

Univerzita Karlova
Přírodovědecká fakulta

Studijní program:
Molekulární biologie a biochemie organismů



Tereza Uhlíková

Nekódující RNA a jejich využití v genovém inženýrství
Non-coding RNAs and their use in genetic engineering

Bakalářská práce

Školitel: Mgr. Adéla Přibyllová, Ph.D.

Konzultant: RNDr. Lukáš Fischer, Ph.D.

Praha, 2025

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně, a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne 27.4.2025

Podpis:

Poděkování:

Ráda bych poděkovala zejména své školitelce Mgr. Adéle Příbylové, Ph.D. za rady, motivaci a konstruktivní připomínky při zpracování bakalářské práce a také konzultantovi RNDr. Lukáši Fischerovi, Ph.D. Dále bych chtěla poděkovat všem, kteří mě při psaní této práce podporovali, zejména pak mým rodičům.

Abstrakt

Nekódující RNA stojí v centru navádění nejmodernějších nástrojů používaných v molekulární biologii například k editaci genů, regulaci genové exprese či při genové terapii. Nejnovější z nich jsou tzv. TIGR-Tas systémy popsané na počátku roku 2025. V těchto systémech jsou efektorové Tas proteiny naváděny k molekule dsDNA či ssDNA za pomoci nekódující RNA. Některé z nich jsou pak schopné katalyzovat štěpení cílové molekuly. Obecně TIGR-Tas systémy vykazují podobnost s RNA-řízenými transpozony rodiny IS110 naznačující možný společný původ. Inzerční sekvence jsou běžnou součástí prokaryotních genomů a aktivně přispívají k jejich variabilitě. Podle svých vlastností jsou rozdělovány do jednotlivých rodin. Mechanismus integrace do genomu se u zástupců z rodiny IS110 dlouho nedařilo popsat. Zlom přišel až v roce 2024, kdy bylo popsáno, že si IS110 exprimují spolu s rekombinázou i strukturní nekódující RNA (bridge RNA), která se s rekombinázou specificky váže a umožňuje proces rekombinace. Oba tyto systémy, TIGR-Tas a bridge RNA, mají potenciál k využití v genovém inženýrství pro zavádění zatím neproveditelných změn v genomu a v některých případech i jako možná alternativa k aktuálně nejrozšířenějším systémům CRISPR/Cas, oproti kterým mají řadu výhod i nevýhod. Tato práce se zabývá popisem všech tří systémů a jejich srovnáním mezi sebou, u systému CRISPR/Cas pak především s variantami založenými na Cas12a a Cas9.

Klíčová slova: bridge RNA, Cas9, Cas12, CRISPR/Cas, IS110, *LbCas12a*, TIGR-Tas

Abstract

Non-coding RNAs are at the forefront of the most modern tools used in molecular biology for gene editing, regulation of gene expression, gene therapy, and more. The newest of them is the TIGR-Tas system, which was described at the beginning of 2025. In these small systems, the effector Tas proteins are directed towards molecules of dsDNA or ssDNA by a non-coding RNA. Some of these proteins are then capable of catalyzing cleavage of the target molecule. TIGR-Tas systems generally show resemblance to transposons from the IS110 family, suggesting their common ancestry. Insertion sequences are a common part of prokaryotic genomes and actively contribute to their variability. They are categorized into individual families based on their characteristics. For a long time, the integration process of the IS110 family members into genomes wasn't successfully described. The breakthrough came in 2024 when it was discovered that, along with their recombinase, the IS110 express a structural non-coding RNA (bridge RNA) which specifically binds to the recombinase and facilitates the recombination process. Both of the systems, TIGR-Tas and bridge RNA, show potential for use in gene engineering in introducing previously unfeasible changes in genomes and, in some cases, as an alternative to the currently most widespread CRISPR/Cas systems, compared to which they have a number of advantages and disadvantages. This work describes all three systems, compares them to one another, and in the matter of CRISPR/Cas systems, particularly with the variants based on Cas12a and Cas9.

Keywords: bridge RNA, Cas9, Cas12, CRISPR/Cas, IS110, *LbCas12a*, TIGR-Tas

Obsah

Seznam použitých zkratek.....	1
1. Úvod.....	3
2. TIGR-Tas	3
2.1 Objev TIGR-Tas	4
2.2 TIGR-asociované (Tas) proteiny	5
2.3 Mechanismus	6
2.3.1 Struktura TasR RNP v komplexu s DNA	7
2.4 Evoluce a přirozená funkce systému	8
3. Bridge RNA.....	9
3.1 Inzerční sekvence z rodiny IS110.....	9
3.1.1 Struktura IS110	10
3.2 Mechanismus bridge RNA	11
3.2.1 Synaptický komplex IS621	11
3.3 Aplikace bridge RNA	13
4. CRISPR/Cas	13
4.1 Mechanismus	13
4.2 Cas12	14
4.2.1 Cas12a.....	15
4.2.1.1 Struktura Cas12a	15
4.2.1.2 Mechanismus Cas12a.....	16
4.2.1.3 <i>Lb</i> Cas12a	17
4.3 Cas9	19
4.3.1 Využití a inovace CRISPR/Cas9	19
5. Diskuse	20
6. Závěr.....	22
7. Seznam použité literatury	22

Seznam použitých zkratek

AMK	aminokyselina
Cas	CRISPR-associated proteins
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DBL	Donor-Binding Loop
dDNA	donor DNA
dsDNA	double-strand DNA
IS	Insertion Sequence (inzerční sekvence)
<i>LbCas12a</i>	<i>Lachnospiraceae bacterium Cas12a</i>
LD	Left Donor
LE	Left End (levý konec)
LT	Left Target
LTG	Left Target Guide
MGEs	Mobile Genetic Elements (mobilní genetické elementy)
Nop	Nucleolar protein (jadéřkový protein)
NUC	Nuclease lobe (nukleázový lalok)
OBD	Oligonucleotide-Binding Domain (oligonukleotid-vázající doména)
PAM	Protospacer-Adjacent Motif
<i>ParTasR</i>	<i>Parcubacteria TasR</i>
PI	PAM-Interacting domain (PAM-interagující doména)
RBD	RNA-Binding Domain (RNA-vázající doména)
RD	Right Donor
RE	Right End (pravý konec)
REC	Recognition lobe (rozpoznávací lalok)
RT	Right Target
SL	Stem Loop
<i>SpCas9</i>	<i>Streptococcus pyogenes Cas9</i>
RNP	ribonucleoprotein

RTG	Right Target Guide
snoRNPs	small nucleolar ribonucleoproteins (malá jadéřkové ribonukleoproteiny)
ssDNA	single-strand DNA
<i>Ta</i> TasR	<i>Thermoproteota archaeon</i> TasR
TBL	Target-Binding Loop
tDNA	target DNA (cílová DNA)
TIGR	Tandem Interspaced Guide RNA
Tas	TIGR-associated proteins
TIRs	Terminal Inverted Repeats (terminální invertované repetice)
tracrRNA	trans-activating CRISPR RNA
WED	Wedge Domain

1. Úvod

Systémy založené na spolupráci nekódujících RNA s efektorovými proteiny v současné době vévodí metodám editace genomů užívaných v genovém inženýrství. Největší převrat začal prací Jinek *et al.*, 2012, v níž se autorům podařilo vytvořit programovatelnou chimérickou RNA, která na základě komplementarity navádí endonukleázu Cas9 do cílové oblasti. Tato chimérická RNA vznikla úpravou crRNA a tracrRNA, které jsou nutné pro funkčnost v prokaryotickém adaptivním imunitním systému, z něhož pochází. CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins) systémy tak v genovém inženýrství postupně nahradily dřívější metody využívající nukleáz spojených s tzv. zinkovými prsty (zinc-finger nuclease, ZNF; Carroll, 2011) či nukleázy podobné transkripčním aktivátorům (transcription activator-like effector nuclease, TALEN; Becker and Boch, 2021).

CRISPR/Cas systémy jsou velmi různorodé, objevují se stále nové a ty již známé jsou upravovány a vylepšovány pro nejrůznější využití, příkladem je třeba bakteriofágní CRISPR/Cas Φ (Cas12j) (Pausch *et al.*, 2020) a její uměle vytvořená verze Cas12j2-STU s vyšší specifitou a efektivitou pro editaci rostlin (Sun *et al.*, 2024). Aktuální možnosti využití CRISPR/Cas u rostlin shrnuje publikace Tuncel *et al.*, 2025, o aplikacích v medicíně pojednává Azeez *et al.*, 2024. Existují však i jiné novější RNA řízené systémy, které mají velký potenciál uplatnění. Cílem této práce je představit dva z nově objevených, TIGR-Tas (Tandem Interspaced Guide RNA-TIGR-associated proteins; Faure *et al.*, 2025) a bridge RNA (Durrant *et al.*, 2024), spolu s jedním z nejpoužívanějších CRISPR systémů založeném na variantě Cas12 (Zetsche *et al.*, 2015). Blíže je konkrétně představena varianta *LbCas12a*, její mechanismus, využití a uměle vytvořené modifikace. Práce se snaží shrnout výhodné a nevýhodné vlastnosti těchto systémů pro genové inženýrství a porovnat je s mechanismem jednoho z nejvýznamnějších nástrojů pro editaci genomu CRISPR/Cas9.

2. TIGR-Tas

Na konci února letošního roku, 2025, byl v časopise Science vědeckou skupinou z Broad Institute MIT a Harvardu pod vedením profesora Zhanga publikován článek pojednávající o dosud nepopsaném mechanismu rekombinace řízeném RNA – TIGR-Tas. Jde o rozmanitou

rodinu RNA řízených systémů, které se přirozeně nacházejí převážně u bakteriofágů, archeálních virů (Hochstein *et al.*, 2016) a parazitických bakterií. Jako efekторы u nich slouží Tas proteiny, jejichž společnou charakteristikou je přítomnost domény jadérekových proteinů (Nucleolar Protein, Nop) a do cílových míst jsou naváděny pomocí nekódující RNA (Faure *et al.*, 2025). Jde zatím o vůbec jediný publikovaný článek zabývající se tímto systémem (duben 2025).

2.1 Objev TIGR-Tas

Svůj výzkum začala vědecká skupina v proteinové databázi AlphaFold (Varadi *et al.*, 2022), metagenomické MGnify (Richardson *et al.*, 2023) a virové databázi BFVD (Kim *et al.*, 2025) s cílem identifikovat dosud neznámé molekulární systémy řízené RNA, které jsou obecně snadno programovatelné, a proto mají velký potenciál pro využití v genovém inženýrství. Jako vstupní protein byla využita nukleáza *SpCas9*, respektive její RNA vázající doména (RNA-binding domain, RBD), k níž pak v databázích hledali strukturní homology. Jako nejvhodnější pro další zkoumání vybrali transpozonem kódovanou transpozázu IS110 (inzerční sekvence rodiny 110) gramnegativní bakterie *Tropicimonas sp. IMCC6043* (TroIS110) (Harwati *et al.*, 2009). IS110 funguje jako rekombináza, která pro své fungování vyžaduje nekódující RNA zvanou bridge RNA (Durrant *et al.*, 2024; Hiraizumi *et al.*, 2024). Právě oblast proteinu vázající bridge RNA vykazuje strukturní podobnost s RBD *SpCas9*. Jde o doménu, která se vyskytuje jen u malé skupiny Cas proteinů, existuje tedy možnost, že došlo k integraci IS110 do předka genu této skupiny a část RBD u nich zůstala zachována (Faure *et al.*, 2025).

Pro vyhledávání dalších strukturně příbuzných domén využili vědci RBD IS110 pocházející z *Escherichia coli* kvůli její menší velikosti a jednodušší struktuře v porovnání s ostatními nalezenými homologními proteiny. Výsledek ukázal podobnost s rodinou Nop domén vázající ribonukleoproteiny (RNPs) charakteristickou pro eukaryotní sestřihový faktor Prp31 a C/D box malých jadérekových ribonukleoproteinů (snoRNPs) eukaryot (Liu *et al.*, 2007) a archeí (Aittaleb *et al.*, 2003). Důkazy o společném evolučním původu IS110 a Nop poskytla již práce Vaysset *et al.*, 2025. Za účelem vymezení rodiny proteinů obsahující Nop doménu prohledali autoři Faure *et al.*, 2025 mikrobiální genomické a metagenomické databáze. Při bližší analýze jedné z takto identifikovaných rodin, sestávající se z velké části z proteinů tvořených pouze samotnou Nop doménou, bylo pozorováno, že geny kódující proteiny Nop domény jsou asociovány s dlouhými řadami repetice s vymezenými spacery. Konkrétně jsou tyto řady nazvané TIGR složené z okrajové repetice, spaceru A, loop repetice a spaceru B

v tomto pořadí. Několik totožných TIGR jednotek opakujících se za sebou tvoří TIGR lokus (Faure *et al.*, 2025).

2.2 TIGR-asociované (Tas) proteiny

V současnosti je známých více než 20 000 různých většinou bakteriofágních Tas proteinů, jejichž obecnou charakteristikou je přítomnost RNA vázající Nop domény. Podle jejich struktury je možné Tas proteiny rozdělit do tří základních skupin: TasA, TasH a TasR (Faure *et al.*, 2025).

Nejjednodušší jsou proteiny TasA, které jsou tvořené pouze Nop doménou, což znamená, že samy o sobě nejsou schopné štěpit DNA. Pravděpodobně jde o původní formu Tas proteinů, z níž se fúzí s nukleázovými doménami později vyvinuly endonuklázy TasH a TasR. Ze všech tří skupin je TasA mezi TIGR systémy nejrozšířenější. Funkce TasA proteinů není zcela jasná, nicméně se zdá, že jsou schopné indukovat ohyb substrátové DNA o téměř 180° (Faure *et al.*, 2025) připomínající ohyby DNA vyvolávané lyzogenními fágy při rekombinaci (Rice *et al.*, 1996). Tato podobnost naznačuje, že by TasA mohly sloužit k usnadnění integrace fága do genomu. Geny *tasA* se často nacházejí ve stejném operonu s geny kódujícími protein homologní k ParB nazvaný jednoduše TasParB. ParB je DNA vázající součást ParABS systému, který se podílí na segregaci chromozomů a plazmidů u bakterií (Surtees and Funnell, 2003), asociace TasA s jemu homologním proteinem by tedy mohla znamenat, že i on hraje roli při segregaci genetické informace. Další možnou funkcí je blokáce transkripce či replikace v rámci kompetice s dalšími mobilními genetickými elementy (MGEs; Faure *et al.*, 2025).

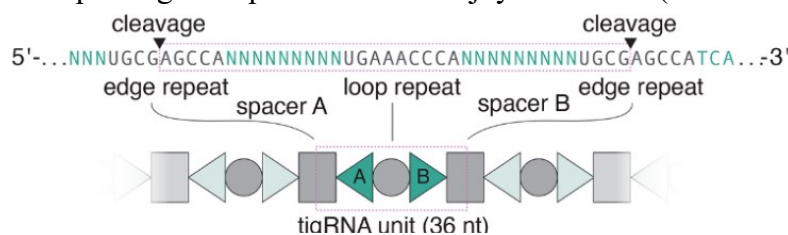
Proteiny ze skupin TasH a TasR jsou RNA řízené endonukleázy, které se od sebe liší svou N-koncovou doménou. TasH mají na N-terminální doméně HNH nukleázu a najdeme je například u bakteriofága *Salicola CGphi29*. HNH proteiny tvoří širokou rodinu nukleázových domén, které mohou být kódovány introny skupin I i II, chromozomální i plasmidovou DNA bakterií a bakteriofágy. Jejich společnou charakteristikou je přítomnost dvou konzervovaných zbytků histidinu (Gorbalenya, 1994) a jednoho asparaginu (Shub, Goodrich-Blair and Eddy, 1994) v aktivním centru enzymu. TasR se vyskytují například u archeálních *Thermoproteota* a na své N-terminální doméně obsahují RuvC endonukleázu, která je však u části zástupců této skupiny inaktivovaná. *In vivo* funguje RuvC jako resolváza Hollidayových struktur, tzn. hraje roli při rekombinaci a opravě DNA (Iwasaki *et al.*, 1991). Kromě štěpení DNA dokáže TasR stejně jako TasA indukovat ohyb substrátové DNA (Faure *et al.*, 2025).

Výhodnou vlastností Tas proteinů i pro potenciální využití v genovém inženýrství je jejich nezávislost na protospacer-adjacent motif (PAM; viz kapitola 2.3) a malá velikost. Autorům práce Faure *et al.*, 2025 se podařilo ukázat, že některé Tas proteiny mohou být naprogramovány pro štěpení DNA v lidských buňkách. Jmenovitě byly pro tuto demonstraci použity nukleázy *TaTasR* (z archea *Thermoproteota*) a *ParTasR* (z bakterie *Parcubacteria*) transfekované do buněk HEK239FT, které pocházejí z linie odvozené z buněk ledvin lidského embrya (Russell *et al.*, 1977). Oba proteiny způsobily v cílových místech vznik krátkých inzercí a delecí, takzvaných indelů. V lokusu, v němž byla pozorována nejvyšší efektivita editace, vzniklo vlivem *ParTasR* až 3,6 % indelů oproti pouhým 0,8 % u *TaTasR* (Faure *et al.*, 2025). Zatím však jde o jediný zdokumentovaný pokus aplikace TIGR-Tas systému v eukaryotních buňkách.

2.3 Mechanismus

TIGR lokus je složen z nejčastěji pěti až patnácti totožných TIGR jednotek, jež jsou tvořeny dvěma rozdílnými repeticemi a dvěma různými spacery, které jsou uspořádány za sebou v následujícím pořadí: okrajová repetice, spacer A, smyčka (loop) repetice a spacer B (Obrázek 1). Délka repetice se pohybuje mezi osmi až dvanácti nukleotidy, spacery mohou být dlouhé až dvanáct nukleotidů, ve většině případů jich však mají pouze devět. V celém genomu může být i více TIGR lokusů, které se mezi sebou liší sekvencemi repetice a délkou spacerů (Faure *et al.*, 2025).

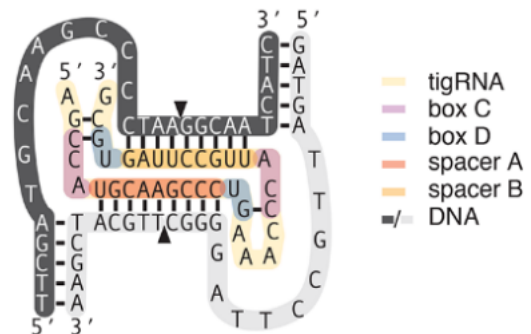
TIGR sekvence jsou transkribovány do delších prekurzorů, z nichž pak štěpením vznikají jednotlivé molekuly nazvané tigRNA dlouhé 36 nukleotidů. Štěpení se téměř vždy odehrává na stejné pozici uprostřed okrajových repetic. Jejich zbylá část pak tvoří ohraničení vnitřní struktury tigRNA, která se dále skládá ze spaceru A, spaceru B a mezi nimi umístěnou smyčkou repetice. Jak přesně samotný proces štěpení probíhá zatím není jasné, nicméně se zdá, že přestože nukleázová aktivita Tas proteinů není pro vyštěpení tigRNA z prekurzoru nutná, bez jejich přítomnosti k němu nedojde. Autoři spekulují, že je za maturaci zodpovědná zatím neznámá vysoce konzervovaná hostitelská RNasa, jelikož i u vzdáleně příbuzných TIGR systémů probíhalo štěpení tigRNA prekurzorů na stejných místech (Faure *et al.*, 2025).



Obrázek 1 - Schéma TIGR jednotky. Převzato z Faure *et al.*, 2025.

Úkolem jednotlivých tigRNA je vázat a následně směřovat Tas proteiny k cílovým místům na DNA. Ty na rozdíl od CRISPR systémů, které vyžadují přítomnost PAM (Mojica *et al.*, 2009), žádné podobné sekvence nepotřebují.

Specifita TIGR-Tas systémů je zajištěna tandemovou spoluprací obou spacerů, kdy spacer A vytváří pár s jemu komplementární sekvencí na jednom vlákně DNA, zatímco spacer B obdobně interaguje s vláknem protilehlým (Obrázek 2).



Obrázek 2 - Schéma vazby tigRNA na DNA pomocí spacerů. Převzato z Faure *et al.*, 2025.

Štěpení DNA řízené tigRNA je tedy závislé

pouze na sekvenci spacerů homologní k cílové sekvenci. Rozdělení naváděcí sekvence do dvou částí zřejmě zvyšuje specifitu systému a zabraňuje zacílení Tas proteinů na svoje vlastní TIGR sekvence. Navíc se zdá, že tato struktura zamezuje rekombinaci a tím i integraci fágů, což vede ke zvýšení stability TIGR jednotek (Faure *et al.*, 2025).

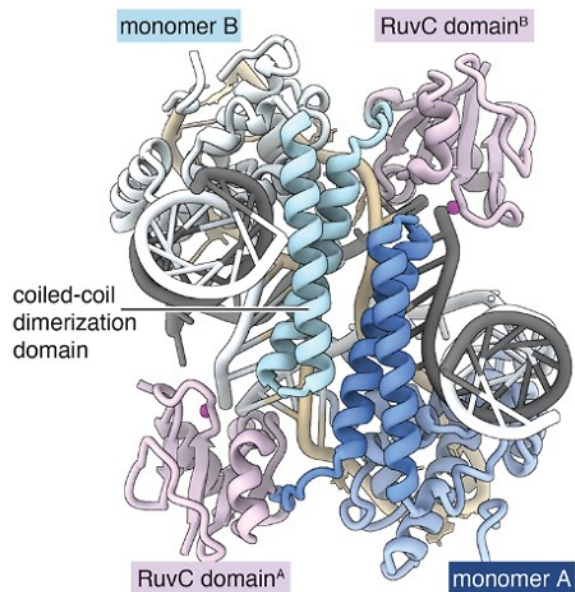
Systém obecně není příliš tolerantní k chybnému párování bází. Mutace na DNA narušující štěpení jednoho z vláken vždy ovlivní i štěpení na druhém, zvláště pak pokud se tato mutace vyskytuje v takzvaném seed regionu, který se nachází v sekvencích komplementárních ke spacerům dva nukleotidy 3' směrem od místa štěpení. Naopak tolerantnější je TIGR-Tas k mutacím na okrajích těchto sekvencí. Zajímavou výjimkou z tohoto pravidla je takzvaná nikázová mutace. Ta je navozena transverzí dvou nukleotidů v 5' místě štěpení a ovlivněno je pouze vlákno, na kterém k ní došlo, kdežto na zbylém vlákně proces probíhá jako za normálních okolností. Tas protein tedy v tomto případě funguje jako nikáza, štěpící pouze jedno z cílových vláken (Faure *et al.*, 2025).

TIGR-Tas systémy za běžných okolností štěpí dvouvláknovou DNA, dokážou nicméně rozštěpit i jednovláknovou DNA, pokud obsahuje sekvenci komplementární k jednomu ze spacerů. Maximální možná délka mezery mezi komplementárními sekvencemi spacerů A a B, při níž je štěpení DNA ještě možné, odpovídá dvěma pářům bází. RNA, ať už jedno či dvouvláknovou, neštěpí (Faure *et al.*, 2025).

2.3.1 Struktura TasR RNP v komplexu s DNA

Propojením dvou TasR protomerů skrze coiled-coil doménu vzniká symetrický dimer, který váže jednu molekulu tigRNA a jednu molekulu substrátové DNA. tigRNA interaguje s oběma protomery a nabývá tvar osmičky. Celá struktura dohromady vykazuje C2 pseudosymetrii.

Při vstupu substrátové DNA do komplexu způsobí TasR její ohyb o zhruba 180° a dochází také k rozvinutí dvoušroubovice DNA v oblasti cílové sekvence. Zbytek substrátové DNA s TasR RNP komplexem nijak neinteraguje a její vlákna po celou dobu zůstávají párována. Spacer A tigrRNA se váže ke své komplementární sekvenci na DNA a společně se stáčí do struktury A-DNA. Nekomplementární vlákno je od nich odděleno takzvanou wedge loop inzercí domény TasR proteinu. Stejná je i interakce spaceru B a jeho komplementární sekvence na druhém vlákně substrátu. Malý žlábek tigrRNA/DNA duplexu spočívá na coiled-coil doméně a vůbec nejbližší s ní interaguje skrz seed region tigrRNA. Okrajová repetice interaguje s Nop doménou jednoho z protomerů, zatímco smyčka repetice interaguje velmi podobným způsobem s tím druhým. Podobnost těchto interakcí je dána konzervovanými motivy UG na začátcích a CCA na koncích repetice u různých TIGR-TasR. Vzhledem k jejich podobnosti s danými sekvencemi na snoRNA nazývají autoři článku CCA motiv jako box C a UG jako box D. Ke štěpení substrátové DNA dochází v místě $5'$ směrem od nukleotidu, který tvoří pár s pátou bází downstream od C boxu na tigrRNA (Faure *et al.*, 2025).



Obrázek 3 - Struktura TasR dimeru v komplexu s DNA (DNA bílošedě, tigrRNA béžově) Převzato z Faure *et al.*, 2025.

2.4 Evoluce a přirozená funkce systému

Není zcela jasné, k čemu přesně TIGR systémy slouží u jejich přirozených nositelů. Je poměrně obtížné určit jejich obvyklé cíle na DNA, jelikož jsou převážně součástí metagenomů. Poněvadž se nacházejí hlavně u MGEs, pracovali autoři s hypotézou, že slouží k obraně proti imunitě hostitelského organismu, zajišťují konkurenceschopnost s jinými MGEs, mobilitu genů a podílejí se na jejich regulaci. Exprese TasR v *Escherichia coli* nicméně podle nich nevedla k ochraně bakterie proti infekci fágem ani proti přenosu plazmidů konjugací, silně negativně však působila na zachování plazmidů. Zřejmě tedy bude mít velkou roli související s dědičností MGEs (Faure *et al.*, 2025).

TIGR systémy vykazují podobnosti s transpozony rodiny IS110 a C/D boxem snoRNA naznačující příbuznost mezi těmito systémy. Autoři článku teoretizují, že tyto systémy sdílejí

společného předka, kterým mohl být protein TasA tvořený pouze Nop doménou. Během evoluce pak docházelo k jeho fúzi s různými doménami za vzniku nových proteinů s odlišnými funkcemi. Kromě TasR skupiny proteinů vzniklé fúzí TasA s RuvC doménou a TasH analogicky s HNH doménou, byl fúzí s transpozázou obsahující DEDD motiv pravděpodobně vytvořen předchůdce rodiny transpozonů IS110. Fúze s TasA dala vzniknout i N-koncové RuvC nukleázové doméně snoRNP. Ta je stejně jako u TasR proteinů CPR bakterií inaktivovaná, což by podle vědců mohlo znamenat, že se právě tato skupina TasR vyvinula jako předchůdce RNA řízených systémů C/D box snoRNA (Faure *et al.*, 2025).

3. Bridge RNA

TIGR systémům příbuznou rodinou transpozonů IS110 se v článku z roku 2024 zabývali vědci pod vedením doktora Patricka D. Hsu z Kalifornské univerzity v Berkeley. Jejich výzkum odhalil, že IS této rodiny kódují kromě transpozázy také nekódující RNA nezbytnou pro transpozici. Tato bridge RNA umožňuje modifikaci genomu prostřednictvím vystřížení, inverze nebo inserce. Autoři původní publikace Durrant *et al.*, 2024 podle svých slov očekávají, že mechanismus rekombinace pozorovaný u bridge RNA odstartuje rozvoj třetí generace programovatelných RNA systémů a přinese ještě lepší možnosti editace genomu než v současnosti hojně využívané mechanismy CRISPR/Cas.

3.1 Inzerční sekvence z rodiny IS110

IS jsou běžnou součástí prokaryotních genomů (Hirsch, Starlinger and Brachet, 1972) a svou aktivitou přispívají k jejich variabilitě. Jde o nejjednodušší transponibilní elementy nesoucí pouze informace nutné pro jejich mobilitu, typicky tedy bývají tvořeny jen krátkými terminálními invertovanými repeticemi (TIRs), mezi kterými se nachází gen pro transpozázu (Mahillon and Chandler, 1998). Jejich inserce často způsobí inaktivaci genů, v některých případech však mohou ovlivňováním sousedících genů, jako jsou například geny resistance k antibiotikům, hostitelskému organismu poskytovat i výhodné vlastnosti (Razavi *et al.*, 2020). Účinek IS nemusí být jen lokální, mohou způsobovat i velké inserce či delece vedoucí ke komplexním změnám genomů (Schneider and Lenski, 2004).

IS jsou velmi diverzifikované a pro snazší orientaci se dělí do rodin primárně v závislosti na podobnosti jejich struktury, organizace čtecích rámců, transpozáz, které kódují a podle TIRs (Siguier *et al.*, 2006). Často se také používá členění IS rodin do širších skupin podle toho,

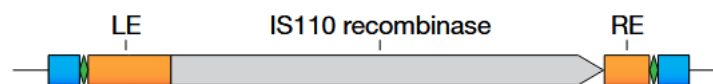
jaké konzervované zbytky se nacházejí v katalytickém místě jejich transpozáz. IS110 je v současnosti jedinou známou rodinou, jejíž transpozáza má katalytický motiv DEDD (aspartát, glutamát, aspartát, aspartát; Siguier *et al.*, 2015). Unikátní je díky tomu i jejich způsob integrace do genomu, při němž je tvořena Hollidayova struktura. Jako resolváza Hollidayových struktur homologní s RuvC resolvázou (Iwasaki *et al.*, 1991) funguje N-terminální doména transpozázy, díky mechanismu integrace běžně označované jako rekombináza (Buchner *et al.*, 2005). Typická je pro IS110 rodinu též nepřítomnost TIRs, které jsou u jiných rodin důležité pro transpozici (Siguier *et al.*, 2015).

IS110 se dlouho nedařilo úspěšně studovat *in vitro* (Siguier, Gourbeyre and Chandler, 2014). S vysvětlením, proč tomu tak je, přišli až autoři práce Durrant *et al.*, 2024, kteří popsali nekódující RNA exprimovanou cirkulární formou IS110 vzniklou vystřížením elementu z genomu (Partridge and Hall, 2003) nezbytnou pro integraci. Jako bridge RNA byla tato molekula pojmenována s odkazem na mechanismus jejího fungování – je totiž bispecifická a při rekombinaci vytváří mezi donorovou a cílovou DNA jakýsi most, což následně umožní průběh samotné rekombinace řízené IS110 rekombinázou (Durrant *et al.*, 2024).

3.1.1 Struktura IS110

Na okrajích IS110 se nacházejí sekvence pro rozpoznání cílové DNA, směrem dovnitř pak následují krátké core sekvence, nekódující levý (left end, LE) a pravý konec (right end, RE), mezi kterými je kódovaná samotná rekombináza (Obrázek 4). Poměr délek jednotlivých částí se u různých zástupců liší. Při

vyštěpení IS110 z genomu dojde k jejímu rozdělení na dvě jednotky: část tvořená genem pro



Obrázek 4 - Struktura IS110 (modře sekvence pro rozpoznání cílové DNA, zeleně core sekvence). Převzato z Durrant *et al.*, 2024.

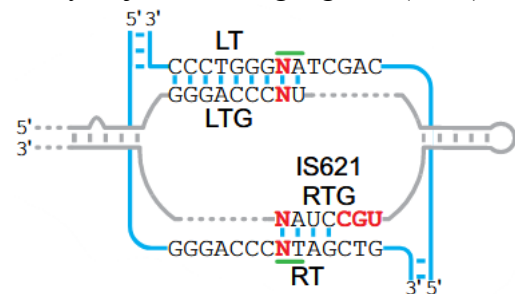
rekombinázu, LE, RE a jednou core sekvencí formují cirkulární strukturu, zatímco zbylá core sekvence společně s oblastmi genů pro rozpoznání cílového místa vytvářejí takzvaný dsDNA target site.

U primárně zkoumané IS621, kterou běžně nacházíme u některých kmenů bakterie *Escherichia coli* (Choi, Ohta and Ohtsubo, 2003), stejně jako u většiny dalších zástupců ze stejné podskupiny IS, je bridge RNA exprimována na LE elementu, u jiných podskupin se ale může nacházet na RE (Durrant *et al.*, 2024).

3.2 Mechanismus bridge RNA

Bridge RNA IS621 a jejích orthologů obsahuje dvě smyčky, jež autoři Durrant *et al.*, 2024 nazývají jednoduše podle sekvence, kterou váží, jako target-binding loop (TBL), nacházející se blíže k 5' konci, a donor-binding loop (DBL). Bylo vypořádováno, že 5' konec obou tvoří páry s bázemi spodního vlákna DNA a délka tohoto úseku je buď osm nebo devět nukleotidů. 3' konce párují s horními vlákny pomocí sekvence dlouhé čtyři až sedm nukleotidů. Obě smyčky bridge RNA mohou být nezávisle na sobě přeprogramované pro vazbu různých DNA sekvencí, zároveň vykazují jen nízkou toleranci chybného párování bází (Durrant *et al.*, 2024).

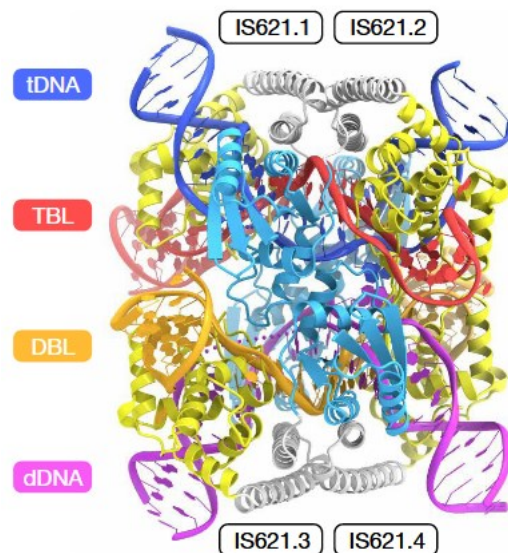
Target-binding loop je tvořena dvěma oblastmi označovanými jako left target guide (LTG) párující s bázemi v levé části spodního vlákna cílové DNA (left target, LT). Oproti tomu right target guide (RTG) cílí na pravou část horního vlákna DNA (right target, RT). Mezi LTG a RTG se nachází stem loop. Struktura donor-binding loop je analogická (Obrázek 5; Durrant *et al.*, 2024).



Obrázek 5 - Schéma vazby bridge RNA (šedě) na cílovou DNA (modře). Převzato u Durrant *et al.*, 2024.

3.2.1 Synaptický komplex IS621

Synaptický komplex je tvořen čtyřmi protomery IS621 rekombinázy, TBL a DBL částmi bridge RNA, donorovou (dDNA) a cílovou DNA (tDNA). Každý IS621 protomer obsahuje tři domény: RuvC, coiled-coil a Tnp. RuvC doména se skládá z β -listu z pěti vláken lemovaného čtyřmi α -helixy, její aktivní místo je pak tvořeno DEDD motivem (D11, E60, D102 and D105). Před zformováním samotného tetrameru dochází nejprve ke vzniku dvou dimerů prostřednictvím coiled-coil domény, která se skládá ze dvou α -helixů. Tnp doména je formována sedmi α -helixy a zajímavá je hlavně tím, že nevykazuje strukturní podobnost s žádným známým proteinem (Hiraizumi *et al.*, 2024).



Obrázek 6 - Synaptický komplex IS621. Převzato z Hiraizumi *et al.*, 2024.

Jeden z protomerů v každém dimeru má uspořádanou smyčku S241 v Tnp doméně tak, aby interagovala se smyčkou D102 RuvC domény protomeru s neuspořádanou S241 smyčkou v opačném dimeru. Toto propojení obou domén zřejmě zabraňuje tomu, aby byla DNA štěpena ještě předtím, než dojde k vytvoření synaptického komplexu.

TBL a DBL bridge RNA rozpoznané Tnp doménou interagují s rekombinázou skrz vazby cukr-fosfátové kostry za vzniku dvou podobných ribonukleoproteinových komplexů. Bridge RNA váže cílovou a donorovou DNA mechanismem popsáním výše, což vede k formaci synaptického komplexu složeného z dimeru rekombinázy, TBL a tDNA, které spolu s druhým dimerem rekombinázy, DBL a dDNA komunikují pomocí svých RuvC domén. Celý komplex je stabilizován vodíkovými můstky mezi zbytky RuvC a Tnp domén a také interakcemi mezi DBL stem loop (DBL-SL) a RuvC.1. S tím souvisí i fakt, že ve výjimečných případech byla *in vitro* pozorována rekombinace mezi dvěma donorovými DNA, ale nikdy mezi dvěma cílovými, jelikož postrádají interakci DBL-SL a RuvC, která se zdá být pro formaci synaptického komplexu nezbytná. Tím je zabráněno mnoha nechtěným editacím genomu.

Horní vlákno tDNA je štěpeno v kompozitním aktivním místě RuvC.1-Tnp.4, horní vlákno dDNA pak v místě RuvC.3-Tnp3. Štěpení probíhá v obou případech mezi devátým a desátým nukleotidem, které následují po CT jádrové oblasti, za vzniku 5'-fosfoserinových intermediátů, jež jsou stabilizovány koordinovanými atomy Mg^{2+} a vodíkovými můstky. Aby mohlo dojít k prohození vláken tDNA a dDNA, musí být zformovány nové vazby mezi bázemi v oblastech zvaných handshake guides (HSGs) v pozicích 81-82 a 166-167 na bridge RNA. Jinými slovy dochází nejprve k uvolnění dosavadních vazeb mezi TBL/DBL bridge RNA a molekulami substrátu, a poté vznikají vazby mezi tDNA/dDNA a opačnou smyčkou, tedy TBL-dDNA a DBL-tDNA. Po prodloužení komplementární sekvence mezi RTG a jeho cílovým místem na DNA probíhala rekombinace v *Escherichia coli* s vyšší efektivitou než za běžných podmínek. RDG zůstává navázaný na dDNA i ve chvíli, kdy se již začíná párovat s přicházejícím CT regionem horního vlákna tDNA. RTG se chová analogicky.

Fosfodiesterové vazby mezi nukleotidy T9 dDNA a A10 tDNA a mezi T9 tDNA a A10 dDNA spolu s disociací zbytků z S.241.4 z A10 tDNA a S241.3 z T10 dDNA vedou ke vzniku intermediární Hollidayovy struktury. Následuje organizace smyček S241 v Tnp.1 a 3 a štěpení spodních vláken dDNA a tDNA, opět mezi devátým a desátým nukleotidem, v RuvC.2-Tnp.3 a RuvC.4-Tnp1 aktivních místech. Toto štěpení vede k rozložení Hollidayovy struktury a uvolnění produktů (Hiraizumi *et al.*, 2024).

3.3 Aplikace bridge RNA

Při zkoumání specifity systému přeprogramovali autoři Durrant *et al.*, 2024 bridge RNA tak, aby cílila na sekvence, které se v genomu *Escherichia coli* nacházejí pouze jednou, u části molekul navíc uměle prodloužili RTG, aby byla zvýšena přesnost. U bridge RNA s normální délkou RTG se specifita inzerce do cílového místa pohybovala mezi 51.2–89.4 %, zatímco s prodlouženým RTG byla 65.4–94.0 % a navíc došlo i k významné redukci off-target inzerce. Přítomnost excizí a inverzí, které jsou taktéž programovatelné, byla za stejných podmínek u všech buněk v tomto pořadí 32.2–98.9 % a 4.54–98.2 %. Víme tedy, že u prokaryotních organismů všechny tři procesy mohou probíhat i s vysokou efektivitou, zatím však neexistují žádné důkazy, že by systém fungoval i v buňkách eukaryot (Durrant *et al.*, 2024).

4. CRISPR/Cas

TIGR-Tas i bridge RNA slibují mnohé, jejich využití v praxi však zatím zůstává pouze teoretické. Jiné RNA řízené systémy jsou pro editaci genomů nicméně používány běžně, přičemž velkou zásluhu na tom mají právě systémy CRISPR/Cas.

Termín CRISPR byl zaveden skupinou pod vedením doktora Ruuda Jansena v roce 2002 v článku, který se konkrétně zabýval Cas proteiny (Jansen *et al.*, 2002). Jako první však CRISPR lokus pozoroval a zdokumentoval již profesor Yoshizumi Ishino se svým týmem v roce 1987 při zkoumání *iap* genu *Escherichia coli* zodpovědného za přeměnu izoenzymu alkalické fosfatázy (Ishino *et al.*, 1987). Teprve o dvě desítky let později bylo experimentálně potvrzeno, že se jedná o součást bakteriálního a archeálního imunitního systému, když autoři Barrangou *et al.*, 2007 upravovali CRISPR lokusy *Streptococcus thermophilus* přidáváním spacerů, čímž bakterie získala resistenci proti daným fágům, a jejich odebíráním, které naopak způsobilo ztrátu obranyschopnosti. Inaktivací *cas* genů také přišli na to, že pro efektivní imunitu nestačí pouze přítomnost spacerů, ale je také nutná aktivita Cas proteinů (Barrangou *et al.*, 2007).

4.1 Mechanismus

Imunitní odpověď zprostředkovaná CRISPR/Cas se může v jednotlivých krocích u různých systémů lišit, princip však zůstává stejný. CRISPR lokus je transkribován do pre-crRNA, na kterou se v rámci maturace může vázat *trans*-aktivační RNA (tracrRNA; Deltcheva *et al.*,

2011), načež dochází k jejímu štěpení na jednotlivé crRNA tvořené spacerem a částí repetice a vzniká komplex s Cas proteinem (Brouns *et al.*, 2008). Při invazi prokaryotní buňky virem, popřípadě plazmidem, slouží Cas k rozpoznání specifické oblasti na této exogenní DNA zvané PAM (Mojica *et al.*, 2009). Pokud je daná sekvence na nukleové kyselině rozpoznána, Cas nukleáza v její bezprostřední blízkosti rozmotá DNA, což umožní vazbu crRNA na komplementární sekvenci DNA, po níž následuje samotné štěpení (Gasiunas *et al.*, 2012). Adaptivní část procesu spočívá v tom, že po odštěpení protospaceru z cizorodé DNA dochází k jeho začlenění do CRISPR lokusu v podobě nového spaceru, což urychlí a zefektivní imunitní odpověď organismu, pokud se s daným virem znovu setká (Pourcel, Salvignol and Vergnaud, 2005).

CRISPR lokus je tvořen krátkými přímými identickými repeticemi s vymezenými spacery, které pocházejí z cizorodé DNA (Mojica *et al.*, 2005). Repetice jsou obvykle dlouhé 23 až 48 párů bází a v rámci daného lokusu je jejich struktura velmi konzervovaná (Kunin, Sorek and Hugenholtz, 2007). Délka spacerů se běžně pohybuje v rozmezí 21 až 72 párů bází.

V blízkosti lokusů se typicky nacházejí geny kódující Cas proteiny, což je široká rodina proteinů s různými aktivitami, například nukleázovou, helikázovou nebo polymerázovou (Haft *et al.*, 2005).

CRISPR/Cas systémy vykazují velmi širokou diverzitu. V současné době jsou rozlišovány dvě třídy, pod které je řazeno šest typů a dohromady třicet tři podtypů (Makarova *et al.*, 2020). Ačkoli jsou systémy 1. třídy méně studovány, v přírodě se vyskytují mnohem častěji. Je to zřejmě dáno tím, že je pro ně typická tvorba multipodjednotkových efektorových komplexů z několika Cas proteinů, které mohou mít různou enzymatickou aktivitu. Mohou kromě dsDNA štěpit také ssDNA a RNA nebo se například podílet na syntéze druhých poslů (Liu and Doudna, 2020). Efektorové komplexy se také ve spolupráci s dalšími proteiny podílejí na maturaci pre-crRNA, při níž není potřeba tracrRNA. Patří sem typy I, III a IV. Systémy 2. třídy se oproti tomu vyznačují tím, že veškeré funkce efektorového komplexu zastává pouze jeden protein, proto jsou také vhodnější k využití v genovém inženýrství. Řadí se mezi ně podtypy II, V a VI (Makarova *et al.*, 2020).

4.2 Cas12

CRISPR/Cas12 patří stejně jako Cas9 do 2. třídy CRISPR/Cas systémů, což znamená, že jejich efektorový komplex je tvořen pouze jedním proteinem. Konkrétně se řadí mezi systémy typu V, jejichž charakteristikou je, že příslušné Cas proteiny obsahují pouze RuvC podobnou

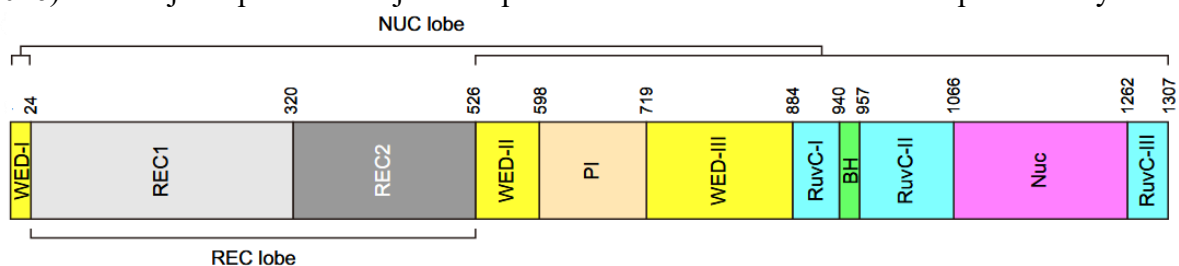
endonukleázovou doménu (Makarova and Koonin, 2015). Je třeba mít na paměti, že v některých publikacích, ale zvláště v popularizačních a nevědeckých článkách, odkazuje termín Cas12 na nukleázu Cas12a dříve označovanou jako Cpf1, která je však jen jedním z dosud známých podtypů (V-A). Těch je v současnosti rozlišováno celkem 10 (Makarova *et al.*, 2020) a liší se mezi sebou mimo jiné velikostí, rozpoznávanými PAM sekvencemi, schopností zpracovat pre-crRNA či substrátovou nukleovou kyselinou. Například proteiny CasΦ (V-J), respektive Cas12j, jsou dlouhé mezi 700 až 800 aminokyselinami (AMK) a rozpoznávají T bohaté PAM sekvence včetně minimální TBN, kdy B = G, T nebo C. RuvC doména CasΦ slouží ke štěpení pre-crRNA i DNA (Pausch *et al.*, 2020). Cas12f (V-F), původně Cas14, tvoří pouhých 400 až 700 AMK, nerozpoznává žádnou PAM sekvenci, není schopná zpracovat vlastní pre-crRNA a štěpí pouze ssDNA (Harrington *et al.*, 2018).

4.2.1 Cas12a

Vůbec prvním objeveným Cas12 proteinem byla právě Cas12a. V databázi proteinových rodin TIGRFAMs byla popsána jako CRISPR-asociovaný protein již v roce 2012 pod svým původním názvem Cpf1, respektive CRISPR from *Prevotella* and *Francisella* 1, který vychází z jejího přirozeného výskytu v některých kmenech těchto bakterií (Schunder *et al.*, 2013). K jejímu přejmenování došlo až po popsání dalších proteinů patřících do rodiny Cas12 (Shmakov *et al.*, 2017).

4.2.1.1 Struktura Cas12a

V závislosti na tom, ze kterého organismu byla endonukleáza Cas12a izolována, se její délka obvykle pohybuje mezi 1100 a 1300 AMK (*LbCas12a* 1228 AMK, *FnCas12a* 1300 AMK, *AsCas12a* 1352 AMK), což z ní dělá v průměru největší ze všech Cas12 proteinů. Její struktura je tvořena dvěma laloky označovanými obdobně jako u Cas9 rozpoznávací (REC) a nukleázový (NUC), které jsou mezi sebou propojeny skrze bridge-helix (BH) motif. α -helikální REC lalok se skládá ze dvou domén: helical I a helical II. NUC lalok obsahuje nukleázovou (Nuc), oligonukleotid-vázající (oligonucleotide-binding, OBD), RuvC doménu a v neposlední řadě také looped-out helikální doménu (LHD; Dong *et al.*, 2016; Gao *et al.*, 2016). V novějších publikacích jsou též používána označení REC1 a REC2 pro domény REC

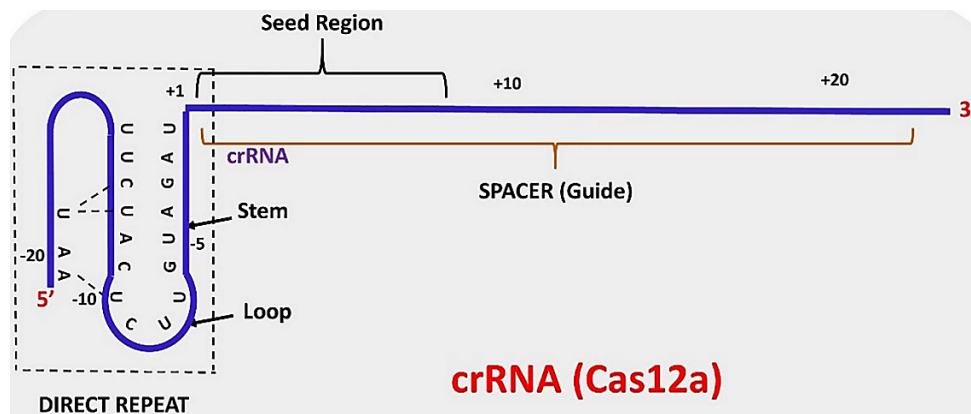


Obrázek 7 - Organizace domén *AsCas12a*. Převzato z Yamano *et al.*, 2016.

laloku, wedge (WED) místo OBD a PAM-interagující (PI) místo LHD kvůli jejich funkční podobnosti se stejnojmennými doménami *SpCas9*. Přestože mají podobnou roli, strukturně nemají Nuc, WED a PI mimo proteinovou rodinu Cas12 homologů. WED doména Cas12a se navíc odlišuje od Cas9 svou RNázovou aktivitou (Yamano *et al.*, 2016; Stella, Alcón and Montoya, 2017a; Jianwei *et al.*, 2023a).

4.2.1.2 Mechanismus Cas12a

CRISPR lokus je transkribován do pre-crRNA, která je následně zpracována na jednotlivé crRNA. Cas12a funguje nejen jako DNáza, ale také jako RNáza a na rozdíl od Cas9 tak ke zpracování pre-crRNA nepotřebuje hostitelskou RNázu a k následné maturaci crRNA nevyžaduje tracrRNA ani jiný typ nekódujících RNA molekul (Fonfara *et al.*, 2016). Každá zralá molekula crRNA je dlouhá 42-44 nukleotidů začíná sekvencí přímých repetit dlouhou 19 nukleotidů, která tvoří kmenovou smyčku (stem loop). Po ní následuje spacer tvořený 23-25 nukleotidy, tedy v opačném pořadí, než je tomu u CRISPR/Cas systémů typu II (Zetsche *et al.*, 2015).



Obrázek 8 – Schéma crRNA Cas12a. Převzato z Bandyopadhyay *et al.*, 2020.

Cas12a štěpí pre-crRNA vysoce sekvencně a strukturně specificky čtyři nukleotidy upstream od stem loop. Aby ke štěpení pre-crRNA vůbec došlo, musí kromě stem loop obsahovat i kompletní sekvenci repetit. Pokud je smyčka příliš krátká, nebo pokud došlo k záměně byt' jen jediného nukleotidu v oblasti mezi ní a místem štěpení (tzv. repeat recognition sequence, RRS), pak Cas12a pre-crRNA téměř nezpracovává (Fonfara *et al.*, 2016).

Štěpení dsDNA zprostředkované Cas12a je částečně tolerantní vůči mutacím, které způsobují zkrácení sekvencí. To bylo konkrétně demonstrováno na *FnCas12a* (Cas12a z *Francisella tularensis* subsp. *novicida*, kmen U112), která zprostředkovává detekovatelné štěpení DNA, pokud je délka vedoucí sekvence, respektive spaceru, alespoň 16 nukleotidů. Minimálně 16, nejlépe však 17 a více nukleotidů musí obsahovat přímé repetice, a to jak pro stabilní vazbu, tak pro samotné štěpení. Oproti tomu Cas9 stačí k vytvoření stabilní vazby nukleotidů pouze 9

a pro průběh štěpení 16 (Singh *et al.*, 2018). Cas12a snáší substituce bází v oblasti smyčky repetice s výjimkou uracilu, na který navazuje spacer. Mutace narušující strukturu stem loop zabraňující vzniku RNA duplexu vedou ke ztrátě endonukleázové aktivity (Zetsche *et al.*, 2015). Enzymatická aktivita Cas12a také není příliš tolerantní k chybnému párování bází a snižují ji i takové chyby, které nijak neovlivňují vznik stabilní vazby crRNA a DNA (Singh *et al.*, 2018).

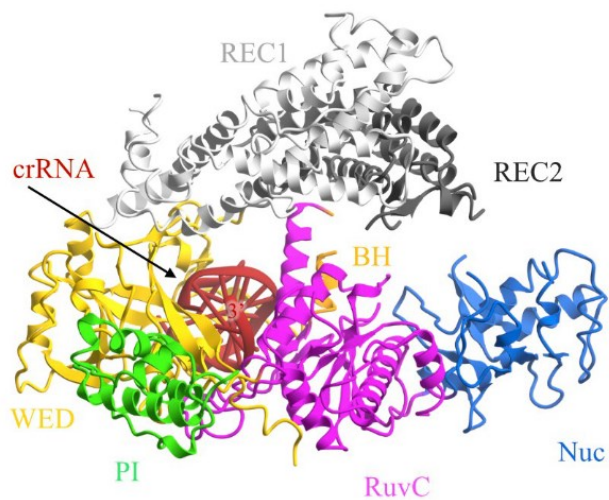
K rozpoznání cílové dsDNA je nezbytná T bohatá PAM sekvence, u většiny včetně *LbCas12a* konkrétně TTTV, kdy V = A, G nebo C. Štěpení DNA generuje kohezivní konce s 5' přesahem čtyři až pět nukleotidů (Zetsche *et al.*, 2015). S o něco nižší efektivitou je Cas12a schopná katalyzovat také nespecifické *trans*-štěpení ssDNA závislé pouze na komplementaritě sekvence substrátové molekuly (Chen *et al.*, 2018).

4.2.1.3 *LbCas12a*

Existuje mnoho různých druhů Cas12a, ovšem ne všechny z nich jsou schopny fungovat v eukaryotických buňkách s dostatečně vysokou efektivitou na to, aby v genovém inženýrství našly stálé využití. Mezi nejvyužívanější patří v současné době kromě *LbCas12a* z bakterie *Lachnospiraceae bacterium* také *AsCas12a* z bakterie *Acidaminococcus sp.*, *Fncas12a* z *Francisella novicida* či *Mb2Cas12a* z *Moraxella bovoculi* (Tuncel *et al.*, 2025).

4.2.1.3.1 Struktura *LbCas12a*

LbCas12a funguje jako monomer. Samotný protein má protáhlou strukturu, na prvním z jeho konců je umístěna REC1 doména, na druhém pak Nuc doména. Oba laloky jsou mezi sebou propojeny dvěma pružnými smyčkami. Navázání crRNA vede ke zhutnění proteinu. Výsledkem této změny konformace je konstrukce trojúhelníkovitého tvaru, uprostřed níž se nachází velký pozitivně nabitý kanál. Jedna stěna se skládá z OBD a C-koncové



Obrázek 9 - Struktura *LbCas12a* v komplexu s crRNA. Převzato z Jianwei *et al.*, 2023.

RuvC domény dělené na tři úseky označované RuvC I-III. Druhá stěna je tvořena helical I a helical II podjednotkami REC domény a třetí vzniká interakcí Nuc domény s helical II. Jedno z β -vláken helical I se páruje s vláknem z OBD a zasahuje do centrálního β -barelu OBD, který je na jedné straně obklopen třemi a na druhé čtyřmi helixy. Naváděcí crRNA se

z největší části nachází mezi dlouhou vlásenkou a dvěma helixy OBD, zbylý úsek interaguje s β -listem a následujícím α -helixem propojujícími RuvC-I a RuvC-II. 3' konec crRNA směřuje do centrálního kanálu proteinu. Navázání přímé sekvence crRNA na *LbCas12a* způsobí její značné zakřivení a vytvoření krátké stem-loop struktury důležité pro zachování konformace crRNA. Struktura je stabilizována vodíkovými můstky mezi nukleovou kyselinou a proteinem a také kationtem $[\text{Mg}(\text{H}_2\text{O})_6]^{2+}$. Absence Mg^{2+} (popřípadě jiných kovových iontů s podobnou geometrií) způsobí významné snížení afinity mezi crRNA a *LbCas12a*, což pochopitelně vede ke ztrátě nukleázové aktivity proteinu (Dong *et al.*, 2016).

Vazbou substrátové DNA vzniká ternární komplex. S crRNA páruje cílové (target, t) vlákno, nekomplementární necílové (non-target, nt) vlákno je vytlačeno směrem k DNA nukleázové oblasti a vytváří R-loop (Stella, Alcón and Montoya, 2017b). V oblastech lemujících R-loop zůstávají řetězce substrátové DNA párovány mezi sebou (Swarts, Oost and Jinek, 2017). RNA-DNA duplex prochází centrálním kanálem proteinu, který je oproti binárnímu komplexu rozšířený. Postavení dvou laloků *LbCas12a* vůči sobě je flexibilní i v porovnání s jinými Cas12a orthology, i když REC lalok jako takový je i u ostatních velmi pružný (Stella *et al.*, 2018). Navázání DNA vede k rotaci REC domény tak, že REC linker, respektive oblast mezi Leu261 a Ser273 spojující REC1 a REC2, směřuje k heteroduplexu RNA-DNA, a dále k rotaci PI domény o 38° , čímž se dostává blíže k PAM sekvenci na DNA. Štěpení obou řetězců DNA probíhá v katalytickém místě RuvC domény, které je tvořené konzervovanými zbytky Asp815, Glu906, and Asp1161. Podílí se na něm i Arg1124 Nuc domény (Jianwei *et al.*, 2023b).

4.2.1.3.2 Využití *LbCas12a*

Oproti *SpCas9* s délkou 1368 AMK a molekulární vahou 158,4 kDa je *LbCas12a* se svými 1228 AMK a 143,7 kDa menší protein (The UniProt Consortium, 2008), který navíc potřebuje kratší naváděcí crRNA a pro její maturaci nevyžaduje tracrRNA – celková guide RNA je tedy kratší. Fakt, že sama zpracovává pre-crRNA svou RNázovou aktivitou, navíc umožňuje využití jednoho transkriptu s více crRNA geny najednou. Cas12a proteiny obecně nacházejí využití jako nástroj k editaci genomu rostlin. Na rýži bylo demonstrováno, že *LbCas12a* v porovnání s Cas9, která generuje indelové mutace o velikosti 1-2 bp (Zhang *et al.*, 2014), vytváří podstatně delší delece v 3' cílové sekvenci. Kromě toho nepůsobí v oblastech mimo cílové sekvence (tzv. off-targets), popřípadě je její působení minimální (Hu *et al.*, 2017; Wang *et al.*, 2017; Xu *et al.*, 2017), zatímco pro Cas9 je tvorba těchto mutací běžná (Josephs *et al.*, 2015). S vyšší zaznamenanou efektivitou než Cas9 edituje *LbCas12a* rostlinný genom

včetně hospodářsky významných plodin jako jsou bavlna (Li *et al.*, 2019), rajče (Vu *et al.*, 2020) či citrusy, respektive grapefruit (Jia, Orbović and Wang, 2019). Další využití najde *LbCas12a* například v diagnostice, kde se uplatňuje její schopnost nesespecificky štěpit ssDNA (Chen *et al.*, 2018; Broughton *et al.*, 2020).

Existují i uměle vytvořené formy jako *LbCas12a-RR* a *LbCas12a-RVR* navržené k rozpoznání různých PAM sekvencí, což značně zvyšuje počet proveditelných mutací proteinem mimo jiné i v savčích buňkách, kde za normálních okolností nemá příliš vysokou efektivitu a přirozeně nachází méně cílových PAM (Gao *et al.*, 2017; Zhong *et al.*, 2018). Další variantou jsou *ttLbCas12a* a *enLbCas12a* odolnější vůči nižším teplotám, které jsou pro kultivaci rostlin nezbytné, nicméně významně snižují enzymatickou aktivitu proteinu. Obě varianty vznikly v reakci na práci Kleinstiver *et al.*, 2019, jejíž autoři substitucí tří AMK vytvořili *enAsCas12a* tolerantnější k nižším teplotám. I přes tuto modifikaci je však *enAsCas12a* k nižším teplotám citlivější než neupravená *LbCas12a*. Substitucí D156, G532 and K538 *LbCas12a*, které korespondují s AMK substituovanými při tvorbě *enAsCas12a*, byla vytvořena *enLbCas12a* a jedinou substitucí D156 *LbCas12a* za R pak *ttLbCas12a*, která vykazuje vysokou aktivitu oproti *LbCas12a* i *enLbCas12a* už při 22 °C (Schindele and Puchta, 2020). Ještě odolnější a efektivnější je pak *ttLbCas12a Ultra* nesoucí kromě substituce D156 za R také změnu E795 na L (Xin *et al.*, 2024).

4.3 Cas9

Endonukleáza Cas9 je součástí CRISPR/Cas systémů typu II a je jedním z nejužívanějších Cas proteinů pro editaci savčího genomu (Xu and Li, 2020). Poprvé byla tato nukleáza popsána již v roce 2005, v té době její autoři článek nazvali Cas5 (Bolotin *et al.*, 2005). Než se ustálilo označení Cas9, byla titulována také jako Csn1 či Csx12 (Deltcheva *et al.*, 2011).

4.3.1 Využití a inovace CRISPR/Cas9

Mechanismus CRISPR/Cas9 se používá pro umlčení (knock-out) nebo naopak vkládání nových genů (knock-in) do DNA prokaryotních i eukaryotních organismů. Jsou tak například vytvářeny transgenní organismy (Kyrou *et al.*, 2018; Waltz, 2021) a uplatní se i v biomedicině při diagnostice, léčbě srpkovité anémie či β -talasémie (Frangoul *et al.*, 2021) nebo při výzkumu léčby rakoviny (Wen *et al.*, 2020).

V genovém inženýrství je využíváno i mnoho modifikovaných verzí Cas9. Například katalyticky inaktivovaná, respektive dead Cas9 (dCas9) získaná mutacemi v RuvC a HNH

doménách, které způsobují ztrátu nukleázové aktivity proteinu, zatímco schopnost vázat specifický úsek DNA zůstává zachována (Qi *et al.*, 2013). Fúzí dCas9 s jedním nebo i více transkripčními aktivátory, popřípadě represory, je možné získat nástroj k aktivaci či inhibici exprese daných genů. Používá se i k zavádění epigenetických modifikací nebo k označení specifických sekvencí fluorescenčními proteiny (Dominguez, Lim and Qi, 2016).

Další modifikace Cas9 například rozšiřují škálu PAM sekvencí, které je systém schopen rozpoznat, což zvyšuje jeho využitelnost (Hu *et al.*, 2018; Nishimasu *et al.*, 2018; Walton *et al.*, 2020), v některých případech však mohou tyto modifikace způsobovat i zvýšení frekvence působení mimo cílové sekvence (Wu *et al.*, 2022). Takzvaný prime editing využívá fúzi Cas9 s narušenou katalytickou funkcí s reverzní transkriptázou. Jako naváděcí RNA slouží prime editing RNA (pegRNA), která zároveň kóduje změnu, jež má být zavedena do genomu. Jde o přesnější metodu, která navíc nevyžaduje DNA templát ani nevytváří dvouvláknové zlomy, ne vždy je ale dostatečně efektivní (Anzalone *et al.*, 2019). SuperFi-Cas9 rovněž vykazuje vysokou přesnost, její aktivita je však silně redukována (Kulcsár *et al.*, 2022). Mechanismem, modifikacemi a obecnou aplikací systému CRISPR/Cas9 se blíže zabývá review Wang and Doudna, 2023, použitím v rostlinách se pak podrobně zabývá práce Příbylová and Fischer, 2024.

5. Diskuse

Systémy TIGR-Tas vykazují pro využití v genovém inženýrství mnoho atraktivních vlastností. V porovnání s *SpCas9* (1368 AMK) jsou velmi malé (okolo 300 AMK). Nepotřebují PAM sekvence, což jim umožňuje cílit na větší množství oblastí v DNA. Jejich vysoká specifita je zajištěna vazbou spacerů A a B na jejich komplementární řetězec dsDNA, popřípadě ssDNA, pokud obsahuje sekvenci komplementární k jednomu ze spacerů. Tato vazba je spíše netolerantní k chybnému párování bází. Správně zavedenou mutací lze indukovat nikázovou aktivitu systému. Velkou nevýhodou TIGR-Tas, kterou sdílí i s bridge RNA, je však jeho neznámost. V současné době existuje pouze jediný vědecký článek zabývající se tímto systémem. Jeho autoři sice demonstrovali, že TIGR-Tas může fungovat i v lidských buňkách, i když s nepříliš vysokou efektivitou, je to ale jediný důkaz, který momentálně máme.

Podobná je i situace bridge RNA, která byla zatím aplikována pouze při editaci genomu prokaryot, a tudíž nevíme, s jakou efektivitou bude v eukaryotních buňkách schopná

fungovat. Mechanismus bridge RNA slibuje modifikaci genomu prostřednictvím excize, inverze nebo inserce. Výhodou efektorového systému IS621 je jeho jednoduchost: tvoří jej pouze protein dlouhý 326 AMK a 1 molekula nekódující bridge RNA. K rozpoznávání cílů navíc nepotřebuje PAM sekvence, ale využívá místo toho komplementarity mezi DBL/TBL a donorovou/cílovou DNA, což mu dává vyšší flexibilitu. Velmi významná je pro genové inženýrství schopnost bridge RNA editovat genom bez vytvoření dvouvláknových zlomů. Nedojde tedy ke vzniku nechtěných mutací způsobených chybnými opravami zlomů ani k vložení aditivních sekvencí. Kromě toho dokáže provést přesnou rekombinaci mezi dvěma sekvencemi DNA, což umožňuje vložení úseku DNA přímo do definovaného místa. Obecně má bridge RNA velmi nízké off-target působení a pro ještě vyšší přesnost je navíc možné komplementární sekvence na DBL/TBL prodloužit, kromě toho lze obě smyčky také nezávisle na sobě přeprogramovat. Není však jasné, jaká může být maximální délka uměle vytvořených komplementárních sekvencí a jestli bude dostatečná k tomu, aby byla cílová místa působení efektivně rozpoznána i v komplikovanějších genomech, kde je výskyt shodných sekvencí dlouhých méně než deset nukleotidů pochopitelně pravděpodobnější. To by posléze mohlo vést ke zvýšení frekvence působení na jiných místech, než bylo původně zamýšleno. Další překážkou by mohla být horší dostupnost eukaryotické DNA kvůli její interakci s histonovými proteiny. Pokud by však bridge RNA byla schopná fungovat v eukaryotních buňkách, využití by mohla mít například při léčbě nemocí způsobených rozšířením repetitivních sekvencí, jako jsou Huntingtonova choroba, amyotrofická laterální skleróza, syndrom fragilního X či myotonická dystrofie. Výzkum zabývající se jejím využitím pro léčbu nemocí centrální nervové soustavy dokonce již probíhá (Li *et al.*, 2025). Teoretizuje se i s využitím bridge RNA v rostlinách (Vollen, Alonso and Stepanova, 2025), kde by se mohla uplatnit také její schopnost modifikovat mnohem delší segmenty DNA, než je možné pomocí CRISPR/Cas systémů. Mohla by tak pomoci například při editaci genomu rýže, která v současné době tvoří téměř čtvrtinu kalorického příjmu veškeré populace. Pro zajištění potravinové bezpečnosti je tedy vývoj odrůd odolnějších vůči biotickým i abiotickým stresorům více než žádoucí (Ramasamy *et al.*, 2025).

Cas12a je protein o délce mezi cca 1100 až 1300 AMK, který má dvojí enzymatickou aktivitu: DNázovou pro štěpení cílové DNA a RNázovou pro vlastní zpracování pre-crRNA. K maturaci crRNA navíc není potřeba žádná další nekódující RNA. Tím se liší od Cas9, jejíž crRNA prekurzor je štěpen hostitelskou RNázou III a k maturaci vyžaduje tracrRNA. Díky tomu je celková délka naváděcí RNA Cas9 i více než dvojnásobná, přestože délka spacerů je

u obou srovnatelná. Dalším rozdílem jsou rozpoznávané PAM sekvence. *SpCas9* rozpoznává G bohaté PAM, naopak *Cas12a* preferuje T bohaté PAM. Štěpením DNA prostřednictvím *Cas12a* vznikají kohezní (lepivé) konce s přesahem 4-5 nukleotidů, což usnadňuje tvorbu nových vazeb, zatímco *Cas9* obvykle generuje konce tupé. I poloha zlomů vzhledem k PAM sekvenci se liší: u *Cas12a* vznikají 18-23 bp downstream od ní, zatímco u *SpCas9* je to 3 bp upstream. *Cas12a* má navíc schopnost nespecificky štěpit jednovláknovou DNA, čehož se využívá v diagnostice pro detekci DNA virů či nádorových buněk.

6. Závěr

Mechanismy TIGR-Tas a bridge RNA představují nové možné způsoby editace genomu, jejich využití v praxi je zatím však pouze teoretické. CRISPR/Cas systémy oproti nim mají obrovskou výhodu v tom, že jsou velmi prozkoumané a existuje pro ně mnoho pokročilých modifikací, které jim umožňují překonat jejich přirozené nedostatky. Dovolují jim například rozpoznat jiné PAM sekvence nebo editovat genom bez zavádění dvouvláknových zlomů. Je pochopitelné, že mnoho výzkumníků dá raději přednost modifikovaným CRISPR/Cas systémům než neznámým TIGR-Tas a bridge RNA, a proto se jejich širší aplikace v dohledné době s největší pravděpodobností nedočkáme. Bude nicméně zajímavé sledovat jejich další výzkum paralelně se vznikem nových CRISPR/Cas modifikací a do jaké míry budou schopné CRISPR/Cas systémům konkurovat.

7. Seznam použité literatury

Aittaleb, M. *et al.* (2003) 'Structure and function of archaeal box C/D sRNP core proteins', *Nature Structural & Molecular Biology*, 10(4), pp. 256–263. Available at: <https://doi.org/10.1038/nsb905>.

Anzalone, A.V. *et al.* (2019) 'Search-and-replace genome editing without double-strand breaks or donor DNA', *Nature*, 576(7785), pp. 149–157. Available at: <https://doi.org/10.1038/s41586-019-1711-4>.

Azeez, S.S. *et al.* (2024) 'Advances in CRISPR-Cas technology and its applications: revolutionising precision medicine', *Frontiers in Genome Editing*, 6, p. 1509924. Available at: <https://doi.org/10.3389/fgeed.2024.1509924>.

Barrangou, R. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712. Available at: <https://doi.org/10.1126/science.1138140>.

- Becker, S. and Boch, J. (2021) 'TALE and TALEN genome editing technologies', *Gene and Genome Editing*, 2, p. 100007. Available at: <https://doi.org/10.1016/j.ggedit.2021.100007>.
- Bolotin, A. *et al.* (2005) 'Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin', *Microbiology*, 151(8), pp. 2551–2561. Available at: <https://doi.org/10.1099/mic.0.28048-0>.
- Broughton, J.P. *et al.* (2020) 'CRISPR–Cas12-based detection of SARS-CoV-2', *Nature Biotechnology*, 38(7), pp. 870–874. Available at: <https://doi.org/10.1038/s41587-020-0513-4>.
- Brouns, S.J.J. *et al.* (2008) 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes', *Science*, 321(5891), pp. 960–964. Available at: <https://doi.org/10.1126/science.1159689>.
- Buchner, J.M. *et al.* (2005) 'Piv Site-Specific Invertase Requires a DEDD Motif Analogous to the Catalytic Center of the RuvC Holliday Junction Resolvases', *Journal of Bacteriology*, 187(10), pp. 3431–3437. Available at: <https://doi.org/10.1128/jb.187.10.3431-3437.2005>.
- Carroll, D. (2011) 'Genome Engineering With Zinc-Finger Nucleases', *Genetics*, 188(4), pp. 773–782. Available at: <https://doi.org/10.1534/genetics.111.131433>.
- Chen, J.S. *et al.* (2018) 'CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity', *Science*, 360(6387), pp. 436–439. Available at: <https://doi.org/10.1126/science.aar6245>.
- Choi, S., Ohta, S. and Ohtsubo, E. (2003) 'A Novel IS Element, IS621, of the IS110/IS492 Family Transposes to a Specific Site in Repetitive Extragenic Palindromic Sequences in *Escherichia coli*', *Journal of Bacteriology*, 185(16), pp. 4891–4900. Available at: <https://doi.org/10.1128/jb.185.16.4891-4900.2003>.
- Deltcheva, E. *et al.* (2011) 'CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III', *Nature*, 471(7340), pp. 602–607. Available at: <https://doi.org/10.1038/nature09886>.
- Dominguez, A.A., Lim, W.A. and Qi, L.S. (2016) 'Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation', *Nature Reviews Molecular Cell Biology*, 17(1), pp. 5–15. Available at: <https://doi.org/10.1038/nrm.2015.2>.
- Dong, D. *et al.* (2016) 'The crystal structure of Cpf1 in complex with CRISPR RNA', *Nature*, 532(7600), pp. 522–526. Available at: <https://doi.org/10.1038/nature17944>.
- Durrant, M.G. *et al.* (2024) 'Bridge RNAs direct programmable recombination of target and donor DNA', *Nature*, 630(8018), pp. 984–993. Available at: <https://doi.org/10.1038/s41586-024-07552-4>.
- Faure, G. *et al.* (2025) 'TIGR-Tas: A family of modular RNA-guided DNA-targeting systems in prokaryotes and their viruses', *Science*, 0(0), p. eadv9789. Available at: <https://doi.org/10.1126/science.adv9789>.
- Fonfara, I. *et al.* (2016) 'The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA', *Nature*, 532(7600), pp. 517–521. Available at: <https://doi.org/10.1038/nature17945>.

- Frangoul, H. *et al.* (2021) ‘CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia’, *New England Journal of Medicine*, 384(3), pp. 252–260. Available at: <https://doi.org/10.1056/NEJMoa2031054>.
- Gao, L. *et al.* (2017) ‘Engineered Cpf1 variants with altered PAM specificities’, *Nature Biotechnology*, 35(8), pp. 789–792. Available at: <https://doi.org/10.1038/nbt.3900>.
- Gao, P. *et al.* (2016) ‘Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition’, *Cell Research*, 26(8), pp. 901–913. Available at: <https://doi.org/10.1038/cr.2016.88>.
- Gasiunas, G. *et al.* (2012) ‘Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria’, *Proceedings of the National Academy of Sciences*, 109(39), pp. E2579–E2586. Available at: <https://doi.org/10.1073/pnas.1208507109>.
- Gorbalenya, A.E. (1994) ‘Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family’, *Protein Science*, 3(7), pp. 1117–1120. Available at: <https://doi.org/10.1002/pro.5560030716>.
- Haft, D.H. *et al.* (2005) ‘A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes’, *PLOS Computational Biology*, 1(6), p. e60. Available at: <https://doi.org/10.1371/journal.pcbi.0010060>.
- Harrington, L.B. *et al.* (2018) ‘Programmed DNA destruction by miniature CRISPR-Cas14 enzymes’, *Science*, 362(6416), pp. 839–842. Available at: <https://doi.org/10.1126/science.aav4294>.
- Harwati, T.U. *et al.* (2009) ‘*Tropicimonas isoalkanivorans* gen. nov., sp. nov., a branched-alkane-degrading bacterium isolated from Semarang Port in Indonesia’, *International Journal of Systematic and Evolutionary Microbiology*, 59(2), pp. 388–391. Available at: <https://doi.org/10.1099/ijs.0.65822-0>.
- Hiraizumi, M. *et al.* (2024) ‘Structural mechanism of bridge RNA-guided recombination’, *Nature*, 630(8018), pp. 994–1002. Available at: <https://doi.org/10.1038/s41586-024-07570-2>.
- Hirsch, H.-J., Starlinger, P. and Brachet, P. (1972) ‘Two kinds of insertions in bacterial genes’, *Molecular and General Genetics MGG*, 119(3), pp. 191–206. Available at: <https://doi.org/10.1007/BF00333858>.
- Hochstein, R.A. *et al.* (2016) ‘Acidianus Tailed Spindle Virus: a New Archaeal Large Tailed Spindle Virus Discovered by Culture-Independent Methods’, *Journal of Virology*, 90(7), pp. 3458–3468. Available at: <https://doi.org/10.1128/jvi.03098-15>.
- Hu, J.H. *et al.* (2018) ‘Evolved Cas9 variants with broad PAM compatibility and high DNA specificity’, *Nature*, 556(7699), pp. 57–63. Available at: <https://doi.org/10.1038/nature26155>.
- Hu, X. *et al.* (2017) ‘Targeted mutagenesis in rice using CRISPR-Cpf1 system’, *Journal of Genetics and Genomics*, 44(1), pp. 71–73. Available at: <https://doi.org/10.1016/j.jgg.2016.12.001>.
- Ishino, Y. *et al.* (1987) ‘Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product’,

- Journal of Bacteriology*, 169(12), pp. 5429–5433. Available at: <https://doi.org/10.1128/jb.169.12.5429-5433.1987>.
- Iwasaki, H. *et al.* (1991) ‘Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure.’, *The EMBO Journal*, 10(13), pp. 4381–4389.
- Josephs, E.A. *et al.* (2015) ‘Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage’, *Nucleic Acids Research*, 43(18), pp. 8924–8941. Available at: <https://doi.org/10.1093/nar/gkv892>.
- Jansen, Ruud. *et al.* (2002) ‘Identification of genes that are associated with DNA repeats in prokaryotes’, *Molecular Microbiology*, 43(6), pp. 1565–1575. Available at: <https://doi.org/10.1046/j.1365-2958.2002.02839.x>.
- Jia, H., Orbović, V. and Wang, N. (2019) ‘CRISPR-LbCas12a-mediated modification of citrus’, *Plant Biotechnology Journal*, 17(10), pp. 1928–1937. Available at: <https://doi.org/10.1111/pbi.13109>.
- Jianwei, L. *et al.* (2023a) ‘Structures of apo Cas12a and its complex with crRNA and DNA reveal the dynamics of ternary complex formation and target DNA cleavage’, *PLOS Biology*, 21(3), p. e3002023. Available at: <https://doi.org/10.1371/journal.pbio.3002023>.
- Jianwei, L. *et al.* (2023b) ‘Structures of apo Cas12a and its complex with crRNA and DNA reveal the dynamics of ternary complex formation and target DNA cleavage’, *PLOS Biology*, 21(3), p. e3002023. Available at: <https://doi.org/10.1371/journal.pbio.3002023>.
- Jinek, M. *et al.* (2012) ‘A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity’, *Science*, 337(6096), pp. 816–821. Available at: <https://doi.org/10.1126/science.1225829>.
- Kim, R.S. *et al.* (2025) ‘BFVD—a large repository of predicted viral protein structures’, *Nucleic Acids Research*, 53(D1), pp. D340–D347. Available at: <https://doi.org/10.1093/nar/gkae1119>.
- Kleinstiver, B.P. *et al.* (2019) ‘Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing’, *Nature Biotechnology*, 37(3), pp. 276–282. Available at: <https://doi.org/10.1038/s41587-018-0011-0>.
- Kulcsár, P.I. *et al.* (2022) ‘SuperFi-Cas9 exhibits remarkable fidelity but severely reduced activity yet works effectively with ABE8e’, *Nature Communications*, 13(1), p. 6858. Available at: <https://doi.org/10.1038/s41467-022-34527-8>.
- Kunin, V., Sorek, R. and Hugenholtz, P. (2007) ‘Evolutionary conservation of sequence and secondary structures in CRISPR repeats’, *Genome Biology*, 8(4), p. R61. Available at: <https://doi.org/10.1186/gb-2007-8-4-r61>.
- Kyrou, K. *et al.* (2018) ‘A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes’, *Nature Biotechnology*, 36(11), pp. 1062–1066. Available at: <https://doi.org/10.1038/nbt.4245>.
- Li, B. *et al.* (2019) ‘Robust CRISPR/Cpf1 (Cas12a)-mediated genome editing in allotetraploid cotton (Gossypium hirsutum)’, *Plant Biotechnology Journal*, 17(10), pp. 1862–1864. Available at: <https://doi.org/10.1111/pbi.13147>.

- Li, F. *et al.* (2025) ‘Bridge RNA-Guided Genetic Recombination Tools for Treating Neurodegenerative Nucleotide Repeat Disorders’, *Neuroscience Bulletin* [Preprint]. Available at: <https://doi.org/10.1007/s12264-025-01358-8>.
- Liu, S. *et al.* (2007) ‘Binding of the Human Prp31 Nop Domain to a Composite RNA-Protein Platform in U4 snRNP’, *Science*, 316(5821), pp. 115–120. Available at: <https://doi.org/10.1126/science.1137924>.
- Mahillon, J. and Chandler, M. (1998) ‘Insertion Sequences’, *Microbiology and Molecular Biology Reviews*, 62(3), pp. 725–774. Available at: <https://doi.org/10.1128/mubr.62.3.725-774.1998>.
- Makarova, K.S. *et al.* (2020) ‘Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants’, *Nature Reviews Microbiology*, 18(2), pp. 67–83. Available at: <https://doi.org/10.1038/s41579-019-0299-x>.
- Makarova, K.S. and Koonin, E.V. (2015) ‘Annotation and Classification of CRISPR-Cas Systems’, in M. Lundgren, E. Charpentier, and P.C. Fineran (eds) *CRISPR*. New York, NY: Springer New York (Methods in Molecular Biology), pp. 47–75. Available at: https://doi.org/10.1007/978-1-4939-2687-9_4.
- Mojica, F.J.M. *et al.* (2005) ‘Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements’, *Journal of Molecular Evolution*, 60(2), pp. 174–182. Available at: <https://doi.org/10.1007/s00239-004-0046-3>.
- Mojica, F.J.M. *et al.* (2009) ‘Short motif sequences determine the targets of the prokaryotic CRISPR defence system’, *Microbiology*, 155(3), pp. 733–740. Available at: <https://doi.org/10.1099/mic.0.023960-0>.
- Nishimasu, H. *et al.* (2018) ‘Engineered CRISPR-Cas9 nuclease with expanded targeting space’, *Science*, 361(6408), pp. 1259–1262. Available at: <https://doi.org/10.1126/science.aas9129>.
- Partridge, S.R. and Hall, R.M. (2003) ‘The IS1111 Family Members IS4321 and IS5075 Have Subterminal Inverted Repeats and Target the Terminal Inverted Repeats of Tn21 Family Transposons’, *Journal of Bacteriology*, 185(21), pp. 6371–6384. Available at: <https://doi.org/10.1128/jb.185.21.6371-6384.2003>.
- Pausch, P. *et al.* (2020) ‘CRISPR-Cas Φ from huge phages is a hypercompact genome editor’, *Science*, 369(6501), pp. 333–337. Available at: <https://doi.org/10.1126/science.abb1400>.
- Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) ‘CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies’, *Microbiology*, 151(3), pp. 653–663. Available at: <https://doi.org/10.1099/mic.0.27437-0>.
- Přibyllová, A. and Fischer, L. (2024) ‘How to use CRISPR/Cas9 in plants: from target site selection to DNA repair’, *Journal of Experimental Botany*, 75(17), pp. 5325–5343. Available at: <https://doi.org/10.1093/jxb/erae147>.

- Qi, L.S. *et al.* (2013) ‘Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression’, *Cell*, 152(5), pp. 1173–1183. Available at: <https://doi.org/10.1016/j.cell.2013.02.022>.
- Ramasamy, S.P. *et al.* (2025) ‘Genome editing for improving disease resistance in rice (*Oryza sativa* L.) - Scope, achievements, limitations and future needs’, *Physiological and Molecular Plant Pathology*, 138, p. 102640. Available at: <https://doi.org/10.1016/j.pmpp.2025.102640>.
- Razavi, M. *et al.* (2020) ‘The Association between Insertion Sequences and Antibiotic Resistance Genes’, *mSphere*, 5(5), p. 10.1128/msphere.00418-20. Available at: <https://doi.org/10.1128/msphere.00418-20>.
- Rice, P.A. *et al.* (1996) ‘Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn’, *Cell*, 87(7), pp. 1295–1306. Available at: [https://doi.org/10.1016/S0092-8674\(00\)81824-3](https://doi.org/10.1016/S0092-8674(00)81824-3).
- Richardson, L. *et al.* (2023) ‘MGnify: the microbiome sequence data analysis resource in 2023’, *Nucleic Acids Research*, 51(D1), pp. D753–D759. Available at: <https://doi.org/10.1093/nar/gkac1080>.
- Russell, W.C. *et al.* (1977) ‘Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5’, *Journal of General Virology*, 36(1), pp. 59–72. Available at: <https://doi.org/10.1099/0022-1317-36-1-59>.
- Schindele, P. and Puchta, H. (2020) ‘Engineering CRISPR/LbCas12a for highly efficient, temperature-tolerant plant gene editing’, *Plant Biotechnology Journal*, 18(5), pp. 1118–1120. Available at: <https://doi.org/10.1111/pbi.13275>.
- Schneider, D. and Lenski, R.E. (2004) ‘Dynamics of insertion sequence elements during experimental evolution of bacteria’, *Research in Microbiology*, 155(5), pp. 319–327. Available at: <https://doi.org/10.1016/j.resmic.2003.12.008>.
- Schunder, E. *et al.* (2013) ‘First indication for a functional CRISPR/Cas system in *Francisella tularensis*’, *International Journal of Medical Microbiology*, 303(2), pp. 51–60. Available at: <https://doi.org/10.1016/j.ijmm.2012.11.004>.
- Shmakov, S. *et al.* (2015) ‘Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems’, *Molecular Cell*, 60(3), pp. 385–397. Available at: <https://doi.org/10.1016/j.molcel.2015.10.008>.
- Shmakov, S. *et al.* (2017) ‘Diversity and evolution of class 2 CRISPR–Cas systems’, *Nature Reviews Microbiology*, 15(3), pp. 169–182. Available at: <https://doi.org/10.1038/nrmicro.2016.184>.
- Siguier, P. *et al.* (2006) ‘ISfinder: the reference centre for bacterial insertion sequences’, *Nucleic Acids Research*, 34(suppl_1), pp. D32–D36. Available at: <https://doi.org/10.1093/nar/gkj014>.
- Siguier, P. *et al.* (2015) ‘Everyman’s Guide to Bacterial Insertion Sequences’, *Microbiology Spectrum*, 3(2), p. 10.1128/microbiolspec.mdna3-0030-2014. Available at: <https://doi.org/10.1128/microbiolspec.mdna3-0030-2014>.

- Siguier, P., Goubeyre, E. and Chandler, M. (2014) 'Bacterial insertion sequences: their genomic impact and diversity', *FEMS Microbiology Reviews*, 38(5), pp. 865–891. Available at: <https://doi.org/10.1111/1574-6976.12067>.
- Singh, D. *et al.* (2018) 'Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a)', *Proceedings of the National Academy of Sciences*, 115(21), pp. 5444–5449. Available at: <https://doi.org/10.1073/pnas.1718686115>.
- Stella, S. *et al.* (2018) 'Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity', *Cell*, 175(7), pp. 1856-1871.e21. Available at: <https://doi.org/10.1016/j.cell.2018.10.045>.
- Stella, S., Alcón, P. and Montoya, G. (2017a) 'Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage', *Nature*, 546(7659), pp. 559–563. Available at: <https://doi.org/10.1038/nature22398>.
- Stella, S., Alcón, P. and Montoya, G. (2017b) 'Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage', *Nature*, 546(7659), pp. 559–563. Available at: <https://doi.org/10.1038/nature22398>.
- Sun, Y. *et al.* (2024) 'Engineer and split an efficient hypercompact CRISPR–CasΦ genome editor in plants', *Plant Communications*, 5(7). Available at: <https://doi.org/10.1016/j.xplc.2024.100881>.
- Surtees, J.A. and Funnell, B.E. (2003) 'Plasmid and Chromosome Traffic Control: How ParA and ParB Drive Partition', in *Current Topics in Developmental Biology*. Academic Press, pp. 145–180. Available at: [https://doi.org/10.1016/S0070-2153\(03\)01010-X](https://doi.org/10.1016/S0070-2153(03)01010-X).
- Swarts, D.C., Oost, J. van der and Jinek, M. (2017) 'Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a', *Molecular Cell*, 66(2), pp. 221-233.e4. Available at: <https://doi.org/10.1016/j.molcel.2017.03.016>.
- The UniProt Consortium (2008) 'The Universal Protein Resource (UniProt)', *Nucleic Acids Research*, 36(suppl_1), pp. D190–D195. Available at: <https://doi.org/10.1093/nar/gkm895>.
- Tuncel, A. *et al.* (2025) 'CRISPR–Cas applications in agriculture and plant research', *Nature Reviews Molecular Cell Biology*, pp. 1–23. Available at: <https://doi.org/10.1038/s41580-025-00834-3>.
- Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1), pp. D439–D444. Available at: <https://doi.org/10.1093/nar/gkab1061>.
- Vaysset, H. *et al.* (2025) 'Evolutionary origins of archaeal and eukaryotic RNA-guided RNA modification in bacterial IS110 transposons', *Nature Microbiology*, 10(1), pp. 20–27. Available at: <https://doi.org/10.1038/s41564-024-01889-2>.
- Vollen, K., Alonso, J.M. and Stepanova, A.N. (2025) 'Beyond a few bases: methods for large DNA insertion and gene targeting in plants', *The Plant Journal*, 121(6), p. e70099. Available at: <https://doi.org/10.1111/tpj.70099>.

- Vu, T.V. *et al.* (2020) ‘Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato’, *Plant Biotechnology Journal*, 18(10), pp. 2133–2143. Available at: <https://doi.org/10.1111/pbi.13373>.
- Walton, R.T. *et al.* (2020) ‘Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants’, *Science*, 368(6488), pp. 290–296. Available at: <https://doi.org/10.1126/science.aba8853>.
- Waltz, E. (2021) ‘GABA-enriched tomato is first CRISPR-edited food to enter market’, *Nature Biotechnology*, 40(1), pp. 9–11. Available at: <https://doi.org/10.1038/d41587-021-00026-2>.
- Wang, J.Y. and Doudna, J.A. (2023) ‘CRISPR technology: A decade of genome editing is only the beginning’, *Science*, 379(6629), p. eadd8643. Available at: <https://doi.org/10.1126/science.add8643>.
- Wang, M. *et al.* (2017) ‘Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System’, *Molecular Plant*, 10(7), pp. 1011–1013. Available at: <https://doi.org/10.1016/j.molp.2017.03.001>.
- Wen, L. *et al.* (2020) ‘CRISPR/Cas9-Mediated TERT Disruption in Cancer Cells’, *International Journal of Molecular Sciences*, 21(2), p. 653. Available at: <https://doi.org/10.3390/ijms21020653>.
- Wu, Y. *et al.* (2022) ‘Genome-wide analyses of PAM-relaxed Cas9 genome editors reveal substantial off-target effects by ABE8e in rice’, *Plant Biotechnology Journal*, 20(9), pp. 1670–1682. Available at: <https://doi.org/10.1111/pbi.13838>.
- Xin, C. *et al.* (2024) ‘Enhanced editing efficiency in Arabidopsis with a LbCas12a variant harboring D156R and E795L mutations’, *ABIOTECH*, 5(2), pp. 117–126. Available at: <https://doi.org/10.1007/s42994-024-00144-w>.
- Xu, R. *et al.* (2017) ‘Generation of targeted mutant rice using a CRISPR-Cpf1 system’, *Plant Biotechnology Journal*, 15(6), pp. 713–717. Available at: <https://doi.org/10.1111/pbi.12669>.
- Xu, Y. and Li, Z. (2020) ‘CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy’, *Computational and Structural Biotechnology Journal*, 18, pp. 2401–2415. Available at: <https://doi.org/10.1016/j.csbj.2020.08.031>.
- Yamano, T. *et al.* (2016) ‘Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA’, *Cell*, 165(4), pp. 949–962. Available at: <https://doi.org/10.1016/j.cell.2016.04.003>.
- Zetsche, B. *et al.* (2015) ‘Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System’, *Cell*, 163(3), pp. 759–771. Available at: <https://doi.org/10.1016/j.cell.2015.09.038>.
- Zhang, Hui *et al.* (2014) ‘The CRISPR / Cas9 system produces specific and homozygous targeted gene editing in rice in one generation’, *Plant Biotechnology Journal*, 12(6), pp. 797–807. Available at: <https://doi.org/10.1111/pbi.12200>.

Zhong, Z. *et al.* (2018) 'Plant Genome Editing Using FnCpf1 and LbCpf1 Nucleases at Redefined and Altered PAM Sites', *Molecular Plant*, 11(7), pp. 999–1002. Available at: <https://doi.org/10.1016/j.molp.2018.03.008>.