-493 (1995).

[4] Parikh SS, Mol CD, Slupphaug G, Bharati S, Krokan HE, Tainer JA. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.*, 17(17), 5214-5226 (1998).

[5] Lindahl T. Instability and decay of the primary structure of DNA. *Nature*, 362(6422), 709-715 (1993).

[6] Tomita F, Takahashi I. DNase Specific for Uracil-Containing Bacteriophage DNA. *J. Virol.*, 15(5), 1081-1087 (1975).

[7] Weil AF, Ghosh D, Zhou Y *et al*. Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc. Natl. Acad. Sci. U.S.A.*, 110(6), E448-E457 (2013).

[8] Martomo SA, Gearhart PJ. Somatic hypermutation: subverted DNA repair. *Curr. Opin. Immunol.*, 18(3), 243-248 (2006).

[9] Yousif AS, Stanlie A, Mondal S, Honjo T, Begum NA. Differential regulation of S-region hypermutation and class-switch recombination by noncanonical functions of uracil DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.*, 111(11), E1016-1024 (2014).

[10] Argnani R, Focher F, Zucchini S *et al.* Herpes simplex virus type 1 (HSV-1) uracil-DNA glycosylase: functional expression in Escherichia coli, biochemical characterization, and selective inhibition by 6-(p-n-octylanilino)uracil. *Virology*, 211(1), 307-311 (1995).

[11] Sun H, Zhi C, Wright GE *et al.* Molecular Modeling and Synthesis of Inhibitors of Herpes Simplex Virus Type 1 Uracil-DNA Glycosylase. *J. Med. Chem.*, 42(13), 2344-2350 (1999).

[12] Jiang YL, Krosky DJ, Seiple L, Stivers JT. Uracil-directed ligand tethering: an efficient strategy for uracil DNA glycosylase (UNG) inhibitor development. *J. Am. Chem. Soc.*, 127(49), 17412-17420 (2005).

[13] Krosky DJ, Bianchet MA, Seiple L, Chung S, Amzel LM, Stivers JT. (2006). Mimicking damaged DNA with a small molecule inhibitor of human UNG2. *Nucleic Acids Res.*, 34(20), 5872–5879.

* Although the compounds in the published report are uracil-analogues, the general method for their discovery and triage could be made generally applicable to Ung targeting compounds such as those proposed in the present article.

[14] Hendricks U, Crous W, Naidoo KJ. Computational Rationale for the Selective Inhibition of the Herpes Simplex Virus Type 1 Uracil-DNA Glycosylase Enzyme. *J. Chem. Inf. Model.*, 54(12), 3362-3372 (2014).

[15] Jiang YL, Chung S, Krosky DJ, Stivers JT. Synthesis and high-throughput evaluation of triskelion uracil libraries for inhibition of human dUTPase and UNG2. *Bioorg. Med. Chem.*, 14(16), 5666-5672 (2006).

[16] Savva R, Pearl LH. Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nat. Struct. Biol.*, 2(9), 752-757 (1995).

[17] Mol CD, Arvai AS, Sanderson RJ *et al.* Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, 82(5), 701-708 (1995).

[18] Baños-Sanz JI, Mojardín L, Sanz-Aparicio J *et al.* Crystal structure and functional insights into uracil-DNA glycosylase inhibition by phage ϕ29 DNA mimic protein p56. *Nucleic Acids Res.*, 41(13), 6761-73 (2013).

[19] Cole AR, Ofer S, Ryzhenkova K, Baltulionis G, Hornyak P, Savva R. Architecturally diverse proteins converge on an analogous mechanism to inactivate Uracil-DNA glycosylase. *Nucleic Acids Res.*, 41(18), 8760-8775 (2013).

[20] Wang HC, Hsu KC, Yang JM *et al. Staphylococcus aureus* protein SAUGI acts as a uracil-DNA glycosylase inhibitor. *Nucleic Acids Res.*, 42(2), 1354-1364 (2014).

[21] Wu Y, Zhou X, Barnes CO *et al.* The DDB1–DCAF1–Vpr–UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nat. Struct. Mol. Biol.*, 23(10), 933-940 (2016).

** Vpr has multiple roles in HIV pathogenesis, but preventing its association with UNG2 would no doubt interfere with the uracil-DNA sensitivity of retroviral integration. This complex structure will offer drug design clues, akin to those used in discovering PPI blockers between D4 and E9 in poxviruses.

[22] Wang Z, Mosbaugh DW. Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.*, 264(2), 1163-1171 (1989).

[23] Pérez-Lago L, Serrano-Heras G, Baños, B *et al.* Characterization of Bacillus subtilis uracil-DNA glycosylase and its inhibition by phage ϕ29 protein p56. *Mol. Microbiol.*, 80(6), 1657-1666 (2011).

[24] Jamrozy D, Coll F, Mather AE *et al.* Evolution of mobile genetic element composition in an epidemic methicillin-resistant Staphylococcus aureus: temporal changes correlated with frequent loss and gain events. *BMC Genomics*, 18, 684 (2017).

* A functional Ung inhibitor is encoded in an ORF within a pathogenicity island linked to antibiotic escape. There are implications for antibiotic potentiation, and also other anti-microbial drugs to target a potential Ung-sensitive mechanism of hypermutation.

[25] Warner HR, Johnson LK, Snustad DP. Early Events After Infection of *Escherichia coli* by Bacteriophage T5. III. Inhibition of Uracil-DNA Glycosylase Activity. *J. Virol.*, 33(1), 535-538 (1980).

[26] Schormann N, Zhukovskaya N, Bedwell G *et al.* Poxvirus uracil-DNA glycosylase-An unusual member of the family I uracil-DNA glycosylases. *Protein Sci.*, 25(12), 2113-2131 (2016).

[27] Czarneckia MW, Traktman P. The vaccinia virus DNA polymerase and its processivity factor. *Virus Res.*, 234, 193–206 (2017).

[28] Nuth M, Guan H, Zhukovskaya N, Saw YL, Ricciardi RP. Design of potent poxvirus inhibitors of the heterodimeric processivity factor required for viral replication. *J. Med. Chem.*, 56(8), 3235–3246 (2013).

** Blocking a PPI involving an Ung can prevent viral replication and plaque formation. This alludes to the fact that far less toxic and far more drug like compounds can be used to target PPIs involving Ung, and uracil analogues do not need to be part of the equation.

[29] Schormann N, Ricciardi R, Chattopadhyay D. Uracil-DNA glycosylases – structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci., 23(12), 1667-1685 (2014).

[30] Sang PB, Srinath T, Patil AG, Woo EJ, Varshney U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.*, 43(17), 8452-8463 (2015).

[31] Li J, Yang Y, Guevara J, Wang L, Cao W. Identification of a prototypical single-stranded uracil DNA glycosylase from Listeria innocua. *DNA Repair (Amst).*, 57, 107-115 (2017).

[32] Kwon E, Pathak D, Chang HW, Kim DY. Crystal structure of mimivirus uracil-DNA glycosylase. *PLoS ONE*, 12(8), e0182382 (2017).

[33] Li J, Chen R, Yang, Y *et al.* An unconventional family 1 uracil DNA glycosylase in Nitratifractor salsuginis. *FEBS J.*, 284(23), 4017-4034 (2017).

[34] Earl C, Bagnéris C, Zeman K, Cole A, Barrett T, Savva R. A structurally conserved motif in γ-herpesvirus uracil-DNA glycosylases elicits duplex nucleotide-flipping. *Nucleic Acids Res.*, 46(8), 4286-4300 (2018).

** The viral Ung variant in γ-herpesviruses uses its modified architecture to signal in AT-rich DNA sequences akin to the viral origin. This structure and mutagenesis study offers drug design insights.

[35] Geoui T, Buisson M, Tarbouriech N, Burmeister WP. New insights on the role of the gamma-herpesvirus uracil-DNA glycosylase leucine loop revealed by the structure of the Epstein-Barr virus enzyme in complex with an inhibitor protein. *J. Mol. Biol.*, 366(1), 117-131 (2007).

[36] Kaushal PS, Talawar RK, Krishna PDV, Varshney U, Vijayan M. Unique features of the structure and interactions of mycobacterial uracil-DNA glycosylase: structure of a complex of the Mycobacterium tuberculosis enzyme in comparison with those from other sources. *Acta Crystallogr. D Biol. Crystallogr.*, 64(pt 5), 551-560 (2008).

[37] Kaushal PS, Talawar RK, Varshney U, Vijayan M. Structure of uracil-DNA glycosylase from Mycobacterium tuberculosis: insights into interactions with ligands. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.*, 66(pt 8), 887-892 (2010).

** The structures reveal that the Ung domain can be open or closed and this may be ligand related. This is of relevance not only this specific pathogen Ung, but to drug design criteria for Ungs in general.

[38] Suksangpleng T, Leartsakulpanich U, Moonsom S *et al.* Molecular characterization of Plasmodium falciparum uracil-DNA glycosylase and its potential as a new anti-malarial drug target. *Malar. J.*, 13, 149 (2014).

[39] Stuart DT, Upton C, Higman MA, Niles EG, McFadden G. A poxvirus-encoded uracil DNA glycosylase is essential for virus viability. *J. Virol.*, 67(5), 2503-2512 (1993).

[40] Varshney U, van de Sande JH. Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by Escherichia coli uracil DNA glycosylase. *Biochemistry*, 30(16), 4055-61 (1991).

[41] Dinner AR, Blackburn GM, Karplus M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature*, 413(6857), 752-755 (2001).

[42] Friedman JI, Majumdar A, Stivers JT. Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res.*, 37(11), 3493-3500 (2009).

[43] Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, 449(7161), 433-437 (2007).

[44] Arif SM, Geethanandan K, Mishra P, Surolia A, Varshney U, Vijayan M. Structural plasticity in Mycobacterium tuberculosis uracil-DNA glycosylase (MtUng) and its functional implications. *Acta Crystallogr. D Biol. Crystallogr.*, 71(pt 7), 1514-1527 (2015).

[45] Imahashi M, Nakashima M, Iwatani Y. Antiviral mechanism and biochemical basis of the human APOBEC3 family. *Front. Microbiol.*, 3, 250 (2012).

[46] Yang B, Chen K, Zhang C, Huang S, Zhang H. Virion-associated Uracil DNA Glycosylase-2 and Apurinic/ Apyrimidinic Endonuclease Are Involved in the Degradation of APOBEC3G-edited Nascent HIV-1 DNA. *J. Biol. Chem.*, 282(16), 11667-11675 (2007).

[47] Chen J, Furano AV. Breaking bad: The mutagenic effect of DNA repair. *DNA Repair (Amst).*, 32, 43-51 (2015).

[48] Yan N, O'Day E, Wheeler LA, Engelman A, Lieberman J. HIV DNA is heavily uracilated, which protects it from autointegration. *Proc. Natl. Acad. Sci. U.S.A.*, 108(22), 9244-9249 (2011).

[49] Eldin P, Chazal N, Fenard D, Bernard E, Guichou JF, Briant L. Vpr expression abolishes the capacity of HIV-1 infected cells to repair uracilated DNA. *Nucleic Acids Res.*, 42(3), 1698-710 (2013).

[50] Hansen EC, Ransom M, Hesselberth JR *et al*. Diverse fates of uracilated HIV-1 DNA during infection of myeloid lineage cells. *Elife*. 5, e18447 (2016).

[51] Bouhamdan M, Benichou S, Rey F *et al.* Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *J. Virol.*, 70(2), 697-704 (1996).

[52] Chen R, Le Rouzic E, Kearney JA, Mansky LM, Benichou S. Vpr-mediated incorporation of UNG2 into HIV-1 particles is required to modulate the virus mutation rate and for replication in macrophages. *J. Biol. Chem.*, 279(27), 28419-28425 (2004).

[53] Schröfelbauer B, Yu Q, Zeitlin SG, Landau NR. Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. *J. Virol.*, 79(17), 10978-10987 (2005).

[54] Langevin C, Maidou-Peindara P, Aas PA *et al.* Human immunodeficiency virus type 1 Vpr modulates cellular expression of UNG2 via a negative transcriptional effect. *J. Virol.*, 83(19), 10256-10263 (2009).

[55] Ahn J, Vu T, Novince Z, Guerrero-Santoro J, Rapic-Otrin V, Gronenborn AM. HIV-1 Vpr Loads Uracil DNA Glycosylase-2 onto DCAF1, a Substrate Recognition Subunit of a Cullin 4A-RING E3 Ubiquitin Ligase for Proteasome-Dependent Degradation. *J. Biol. Chem.*, 285(48), 37333-37341 (2010).

[56] Guenzel CA, Hérate C, Le Rouzic, E *et al.* Recruitment of the nuclear form of uracil DNA glycosylase into virus particles participates in the full infectivity of HIV-1. *J. Virol.*, 86(5), 2533-2544 (2012).

[57] Hérate C, Vigne C, Guenzel CA, Lambele M, Rouyez MC, Benichou S. Uracil DNA glycosylase interacts with the p32 subunit of the replication protein A complex to modulate HIV-1 reverse transcription for optimal virus dissemination. *Retrovirology*, 13, 26 (2016).

[58] Ara A, Love RP, Follack TB, Ahmed KA, Adolph MB, Chelico L. Mechanism of Enhanced HIV Restriction by Virion Coencapsidated Cytidine Deaminases APOBEC3F and APOBEC3G. *J. Virol.*, 91(3) pii: e02230-16 (2017).

[59] Courcelle CT, Courcelle J, Prichard MN, Mocarski ES. Requirement for Uracil-DNA Glycosylase during the Transition to Late-Phase Cytomegalovirus DNA Replication. *J. Virol.*, 75(16), 7592-7601 (2001).

[60] Su M-T, Liu I-H, Wu C-W *et al.* Uracil DNA Glycosylase BKRF3 Contributes to Epstein-Barr Virus DNA Replication through Physical Interactions with Proteins in Viral DNA Replication Complex. *J. Virol.*, 88(16), 8883-8899 (2014).

[61] Pyles RB, Thompson RL. Evidence that the herpes simplex virus type 1 uracil DNA glycosylase is required for efficient viral replication and latency in the murine nervous system. *J. Virol.*, 68(8), 4963-4972 (1994).

[62] Verma SC, Bajaj BG, Cai Q, Si H, Seelhammer T, Robertson ES. Latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus recruits uracil DNA glycosylase 2 at the terminal repeats and is important for latent persistence of the virus. *J. Virol.*, 80(22), 11178-11190 (2006).

[63] Takahashi I, Marmur J. Replacement of Thymidylic Acid by Deoxyuridylic Acid in the Deoxyribonucleic Acid of a Transducing Phage for Bacillus subtilis. *Nature*, 197(4869), 794−795 (1963).

[64] Kiljunen S, Hakala K, Pinta E *et al.* Yersiniophage ϕR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology*, 151(Pt 12), 4093-4102 (2005).

[65] Lavysh D, Sokolova M, Minakhin L *et al.* The genome of AR9, a giant transducing Bacillus phage encoding two multisubunit RNA polymerases. *Virology*, 495, 185-196 (2016).

[66] Katz GE, Price AR, Pomerantz MJ. Bacteriophage PBS2-induced inhibition of uracil-containing DNA degradation. *J. Virol.*, 20(2), 535-538 (1976).

[67] Serrano-Heras G, Bravo A, Salas M. Phage φ29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.*, 105(9), 19044-19049 (2008).

[68] Glukhov AS, Krutilina AI, Shlyapnikov MG *et al.* Genomic Analysis of Pseudomonas putida Phage tf with Localized Single-Strand DNA Interruptions. *PLoS ONE*, 7(12), e51163 (2012).

[69] Yamagishi H. Single Strand Interruptions in PBS 1 Bacteriophage DNA Molecule. *J. Mol. Biol.*, 35(3), 623-633 (1968).

[70] Rhoades M. Localization of single-chain interruptions in bacteriophage T5 DNA. II. Electrophoretic studies. *J. Virol.*, 23(3), 737-50 (1977).

[71] Tye BK, Nyman PO, Lehman IR, Hochhauser S, Weiss B. Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 74(1), 154-157 (1977).

[72] Friedberg EC, Ganesan AK, Minton K. N-Glycosidase activity in extracts of Bacillus subtilis and its inhibition after infection with bacteriophage PBS2. *J. Virol.*, 16(2), 315-321 (1975).

[73] Tomita F, Takahashi I. Changes in DNase Activities in Bacillus subtilis Infected with Bacteriophage PBS 1. *J. Virol.*, 15(5), 1073-1080 (1975).

[74] Vieira G, de Lencastre H, Archer L. Restriction analysis of PBS1-related phages. *Arch. Virol.*, 106(1-2),121-126 (1989).

[75] Ivarie R. Thymine methyls and DNA-protein interactions. *Nucleic Acids Res.* , 15(23), 9975-83 (1987).

[76] Chamberlin M, Ring J. Characterization of T7-specific Ribonucleic Acid Polymerase: I. General properties of the enzymatic reaction and the template specificity of the enzyme. *J. Biol. Chem.*, 248(6), 2235-2211 (1973).

[77] Price AR, Cook SJ. New Deoxyribonucleic Acid Polymerase Induced by Bacillus *subtilis* Bacteriophage PBS2. *J. Virol.*, 9(4), 602-610 (1972).

[78] Kuo TT, Huang TC, Teng MH. 5-Methylcytosine replacing cytosine in the deoxyribonucleic acid of a bacteriophage for *Xanthomonas oryzae*. *J. Mol. Biol.*, 34(2), 373-5 (1968).

[79] Warren RAJ. Modified Bases In Bacteriophage DNAs. *Ann. Rev. Microbiol.*, 34, 137-58 (1980).

[80] Lee Y-J, Dai N, Walsh SE *et al.* Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. *Proc. Natl. Acad. Sci. U.S.A.*, 115(14), E3116–E3125 (2018).

[81] Shen JC, Rideout WM 3rd, Jones PA. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.*, 22(6), 972-6 (1994).

[82] Mahler I, Cahoon M, Marmur J. *Bacillus Subtilis* Deoxyribonucleic Acid Transfer In PBS2 Transduction. *J. Bacteriol.*, 87(Jun), 1423-1428 (1964).

[83] Lovett PS, Young FE. Genetic Analysis in *Bacillus pumilus* by PBS1-Mediated Transduction. *J. Bacteriol.*, 101(2), 603-608 (1970).

[84] Uchiyama J, Takemura-Uchiyama I, Sakaguchi Y *et al.* Intragenus generalized transduction in Staphylococcus spp. by a novel giant phage. *The ISME journal*, 8(9), 1949-52 (2014).

[85] Belyaeva NN, Azizbekyan RR. Fine structure of new *Bacillus subtilis* phage AR9 with complex morphology. *Virology*, 34(1), 176-9 (1968).

[86] Maisch WF, Wachsman JT. Properties of thymineless strains of *Bacillus megaterium*. *Appl. Microbiol.*, 24(5), 717-20 (1972).

[87] Keen EC, Bliskovsky VV, Malagon F *et al.* Novel "Superspreader" Bacteriophages Promote Horizontal Gene Transfer by Transformation. *mBio*, 8(1), pii: e02115-16 (2017).

[88] Wang HC, Ho CH, Chou CC *et al.* Using structural-based protein engineering to modulate the differential inhibition effects of SAUGI on human and HSV uracil DNA glycosylase. *Nucleic Acids Res.*, 44(9), 4440-4449 (2016).

[89] Martin SA, Lord CJ, Ashworth A. DNA repair deficiency as a therapeutic target in cancer. *Curr. Opin. Genet. Dev.*, 18(1), 80-86 (2008).

[90] Chen R, Wang H, Mansky LM. Roles of uracil-DNA glycosylase and dUTPase in virus replication. *J. Gen. Virol.*, 83(Pt 10), 2339-2345 (2002).

[91] Qin L, Favis N, Famulski J, Evans DH. Evolution of and Evolutionary Relationships between Extant Vaccinia Virus Strains. *J. Virol.*, 89(3), 1809-1824 (2015).

[92] Lu CC, Huang HT, Wang JT *et al.* Characterization of the uracil-DNA glycosylase activity of Epstein-Barr virus BKRF3 and its role in lytic viral DNA replication. *J. Virol.*, 81(3), 1195–1208 (2007).

[93] Xu Y, Rodriguez-Huete A, Pari GS. Evaluation of the Lytic Origins of Replication of Kaposi's Sarcoma-Associated Virus/Human Herpesvirus 8 in the Context of the Viral Genome. *J. Virol.*, 80(19), 9905-9909 (2006).

[94] Sattler C, Steer B, Adler H. Multiple Lytic Origins of Replication Are Required for Optimal Gammaherpesvirus Fitness In Vitro and In Vivo. *PLoS Pathog.*, 12(3), e1005510 (2016).

[95] Itzhaki RF. Corroboration of a Major Role for Herpes Simplex Virus Type 1 in Alzheimer's Disease. *Front. Aging Neurosci.*, 10, 324 (2018).

[96] Priet S, Sire J, Quérat G. Uracils as a cellular weapon against viruses and mechanisms of viral escape. *Curr. HIV Res.*, 4(1), 31-42 (2006).

[97] Kogan M, Rappaport J. HIV-1 Accessory Protein Vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology*, 8, 25 (2011).

[98] Guenzel CA, Hérate C, Benichou S. HIV-1 Vpr – a still "enigmatic multitasker". *Front. Microbiol.*, 5, 127 (2014).

[99] Willetts KE, Rey F, Agostini I *et al.* DNA repair enzyme uracil DNA glycosylase is specifically incorporated into human immunodeficiency virus type 1 viral particles through a Vpr-independent mechanism. *J. Virol.*, 73(2), 1682-1688 (1999).

[100] Mansky LM, Preveral S, Selig L, Benarous R, Benichou S. The Interaction of Vpr with Uracil DNA Glycosylase Modulates the Human Immunodeficiency Virus Type 1 In Vivo Mutation Rate. *J. Virol.*, 74(15), 7039-47 (2000).

[101] Win NN, Ngwe H, Abe I, Morita H. Naturally occurring Vpr inhibitors from medicinal plants of Myanmar. *J. Nat. Med.*, 71(4), 579-589 (2017).

[102] González ME. The HIV-1 Vpr Protein: A Multifaceted Target for Therapeutic Intervention. *Int. J. Mol. Sci.*, 18(1), pii: E126 (2017).

[103] Rossenkhan R, MacLeod IJ, Brumme ZL *et al.* Transmitted/Founder HIV-1 Subtype C Viruses Show Distinctive Signature Patterns in Vif, Vpr, and Vpu That Are Under Subsequent Immune Pressure During Early Infection. *AIDS Res Hum Retroviruses*, 32(10-11), 1031-1045 (2016).

[104] Yan N, Cherepanov P, Daigle JE, Engelman A, Lieberman J. The SET complex acts as a barrier to autointegration of HIV-1. *PLoS Pathog.*, 5(3), e1000327 (2009).

[105] Rustagi A, Gale M Jr. Innate antiviral immune signaling, viral evasion and modulation by HIV-1. *J. Mol. Biol.*, 426(6), 1161-1177 (2014).

[106] Sumner RP, Thorne LG, Fink DL, Khan H, Milne RS, Towers GJ. Are Evolution and the Intracellular Innate Immune System Key Determinants in HIV Transmission? *Front. Immunol.*, 8, 1246 (2017).

[107] Le Rouzic E, Benichou S. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology*, 2, 11 (2005).

[108] Feng Y, Love RP, Chelico L. HIV-1 Viral Infectivity Factor (Vif) Alters Processive Single-stranded DNA Scanning of the Retroviral Restriction Factor APOBEC3G. *J. Biol. Chem.*, 288(9), 6083-6094 (2013).

[109] Morellet N, Bouaziz S, Petitjean P, Roques BP. NMR structure of the HIV-1 regulatory protein VPR. *J. Mol. Biol.*, 327(1), 215-27 (2003).

[110] Beger RD, Balasubramanian S, Bennett SE, Mosbaugh DW, Bolton PH. Tertiary structure of uracil-DNA glycosylase inhibitor protein. *J. Biol. Chem.*, 270(28), 16840-7 (1995).

[111] Asensio JL, Pérez-Lago L, Lázaro JM, González C, Serrano-Heras G, Salas M. Novel dimeric structure of phageϕ29-encoded protein p56: insights into uracil-DNA glycosylase inhibition. *Nucleic Acids Res.*, 39(22), 9779-88 (2011).

[112] Cierpicki T, Grembecka J. Targeting protein–protein interactions in hematologic malignancies: still a challenge or a great opportunity for future therapies? *Immunol. Rev.*, 263(1), 279-301 (2015).

[113] Cardote TAF, Ciulli A. Structure-Guided Design of Peptides as Tools to Probe the Protein-Protein Interaction between Cullin-2 and Elongin BC Substrate Adaptor in Cullin RING E3 Ubiquitin Ligases. *ChemMedChem*, 12(18), 1491-1496 (2017).

[114] Nevola L, Giralt E. Modulating protein-protein interactions: the potential of peptides. *Chem. Commun. (Camb).*, 51(16), 3302-3315 (2015).

* PPIs are druggable, and there are examples highlighted here.

[115] Juritz E, Palopoli N, Fornasari MS, Fernandez-Alberti S, Parisi G. Protein conformational diversity modulates sequence divergence. *Mol. Biol. Evol.*, 30(1), 79-87 (2013).

[116] Ziach K, Chollet C, Parissi V *et al.* Single helically folded aromatic oligoamides that mimic the charge surface of double-stranded B-DNA. *Nat. Chem.*, 10(5), 511-518 (2018).

** This kind of novel research could be applicable to design of Ung DNA-binding cleft inhibitors.