bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sir2-domain associated short prokaryotic Argonautes provide defence against invading mobile genetic elements through NAD^+ depletion

Mindaugas Zaremba[1*], Donata Dakineviciene[1], Edvardas Golovinas[1], Edvinas Stankunas[1†], Anna Lopatina[2‡], Rotem Sorek[2], Elena Manakova[1], Audra Ruksenaite[1], Arunas Silanskas[1], Simonas Asmontas[1], Rokas Grigaitis[1], Kęstutis Timinskas[1], Česlovas Venclovas[1], Virginijus Siksnys[1*]

[1] Institute of Biotechnology, Life Sciences Center, Vilnius University, Sauletekio av. 7, LT-10257, Vilnius, Lithuania.

[2] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel.

* To whom correspondence should be addressed. Tel: +370-5-2234357; Fax: +370-5-2234367; Email: zare@ibt.lt. Correspondence may also be addressed to Virginijus Siksnys. Tel: +370-5-2234359; Fax: +370-5-2234367; Email: siksnys@ibt.lt.

The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.

Present address: †Edvinas Stankunas, Max F. Perutz Laboratories, Medical University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria.

‡Anna Lopatina, Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Center for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Argonaute (Ago) proteins are found in all three domains of life. The so-called long Agos are composed of four major domains (N, PAZ, MID, and PIWI) and contribute to RNA silencing in eukaryotes (eAgos) or defence against invading mobile genetic elements in prokaryotes (pAgos). Intriguingly, the majority (~60%) of prokaryotic Agos (pAgos) identified bioinformatically are shorter (comprised of only MID and PIWI domains) and are typically associated with Sir2, Mrr or TIR domain-containing proteins. The cellular function and mechanism of short pAgos remain enigmatic. Here, we show that short pAgos from *Geobacter sulfurreducens*, *Caballeronia cordobensis* and *Paraburkholderia graminis*, together with the NAD^+-bound Sir2-proteins form a stable heterodimeric Sir2/Ago complex that recognizes invading plasmid or phage DNA through the pAgos subunit and activates Sir2 subunit triggering the endogenous NAD^+ depletion and cell death thus preventing the propagation of invading DNA. This is the first demonstration that short Sir2-associated pAgos provide defence against phages and plasmids and underscores the diversity of mechanisms of prokaryotic Agos.

Being at the core of RNA interference eAgos are involved in the regulation of gene expression, silencing of mobile genome elements, and defence against viruses[1,2]. All eAgos use small RNA molecules as guides for target RNA recognition and are similar both structurally and mechanistically [2–5]. Monomeric eAgos are composed of four major N, PAZ, MID, and PIWI domains (Fig. 1A) and share a bilobed structure, where the N- and C-terminal lobes are formed by conserved N/PAZ and MID/PIWI domains, respectively[4–7]. The N-domain acts as a wedge that separates guide and target strands[3,7], while the MID and PAZ domains bind, respectively, the 5′- and 3′-terminus of the guide RNA, located between the N- and C-terminal lobes[4,6]. eAgos can slice target RNA through endonucleolytic cleavage by the PIWI domain or inhibit translation through RNA binding by the catalytically inactive eAgos that may also trigger RNA decay by auxiliary cellular nucleases[2,4,5].

pAgos are quite widespread and are present in 17% of sequenced bacterial and 25% of archaeal genomes[8]. To date, more than ~1000 pAgos have been identified bioinformatically, revealing a striking diversity. Moreover, pAgos are often associated with additional putative nucleases, helicases, and DNA binding proteins that are not linked to eAgos[8]. pAgos are divided into full-length or long pAgos (~40%) sharing conserved N, PAZ, MID, and PIWI domain architecture with eAgos, and short pAgos (~60%) composed only of MID and PIWI domains (Fig. 1A)[8,9]. Long pAgos are relatively well-characterized both structurally and functionally[10,11] and, similarly to eAgos, contain either catalytically active or inactive PIWI domain. Long pAgos with the catalytically active PIWI, as exemplified by CbAgo and TtAgo, guided by DNA cleave DNA target providing defence against invading phages or plasmids, or contributing to chromosome segregation after replication, respectively[12–14]. Meanwhile, long RsAgo with an inactive PIWI domain guided by small RNA is thought to mobilize an unknown cellular nuclease(s) for degradation of invading plasmids and mobile genetic elements[9,15]. In contrast to long pAgos, all short pAgos possess a catalytically inactive PIWI domain; however, they are typically associated with proteins containing an APAZ (Analog of PAZ) domain fused to predicted Sir2, Mrr nucleases or TIR domains[8,16]. Sir2-APAZ and short pAgo are sometimes fused into a single-chain protein[8]. Despite the fact that short pAgos make up the majority of all pAgos, to date, their function in the cell and *in vitro* remain to be established.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

In this study, we aimed to explore whether short pAgos can act as prokaryotic defence systems against viruses or plasmids. To this end, we selected two short pAgos, GsSir2/Ago from *Geobacter sulfurreducens* and CcSir2/Ago from *Caballeronia cordobensis*, each encoded in a putative operon together with a Sir2 domain protein, and PgSir2-Ago from *Paraburkholderia graminis*, encoded as a single-chain variant, comprised of Sir2 and pAgo domains (Fig. 1A). The coding regions of the Sir2 and pAgo proteins in GsSir2/Ago and CcSir2/Ago systems overlap by 11 and 8 bp, respectively, indicating that they belong to the same operon (Supplementary Text). Next, we engineered heterologous *E. coli* cells by cloning GsSir2/Ago and CcSir2/Ago genes, or a single gene, encoding PgSir2-Ago into pBAD expression vectors under a P_{BAD} promoter (Supplementary Table 2) and challenged them with phages or plasmids.

Sir2/Ago systems provide defence against phages. To test whether the GsSir2/Ago system provides defence against phages, we challenged *E. coli* host carrying the GsSir2/Ago system with a set of six *E. coli* phages spanning four morphological families including *Podoviridae* (T7), *Siphoviridae* (lambda-vir, SECphi27, SECphi18), *Myoviridae* (T4) and *Microviridae* (SECphi17, a ssDNA phage) (Supplementary Table 2). We measured the efficiency of plating (EOP) of these phages with and without L-arabinose induction of the GsSir2/Ago system. The system showed protection against two out of six phages – lambda-vir (~100-fold) and SECphi27 (~1000-fold) (Fig. 1C).

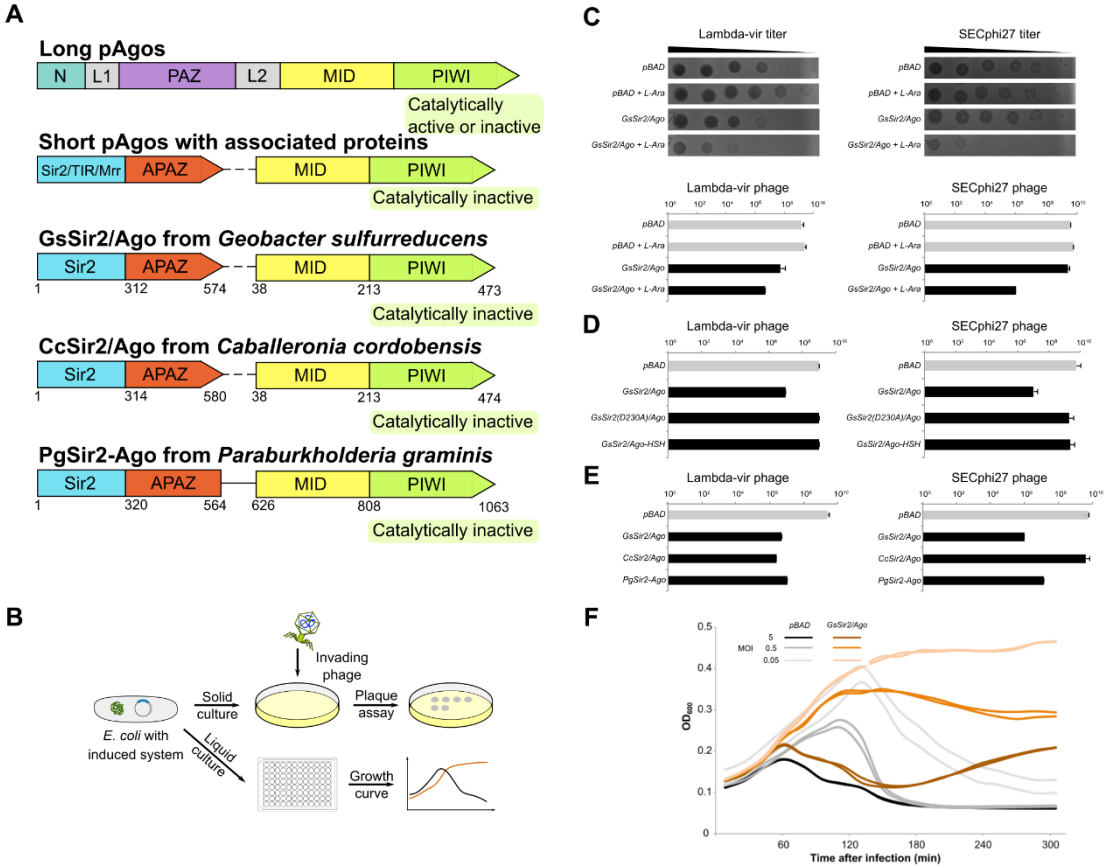


Fig. 1 Sir2/Ago provide defence against phages. A, Schematic representation of different pAgo systems. Individual domains are colour-coded. Catalytically active pAgos contain a conserved catalytic DEDX tetrad that is mutated in the inactive pAgos. Catalytically inactive short pAgos lack the canonical PAZ (PIWI-Argonaute-Zwille) domain, however,

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an APAZ (analogue of PAZ) domain is identified in putative Sir2, TIR or Mrr proteins associated with short pAgos. MID indicates middle; L – linker domain; N – N-terminal domain. Short pAgos from *Geobacter sulfurreducens*, *Caballeronia cordobensis*, and *Paraburkholderia graminis* associated with Sir2 protein were studied in this work. **B**, Schematic diagram of phage restriction assays. **C**, Efficiency of plating (EOP) of 2 phages infecting *E. coli* cells with and without the GsSir2/Ago system. The x-axis represents the number of p.f.u. Shown are the means of three replicates in the absence and in the presence of the inducer L-arabinose (L-Ara), and error bars are s.d. of the mean. Grey bars represent efficiency of plating (EOP) on pAgo-lacking cells and black bars are EOP in pAgo-containing cells. Representative images of plaque assays for lambda-vir and SECphi27 are also presented. **D**, EOP of lambda-vir and SECphi27 phages infecting *E. coli* cells with the wt and mutant GsSir2/Ago systems. Data represent p.f.u./ml in the presence of the inducer L-arabinose, an average of three replicates with error bars representing s.d. GsSir2(D230A)/Ago and GsSir2/Ago-HSH are variants that contain D230A mutation in the Sir2 domain or His_6-StrepII-His_6-tag (HSH-tag) on the C-terminus of pAgo. **E**, EOP of lambda-vir and SECphi27 phages infecting pAgo-lacking and pAgo-containing cells from *Geobacter sulfurreducens* (GsSir2/Ago), *Caballeronia cordobensis* (CcSir2/Ago), and *Paraburkholderia graminis* (PgSir2-Ago, a single-chain protein). Data represent p.f.u./ml in the presence of the inducer L-arabinose, an average of three replicates with error bars representing s.d. **F**, Lambda phage infection in liquid cultures of *E. coli* cells containing the GsSir2/Ago system. GsSir2/Ago-lacking *E. coli* MG1655 (shown in grey) or GsSir2/Ago-containing cells (shown in orange) were infected at t=0 at multiplicities of infection (MOI) of 0.05, 0.5 and 5. Each curve represents one individual replicate; two replicates for each MOI are shown. The collapse of GsSir2/Ago-containing cultures at high MOI suggests that most of the infected GsSir2/Ago-containing cells undergo abortive infection, while at low MOI a minority of cells produce viable phage progeny and the culture survives.

To probe the role of individual Sir2 and Ago proteins in antiviral defence, we engineered *E. coli* cells, carrying mutant variants of either the Sir2 or the Ago protein of the GsSir2/Ago system and performed small drop plaque assays using the lambda-vir and SECphi27 phages. In Sir2 variants, the highly-conserved D230 residue, presumably involved in NAD^+ binding was replaced by Ala residue (Supplementary Fig. 1D)[16]. Phage challenge assay revealed that the D230A mutation completely abolished defence against both phages (Fig. 1D). The Ago protein is catalytically inactive due to the active site mutations in the PIWI domain, therefore, to obtain a binding-deficient Ago variant, we fused a bulky 29 aa **H**is$_6$-**S**trepII-**H**is$_6$-tag (HSH-tag) at the C-terminus that is important for nucleic acid binding in other Agos[3,6,8,17]. We found that Ago C-terminal modification abolished protection from both phages (Fig. 1D). Thus, both the Sir2 and the Ago proteins are required for protection against phages by the GsSir2/Ago system.

Additionally, we probed the ability of homologous CcSir2/Ago and PgSir2-Ago systems to restrict lambda-vir and SECphi27 phages (Fig. 1E). After L-arabinose induction, the PgSir2-Ago system showed ~500-fold protection against lambda phage and ~400-fold against SECphi27, while CcSir2/Ago system showed ~1000-fold protection against lambda and no protection against SECphi27 phage. Thus, homologous Sir2/Ago systems from three phylogenetically distant bacteria showed protection against phage infection, albeit with different efficiency and specificity.

Next, *E. coli* MG1655 cells lacking or containing the GsSir2/Ago system were infected in liquid culture with lambda phage at a multiplicity of infection (MOI) 0.05, 0.5 and 5 (Fig. 1F). At high MOI where, on average, a single bacterial cell is infected by a single phage, the culture collapses, while at low MOI the culture survives. This phenotype implies that GsSir2/Ago-mediated defence triggers cell death approximately at the same time when the phage-induced lysis occurs.

Sir2/Ago systems interfere with plasmid transformation. Further, to test whether heterologous expression of the GsSir2/Ago, CcSir2/Ago and PgSir2-Ago systems in *E. coli* cells (strains BL21-AI and DH10B) provides a barrier for plasmid transformation, four plasmids (pCDF, pCOLA, pACYC184, and pRSF) with different *ori* regions and copy numbers were used in a plasmid interference assay (Supplementary Table 2, Fig. 2, and Supplementary Fig. 2). We found that GsSir2/Ago system prevented only pCDF plasmid transformation reducing its efficiency nearly ~100-fold (Fig. 2B). Next, we tested pCDF transformation efficiency in *E. coli* cells expressing the GsSir2/Ago mutants (Fig. 2C). Both Sir2 D230A mutation and pAgo HSH-tag modification that

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

impaired phage restriction also abolished plasmid interference making cells permissive for pCDF plasmid transformation (Fig. 2C). CcSir2/Ago system like GsSir2/Ago provided resistance only for pCDF plasmid transformation (Supplementary Fig. 2E) while cells carrying the single-chain PgSir2-Ago system was permissive for transformation of all four plasmids (Supplementary Fig. 2E).

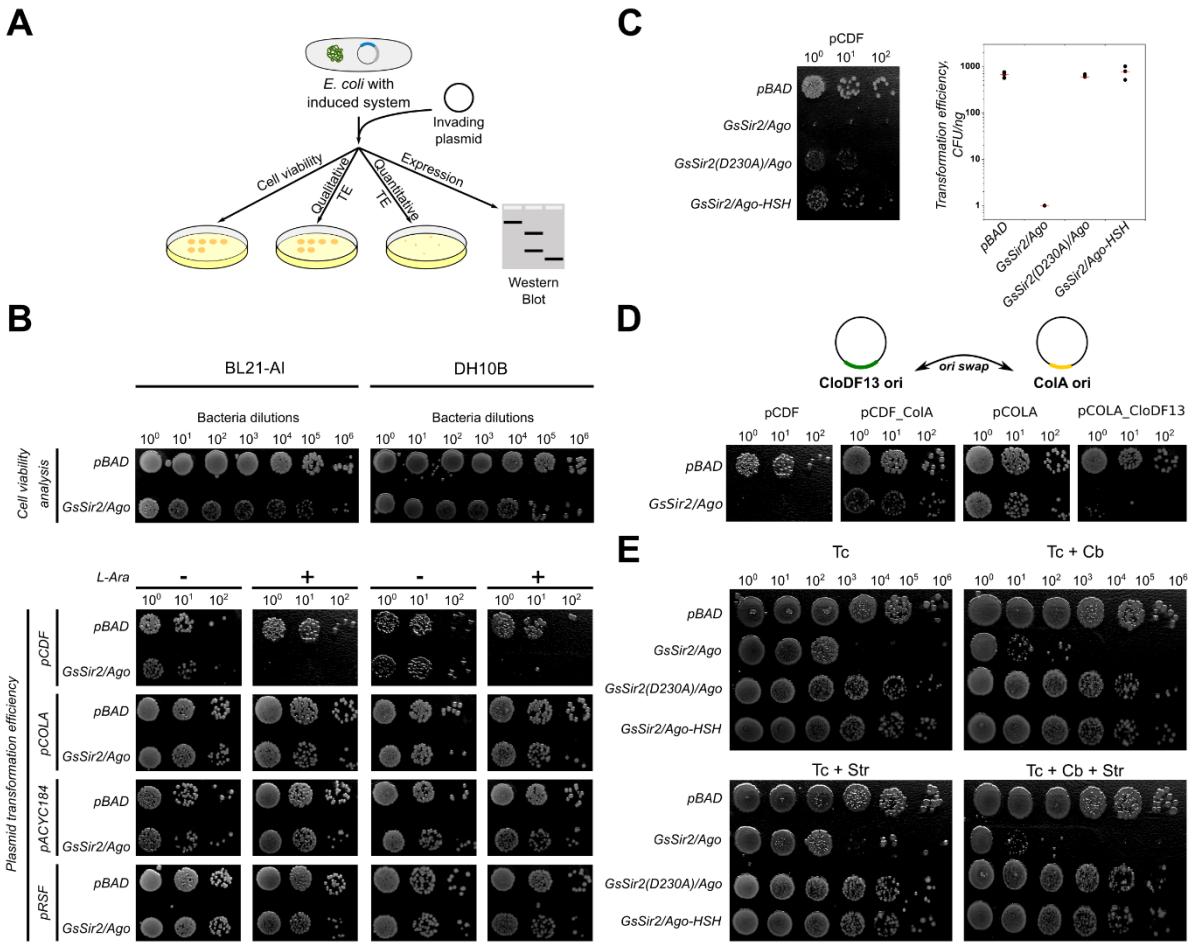


Fig. 2 The GsSir2/Ago system interfere with plasmid transformation. **A**, Schematic representation of the experiment. **B**, Qualitative evaluation of plasmid transformation efficiency in *E. coli* cells carrying GsSir2/Ago system. Top: comparison of cell viability in the presence or absence of plasmid-borne GsSir2/Ago expression. Bottom: comparison of plasmid transformation efficiencies in the presence or absence of plasmid-borne GsSir2/Ago expression. **C**, Left – comparison of pCDF transformation efficiency between cells expressing wt and mutant GsSir2/Ago complexes; right - quantification of transformation efficiencies (three independent replicates, the red line represents average transformation efficiency). **D**, Top - schematic representation of *ori* exchange between pCDF and pCOLA plasmids; bottom - comparison of plasmid transformation efficiencies. pCDF, pCDF with CloDF13 *ori* exchanged with ColA *ori* (pCDF_ColA), pCOLA and pCOLA with ColA *ori* exchanged with CloDF13 *ori* (pCOLA_CloDF13) plasmids were used for transformation of *E. coli* cells carrying GsSir2/Ago system. **E**, Cell viability in the absence of antibiotic selection. In the case of the wt GsSir2/Ago system, the cell viability decreases on the plates even in the absence of Cb and Str antibiotics, for which the resistance genes are in pBAD and pCDF plasmids, respectively. The GsSir2/Ago system causes cell death in the presence of the target (pCDF plasmid).

Interestingly, although the GsSir2/Ago system in *E. coli* interfered with the pCDF plasmid transformation, pCOLA plasmid that differs mainly in the *ori* region and antibiotic resistance gene was permissive. To test whether the *ori* region determines differences in the transformation efficiency between pCDF and pCOLA plasmids, we swapped the *ori* sequences of pCDF and pCOLA (Fig. 2D). The pCDF plasmid with ColA *ori* instead of CloDF13 became permissive in *E.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

coli cells expressing the GsSir2/Ago system, whereas transformation of pCOLA plasmid bearing CloDF13 *ori* instead of ColA *ori* was prevented. These results indicate that CloDF13 *ori* is a key element that controls plasmid transformation efficiency in *E. coli* cells expressing GsSir2/Ago system.

The plasmid interference by the GsSir2/Ago system could be due to either the plasmid entry exclusion, replication inhibition or plasmid degradation. To eliminate the possible role of the plasmid entry barriers on the pCDF plasmid transformation efficiency, we engineered heterologous *E. coli* cells carrying two plasmids: pBAD plasmid expressing GsSir2/Ago (or its mutants) under control of P_{BAD} inducible promoter and the pCDF plasmid providing antibiotic resistance, and tested cell viability in the presence or absence of inducer. In this case, the pCDF plasmid is already in the cell and provides streptomycin resistance, however antibiotic resistance should be lost if the plasmid is restricted after GsSir2/Ago expression. In the absence of induction, cell viability of *E. coli* cells carrying wt GsSir2/Ago (or its mutants) and an empty pBAD vector (Supplementary Fig. 2C) was identical. In the presence of inducer the viability of cells expressing the wt GsSir2/Ago system, but not its mutants, significantly decreased (Fig. 2E), indicating that GsSir2/Ago interferes with the pCDF plasmid already present in the cell. Notably, a decrease in cell viability is observed in *E. coli* BL21-AI cells (Tc-resistant) without the cell selection for Str (pCDF) and Cb (pBAD) resistance. It cannot be excluded that upon recognition of pCDF the GsSir2/Ago system becomes activated and triggers cell death. A similar cell death phenotype triggered by the GsSir2/Ago has been observed during phage infection in liquid cultures (Fig. 1F). Taken together, these data show that the GsSir2/Ago system acts as a defence system against phages and plasmids via cell death or suicidal mechanism.

Short pAgo and associated Sir2 protein form a stable heterodimeric complex. To characterize the Sir2/Ago systems biochemically, we aimed to express individual Sir2 and Ago proteins in *E. coli*. The GsSir2 and CcSir2 proteins (but not PgSir2-Ago) were expressed and purified by liquid chromatography (Fig. 3A and Supplementary Fig. 3). The N-terminal His_6-tagged GsSir2 and CcSir2 proteins co-expressed with Ago proteins co-purified on the Ni^{2+}-affinity column (Supplementary Fig. 3), indicating that Sir2 and pAgo proteins form a stable complex. Functionally compromised GsSir2(D230A)/Ago and GsSir2/Ago-HSH variants also formed a complex indicating that while the introduced mutations abolished the activity *in vivo*, they did not affect the protein complex structure (Fig. 1, Supplementary Fig. 3). Further analysis of the oligomeric state of GsSir2/Ago and CcSir2/Ago complexes in solution using SEC-(MALS), mass photometry and small-angle X-ray scattering (SAXS) demonstrated that the complexes are heterodimeric in the wide range of protein concentrations (from 20 nM to 6.5 μM) (Fig. 3B-D, and Supplementary Fig. 4). According to the SAXS data, the heterodimeric GsSir2/Ago complex acquires a notably asymmetric shape (Fig. 3D, Supplementary Fig. 4, Supplementary Table 3) that is consistent with the structural model of the heterodimer (Supplementary Text). In summary, the results show that pAgos and associated Sir2 proteins encoded by a single operon form a stable heterodimeric complex.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

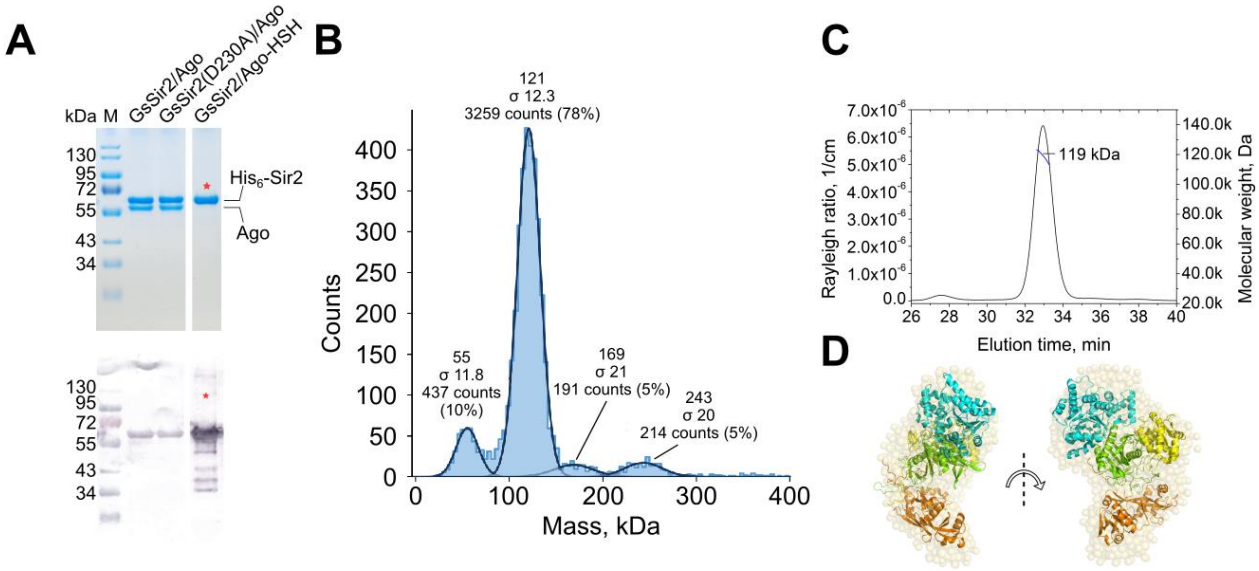


Fig. 3 The GsSir2 and GsAgo proteins form a heterodimeric complex. A, Top – SDS PAGE of purified wt GsSir2/Ago, a D230A mutant, and C-terminal HSH tag-bearing GsSir2/Ago. Red star in the HSH-tagged sample lane marks an overlap of bands in the gel due to similar mass. Bottom – anti-His-tag Western blot of same samples. The red star shows the lane where the His-tag is on the C-terminus of Ago, rather than the N-terminus of Sir2. **B**, Mass photometry data of the GsSir2/Ago complex, with masses and respective particle population counts indicated. According to mass spectrometry of the purified GsSir2/Ago complex (Supplementary Fig. 3F), the molar mass of the GsSir2/Ago heterodimer is 121 kDa. **C**, SEC-MALS data of GsSir2/Ago, showing the chromatography peak and molar weight of the Sir2/Ago heterodimer. **D**, *Ab initio* model of GsSir2/Ago calculated from SAXS data by GASBOR. Colour coding: Sir2 domain – cyan, APAZ – brown, MID – yellow, PIWI – green.

Sir2/Ago prefers an RNA guide to bind a DNA target *in vitro*. Long pAgos use ssDNA and/or ssRNA guides to recognize their complementary DNA and/or RNA targets[10]. To establish guide preference of Gs and Cc Sir2/Ago, we analysed by EMSA binding of single- or double-stranded DNA or RNA (Fig. 4, Supplementary Fig. 6). Both Sir2/Ago heterodimers showed a strong preference for ssDNA and ssRNA binding. RNA/DNA heteroduplex was bound with an intermediate affinity, while dsRNA or dsDNA showed only weak binding (Fig. 4A, Supplementary Fig. 6, Table 1). Surprisingly, neither the 5'-terminal phosphate nor the 3'-OH end are important for ssDNA binding by both Gs and Cc complexes, as both bind the circular ssDNA and linear oligonucleotides with a similar affinity (Fig. 4, Supplementary Fig. 6). Binding affinity of the GsSir2(D230A)/Ago and GsSir2/Ago-HSH mutants that are functionally compromised in cellular assays was similar to that of the wt complex (Supplementary Fig. 6C, D). In summary, EMSA experiments suggest that *in vitro* ssRNA or ssDNA are preferable GsSir2/Ago guides.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

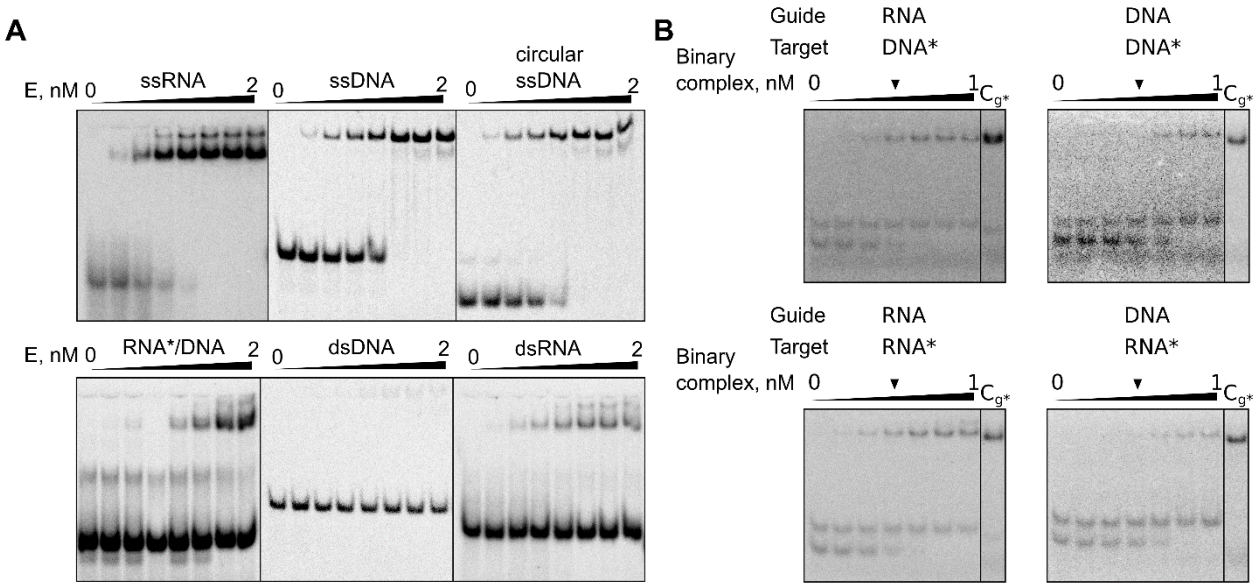


Fig. 4. Nucleic acid binding by GsSir2/Ago *in vitro*. A, EMSA of GsSir2/Ago binding to various DNA and RNA oligonucleotides. The asterisk denotes the radiolabelled strand. **B**, Binding of different targets by binary complexes pre-loaded with RNA or DNA guides. The pre-formed GsSir2/Ago complex with the guide strand (binary complex) was mixed with a radiolabelled target strand (see experimental details in "Materials and Methods"). Cg* denotes a control lane, equivalent to experimental lane marked by a black triangle, but with the guide, rather than the target, bearing the radioactive label. No displacement of the radiolabelled guide by the target strand is observed.

Next, we analysed ssDNA or ssRNA target binding by the binary GsSir2/Ago complexes pre-loaded with ssRNA or ssDNA guides. GsSir2/Ago-ssRNA binary complex showed tight binding to both complementary ssDNA and ssRNA targets. Importantly, RNA-guided GsSir2/Ago-ssRNA complex bound to complementary ssDNA target ~140-fold better than apo-GsSir2/Ago bound pre-annealed gRNA/tDNA heteroduplex (Fig. 4B, Table 1). On the other hand, DNA-guided GsSir2/Ago-ssDNA complex binding to the complementary ssRNA target was only 25-fold higher than apo-GsSir2/Ago binding to pre-annealed tDNA/gRNA heteroduplex (Fig. 4B, Table 1). Given these observations, we assume that GsSir2/Ago uses ssRNA as a guide for the recognition of the ssDNA target.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. K_d of wt GsSir2/Ago and CcSir2/Ago for various nucleic acids, calculated from EMSA data. K_d of a pre-formed GsSir2/Ago-guide complex was determined for two putative guides and two targets. Values are averages ± standard deviation of three independent replicates.

Oligonucleotide	**K_d, nM**	
	GsSir2/Ago	**CcSir2/Ago**
ssDNA	0.033 ± 0.016	0.038 ± 0.015
ssRNA	0.042 ± 0.015	0.36 ± 0.098
circular ssDNA	0.072 ± 0.018	0.18 ± 0.068
dsRNA/DNA	4.31 ± 0.37	2.4 ± 0.69
dsDNA	51 ± 20	330 ± 87
dsRNA	73 ± 29	81 ± 15

Guide-target mechanism GsSir2/Ago		
Guide	**Target**	**K_d, nM**
ssRNA	ssDNA	0.031 ± 0.003
	ssRNA	0.047 ± 0.007
ssDNA	ssDNA	0.121 ± 0.021
	ssRNA	0.172 ± 0.031

The GsSir2/Ago complex binds NAD^+ and causes its depletion. Computational analysis of Sir2 domains showed that they possess a conserved NAD^+-binding pocket (Supplementary Fig. 1D and Supplementary Fig. 5). To determine whether Sir2 domains can indeed bind endogenous NAD^+, purified GsSir2/Ago and CcSir2/Ago complexes were heat-treated, protein aggregates removed by centrifugation, and the supernatant analysed by MS-HPLC (Fig. 5 and Supplementary Fig. 7). The quantitative analysis showed that both the GsSir2/Ago and CcSir2/Ago complexes co-purified with bound endogenous NAD^+ in approx. 1:1 molar (Sir2:NAD^+) ratio (Fig. 5A, B, Supplementary Fig. 3). However, in the case of the functionally inactive GsSir2(D230A)/Ago mutant, only 0.6% of all complexes were NAD^+-bound indicating that the mutation severely compromised NAD^+ binding by the Sir2 domain (Fig. 5B). NAD^+ binding by the GsSir2 subunit and the similarity of the GsSir2 to the N-terminal NADase of the ThsA protein from the anti-phage Thoeris system[18] prompted us to investigate the level of endogenous NAD^+ in the presence of the induced GsSir2/Ago system and its target pCDF plasmid. In these experiments the corresponding *E. coli* cells were lysed, proteins were removed, and the amount of NAD^+ in the supernatant was examined by MS-HPLC. When the wt GsSir2/Ago system was induced in the presence of pCDF plasmid NAD^+ was depleted (~120-fold decrease), whereas in the case of the functionally inactive GsSir2(D230A)/Ago and GsSir2/Ago-HSH mutants the level of endogenous NAD^+ was similar to that of the empty pBAD vector (Fig. 5C, Supplementary Fig. 7). It should be noted that a significant ~30-fold decrease of the NAD^+ level was observed when wt GsSir2/Ago system expression was induced even in the absence of pCDF suggesting that the heterologously expressed system may be toxic to the cells resulting in their slower growth (Fig. 5C, Supplementary Fig. 7). To test the hypothesis that the activated

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GsSir2/Ago system similarly to the Thoeris anti-phage system depletes the endogenous NAD^+ through hydrolysis, we attempted to identify possible hydrolysis products (ADPR, cADPR, AMP, cAMP, ADP, cADP, nicotinamide, and adenine) using MS-HPLC albeit without success. Taken together, these results show that the wt GsSir2/Ago complex causes NAD^+ depletion when activated by the target pCDF plasmid.

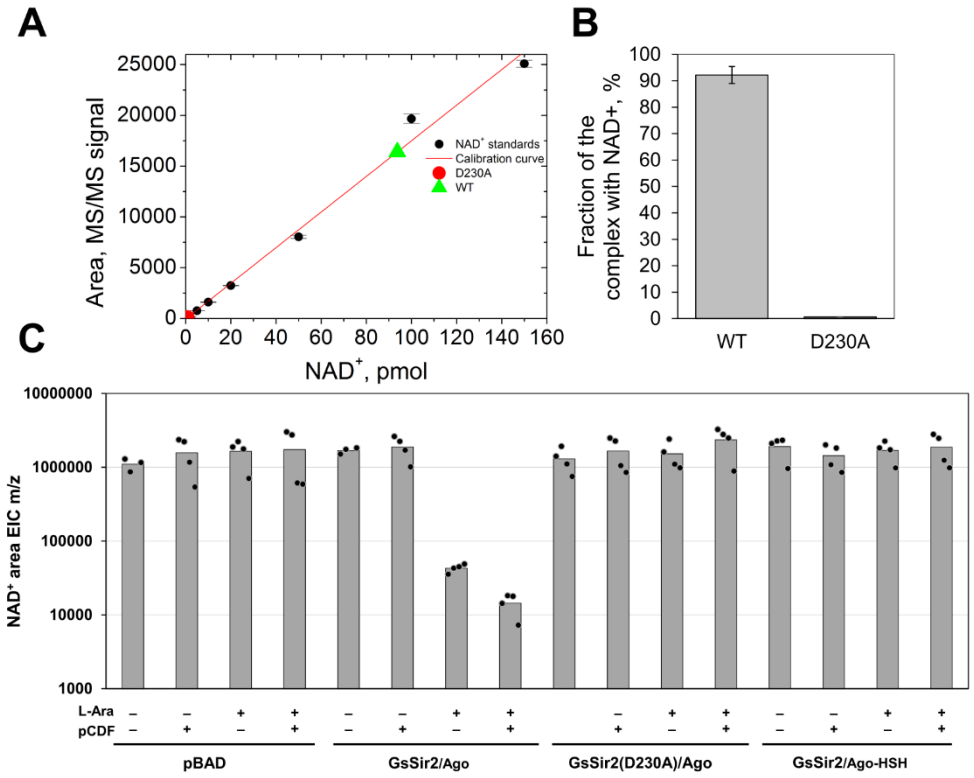


Fig. 5 The GsSir2/Ago complex binds NAD^+ *in vitro* and causes NAD^+ depletion in *E. coli* cells. A, MS/MS calibration curve of NAD^+ standard (marked in black) and the observed amount of NAD^+ in three samples: 93.7 pmol in the wt GsSir2/Ago sample (marked in green), 0.7 pmol in the sample D230A (marked in red). **B**, the D230A mutation within the Sir2 protein abolished NAD^+ binding. **C**, NAD^+ amount in *E. coli* cells in the presence of the (non)induced wt and mutant GsSir2/Ago systems and in the presence and absence of the pCDF plasmids. NAD^+ amounts were estimated according to the EIC areas of NAD^+ (m/z 662.1018).

Discussion

Association of Sir2-like domains with short pAgo proteins has been identified bioinformatically in the pioneering Makarova et al. paper[16]. It has been speculated that Sir2-domain proteins can act as nucleases, however, the structure and function of Sir2 proteins remained to be elucidated. Here, we show that the Sir2 and short pAgo proteins form a heterodimeric complex (Fig. 3) similar to the heterodimeric complex between the short pAgo and the Mrr nuclease domain-containing protein[19]. It has been suggested[16] that PAZ domain that is present in long pAgos is replaced by APAZ domain in short pAgo systems, however, sequence comparison and structure modelling suggest that the canonical N domain and the L1 and L2-like subdomains are part of the Sir2 proteins, while the PAZ domain is missing (Fig. 1A, Supplementary Fig. 5). In this respect, single-chain PgSir2-Ago protein is similar to long pAgos that acquired Sir2 domain but lost the PAZ domain. We further show that the Sir2 domain isolated from the heterologous *E. coli* cells contains bound NAD^+ (Fig. 5). Next, we provide the first experimental evidence that the Sir2/Agos complex functions as a defence system against invading phages and plasmids (Fig. 1 and Fig. 2). Intriguingly, plasmid interference

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assay using four plasmids with different replicons (Supplementary Table 2) revealed that GsSir2/Ago system prevents transformation only of the pCDF plasmid that contains CloDF13 *ori* (Fig. 2), suggesting that GsSir2/Ago may recognize specific replicon elements or structures. Indeed, *ori* swap between pCDF and permissive pCOLA plasmid made the latter sensitive to GsSir2/Ago interference. It has been previously shown that ColE1-like origins, including CloDF13, use two sense/antisense RNAs for priming the replication process that proceeds through the R-loop intermediate[20–22]. RNA is also involved in the priming of lambda phage theta replication that also involves R-loop formation, suggesting that priming RNAs/replication intermediate could be a recognition determinant for Sir2/Ago systems[23]. It is possible that Ago subunit in GsSir2/Ago complex, which shows a preference for ssRNA, binds ssRNA to interact with an ssDNA target, interfering with replication. Such GsSir2/Ago interaction with an *ori* region during the lambda phage infection or the pCDF plasmid transformation could activate the Sir2 effector subunit that depletes NAD^+ leading to the cell death, thereby, restricting phage propagation (Fig. 6, the accompanying paper). A similar anti-phage defence mechanism based on NAD^+ exhaustion has been shown for the Thoeris and the Pycsar systems[24,25]. In the Thoeris system of *Bacillus cereus* MSX-D12, the activated Sir2 domain is similar to that of the GsSir2/Ago system and performs the hydrolysis of NAD^+ to ADPR and nicotinamide[24]. However, in the case of the GsSir2/Ago system, we were unable to detect ADPR and nicotinamide suggesting that despite the similarity of Sir2 domains in both systems, the fate of NAD^+ might be different. It cannot be excluded that instead of the NAD^+ hydrolysis, GsSir2/Ago may catalyze the transfer or polymerization of NAD^+ (or its cleavage products) on other cellular components (e.g. proteins or nucleic acids). Further structural and biochemical studies are underway to establish the structure of the heterodimeric Sir2/Ago complex and mechanisms of *ori* recognition and Sir2 activation that triggers NAD^+ transformation.

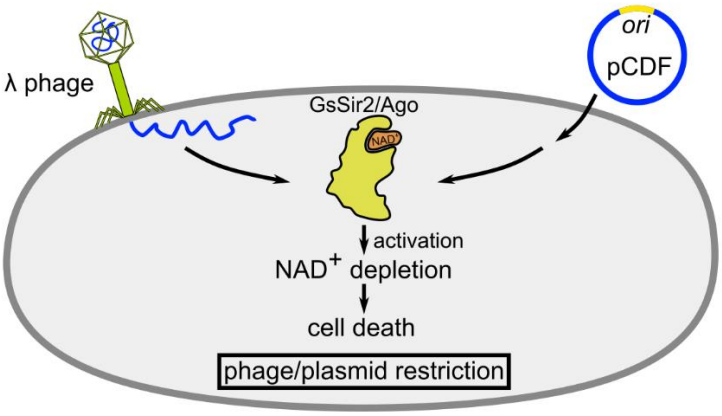


Fig. 6 Putative model of GsSir2/Ago defence against mobile genetic elements. After lambda phage infection or pCDF plasmid transformation, the GsSir2/Ago system becomes activated and triggers NAD^+ depletion and cell death. Recognition of the CloDF13 replication origin of pCDF plasmid seems to be a major factor that triggers NAD^+ depletion by the Sir2 domain.

Methods

Oligonucleotides used in this work. All synthetic DNA oligonucleotides used for cloning and site-specific mutagenesis were purchased from Metabion (Germany) and are listed in Supplementary Table 1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cloning and mutagenesis. A whole operon of the GsSir2/Ago system, composed of the genes encoding the Sir2 (GSU1360, NP_952413.1) and Ago (GSU1361, NP_952414.1) proteins, was amplified by PCR using the oligonucleotides MZ-239 and MZ-240 (Supplementary Table 1), respectively, from the genomic DNA of *Geobacter sulfurreducens* Caccavo (ATCC No. 51573, LGC Standards cat#51573D-5). The resulting DNA fragment was digested by Eco31I (ThermoFisher cat#FD0293) and XhoI (ThermoFisher cat#FD0694) and using T4 DNA ligase (ThermoFisher cat#EL0014) was cloned into pBAD/HisA expression vector (ThermoFisher cat#V43001) precleaved with NheI (ThermoFisher cat#FD0973) and XhoI and dephosphorylated using FastAP (ThermoFisher cat#EF0651). The GsSir2 protein contains a His_6-Tag at its N terminus. The D230A mutant of the GsSir2 protein was produced by QuikChange Site-Directed Mutagenesis[26] using respective mutagenic oligonucleotides (Supplementary Table 1). To inactivate the GsAgo protein a bulky **H**is$_6$-**S**trepII-**H**is$_6$-tag (HSH-tag, 29 aa.: LEGHHHHHHSSWSHPQFEKGVEGHHHHHH) was fused to its C terminus. For this, a whole operon of the GsSir2/Ago system was amplified by PCR using the oligonucleotides MZ-325 and MZ-326 (Supplementary Table 1), respectively, from the genomic DNA. The resulting DNA fragment was digested by Eco31I and XhoI and using T4 DNA ligase was cloned into pBAD24-HSH expression vector precleaved with NcoI (ThermoFisher cat#FD0573) and XhoI and dephosphorylated using FastAP. In this case, the GsSir2 protein does not contain any tag at its N-terminus.

A PgSir2-Ago gene (BgramDRAFT_6510, WP_052303232.1) was amplified by PCR using the oligonucleotides MZ-915 and MZ-916 (Supplementary Table 1), respectively, from the genomic DNA of *Paraburkholderia graminis* C4D1M (ATCC No. 700544) purchased from LGC Standards (UK). The resulting DNA fragment was digested by BveI (ThermoFisher cat#FD1744) and HindIII (ThermoFisher cat#FD0504) and using T4 DNA ligase was cloned into pBAD24 expression vector precleaved with NcoI and HindIII and dephosphorylated using FastAP.

The genes (codons were optimized for *E. coli* by IDT Codon Optimization Tool) encoding the Sir2 (WP_053571900.1) and Ago (WP_053571899.1) of the CcSir2/Ago system (from *Caballeronia cordobensis*, NCBI taxon_id 1353886) were synthesized and cloned into pBAD/HisA expression vector by Twist Bioscience. The CcSir2 protein contains at its N-terminus a His_6-tag that can be cleaved by TEV protease. For purification of the CcSir2/Ago complex, a TwinStrep-tag (35 aa.: MGGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKGS) was additionally fused to the N-terminus of the CcSir2 protein already containing a His_6-tag.

To swap *ori* regions between pCOLA and pCDF plasmids the DNA fragments containing ColA and CloDF13 *ori* were amplified by PCR using the oligonucleotides MZ-1217/MZ-1218 and MZ-1230/MZ-123, respectively (Supplementary Table 1). The resulting DNA fragments were digested by NheI and XbaI (ThermoFisher cat#FD0684)and using T4 DNA ligase were cloned into pCOLA and pCDF vectors precleaved with NheI and XbaI and dephosphorylated using FastAP.

To swap streptomycin resistance for kanamycin in the pCDF plasmid, for use in *E. coli* strain DH10B, plasmids pCDF and pCOLA were cleaved with NheI and Eco81I (ThermoFisher cat#FD0374). Subsequently, the appropriate fragments – the longer from pCDF and the shorter, harbouring the Kn resistance, from pCOLA – were purified separately using a runView electrophoresis system (Cleaver Scientific). The purified fragments were then ligated appropriately, to yield a pCDF_Kn plasmid.

All complete gene sequences were confirmed; links to DNA and protein sequences are presented in Supplementary Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phage restriction assay. *E. coli* MG1655 (ATCC 47076) cells transformed with wt or mutated Sir2/Ago systems, cloned into the pBAD vector, were used for phage infection assays as described below. Whole-genome sequencing was applied to all transformed *E. coli* clones as described previously[27], to verify the integrity of the system and lack of mutations.

E. coli phages (T4, T7, lambda-vir) were kindly provided by U. Qimron. Phages SECphi17, SECphi18 and SECphi27 were isolated by the Sorek lab as described before[28]. Small drop plaque assay was performed as described by Mazzocco et al.[29]. Overnight culture of *E. coli* bacteria was diluted 1:100 in MMB medium (LB + 0.1 mM $MnCl_2$ + 5 mM $MgCl_2$ + 5 mM $CaCl_2$) supplied with 0.1% L-arabinose for expression induction. Bacterial cultures were incubated at 37 °C until early log phase (OD_{600} = 0.3), and 500 µl of bacteria were mixed with 25 ml of MMB agar (LB + 0.1 mM $MnCl_2$ + 5 mM $MgCl_2$ + 5 mM $CaCl_2$ + 0.5% agar + 0.1% L-arabinose) and poured into the square Petri dish. Serial dilutions of phage lysate in MMB were dropped on top of cell lawn. After the drops dried up, plates were incubated at room temperature for 24 hours. Efficiency of plating (EOP) was determined via comparison of phage lysate titer on control bacteria and bacteria containing the Argonaute system with and without induction with L-arabinose.

Liquid culture phage infection experiments were performed as described previously by Ofir et al.[27]. After overnight incubation, the liquid suspension of pAgo-lacking and pAgo-containing *E. coli* cells were diluted 1:100 in MMB medium supplied with 0.2% L-arabinose and dispensed into a 96-well plate by 180-µl volume. Plates were incubated at 37 °C until the early log phase (OD_{600} = 0.3), then 20 µl of phage lysate was added to each well in multiplicity of infection 5, 0.5, or 0.05, and each experiment was performed in three replicates. Optical density measurements at a wavelength 600 nm were taken every 15 min using a TECAN Infinite 200 plate reader.

Plasmid restriction assay. *E. coli* BL21-AI and DH10B strain cells were pre-transformed with pBAD/HisA constructs encoding a GsSir2/Ago, CcSir2/Ago or PgSir2-Ago operon - with GsSir2/Ago being either wild-type or harbouring genetically engineered mutations - under the control of araBAD promoter. After 2 hours of induction with either 0.01% (w/v) (CcSir2/Ago) or 0.1% (others) L-arabinose at 37°C 200 RPM, protein-expressing cells were transformed with pCDF (pCDF_Kn in the case of PgSir2-Ago in DH10B), pCOLA, pACYC184 and pRSF plasmids by following standard heat-shock transformation protocol. After recovery, cells were either serially diluted and aliquots spotted on selection medium or undiluted suspensions spread on selection medium and CFUs counted manually. In parallel, viability and over-expression of pAgo-containing *E. coli* cells were monitored using serial dilutions and Western blot, respectively.

In separate experiments, *E. coli* BL21-AI strain cells were heat-shock transformed with both GsSir2/Ago operon-encoding pBAD/HisA construct and pCDF plasmid. Protein expression in selected double transformants was then induced by the addition of L-Ara (final concentration of 0.1% (w/v)) into liquid LB culture. After a 2-hour induction at 37°C 200 RPM, optical density (OD_{600}) equalized, cultures were serially diluted and aliquots were spotted on a selection medium containing different antibiotics.

Expression and purification of GsSir2/Ago complexes. Expression vector constructs of the GsSir2/Ago system were used to transform *E. coli* TOP10 strain. Transformed bacteria were grown at 37°C in LB medium in the presence of 50 µg/ml ampicillin until OD_{600} = 0.7 was reached. Then, the medium was cooled to 16°C temperature and proteins were expressed for 16 h by adding 0.2% w/v L-arabinose. Harvested cells were disrupted by sonication in buffer A (20 mM Tris–HCl (pH 8.0 at 25°C), 500 mM NaCl, 2 mM phenylmethylsulfonyl fluoride, 5 mM 2-mercaptoethanol), and cell debris was removed by centrifugation. GsSir2/Ago complexes were purified to > 90% homogeneity by chromatography through HisTrap HP chelating, HiTrap Heparin HP and HiLoad

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Superdex 200 columns (GE Healthcare). Purified proteins were stored at −20 °C in a buffer containing 20 mM Tris-HCl (pH 8.0 at 25°C), 200 mM KCl, 1 mM DTT and 50% v/v glycerol. The identity of the purified proteins was confirmed by mass spectrometry. Protein concentrations were determined from OD_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/. GsSir2/Ago complex concentrations are expressed in terms of heterodimer.

SEC-(MALS) and mass photometry. Size-exclusion chromatography of GsSir2/Ago complexes was carried out at room temperature using Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with a buffer (20 mM Tris–HCl (pH 8.0 at 25°C), 500 mM NaCl). A calibration curve was generated by measuring the elution volumes of a series of standard proteins of known molecular mass (Bio-Rad). The molecular masses of pAgos complexes were calculated by interpolating its elution volume onto the calibration curve. SEC-MALS of GsSir2/Ago and CcSir2/Ago complexes was performed at room temperature using Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated with a buffer (20 mM Tris–HCl (pH 8.0 at 25°C), 500 mM NaCl, 0.03% NaN_3, 1 mM DTT), at 0.4 ml/min flow rate. Sample concentrations were 6 µM and 6.5 µM for GsSir2/Ago and CcSir2/Ago, respectively. The light scattering signals were monitored on a miniDawn TREOS II detector, concentrations of protein samples were measured using an Optilab T-rEX refractive index detector (Wyatt Technologies). Data were analysed in Astra software (Wyatt Technologies) using dn/dc value of 0.185 mL/g. Mass photometry of the GsSir2/Ago complex was performed using a Refeyn OneMP system (Refeyn). The protein complex was diluted to 20 nM in a buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl before measurement.

Small-angle X-ray scattering (SAXS) analysis. The synchrotron SAXS data were collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany)[30]. GsSir2/Ago sample was transferred from the storage buffer to the sample buffer using gel-filtration NAP column (GE Healthcare) and concentrated by ultrafiltration to the concentrations 1.2, 1.3, 1.6 and 5.5 mg/ml. The sample buffer contained 20 mM Tris-HCl pH7.5, 200 mM NaCl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol. The data were collected at the wavelength of 0.124 nm and the distance to the detector (Pilatus 2M, Dectris) was set to 3 m. Samples in the sample changer were kept at 10°C, capillary temperature was set to 20°C. Twenty frames exposed 0.045 sec were averaged for each concentration. The s-range of collected data was from 0.0133796 to 3.7925 nm^{-1}. The data were analysed using programs of ATSAS 2.8.4 (r10552) suite[31]. Data were normalized to an absolute scale with water as standard. As the data collected for the sample with concentration 1.22 mg/ml were noisy at higher s (Supplementary Fig. 4), and higher concentration data showed more aggregation at low s, we used a merged dataset produced with PRIMUS[32]. Scattering data were parameterized and indirectly Fourier transformed with GNOM5[33]. Structural parameters of this dataset are summarized in Supplementary Table 3. Dimensionless Kratky plot in Supplementary Fig. 4 was calculated as described previously[34]. The *ab initio* models were calculated by GASBOR[35] software. Molecular mass estimations of apo GsSir2/Ago complex in solution assessed by ATSAS tools (DATVC, DATMW) and server SAXSMoW (http://saxs.ifsc.usp.br/)[36] are presented in the Supplementary Table 3.

Nucleic acid binding assay. The oligonucleotide substrates (Supplementary Table 1) were 5'-labelled with [γ-^{32}P]ATP (PerkinElmer) and T4 polynucleotidyl kinase (PNK) (ThermoFisher cat#EK0031). The 3'-labelled substrate was prepared with [α-^{32}P]cordycepin-5'-triphosphate (Hartmann Analytics) and terminal deoxynucleotidyl transferase. An aliquot of the 3'-labelled substrate was subsequently phosphorylated with cold ATP and T4 PNK to obtain a 3'-labelled-5'P oligonucleotide. Before annealing complementary strands, EDTA was added to 50 mM. Annealing was performed in the same PNK reaction buffer at total single-stranded oligonucleotide concentrations of 2 µM. Circular ssDNA substrate was prepared from TK-49 5'-^{32}P using

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CircLigase II (Lucigen cat#CL9021K), according to manufacturer recommendations, then purified from a denaturing PAA gel (21% 29:1 acrylamide/bis-acrylamide in TBE supplemented with 8 M urea) by phenol-chloroform extraction, precipitated in 96 % ethanol with 0.45 M sodium acetate, washed with 75% ethanol, and resuspended in water.

For EMSA experiments, appropriate substrates and enzymes were pre-diluted to 2x the final binding reaction concentration in 40 mM Tris-acetate (pH 8.3 at 23°C), 1 mM EDTA (TAE, Invitrogen, 24710-030), supplemented with 5 mM magnesium acetate, 0.1 mg/ml BSA, 1 mM DTT, and 10% glycerol. The binding reactions were conducted by mixing equal volumes of enzyme and radiolabelled substrate. In all cases, final binding reactions contained 0.1 nM of radiolabelled substrate at 0-2 nM (0; 0.02; 0.05; 0.1; 0.2; 0.5; 1; 2) or 0-500 nM (0; 5; 10; 20; 50; 100; 200; 500) of GsSir2/Ago and CcSir2/Ago complexes. Three independent replicates were performed.

The guide-target experiments for GsSir2/Ago were conducted by first pre-mixing 5'-phosphorylated guide NA:protein complex at a ratio of 1:1 using the same buffer as above. The pre-mixed GsSir2/Ago:guide complex was then diluted to 2x final reaction concentration (according to guide) in the same buffer and mixed with a 5'-^{32}P-target oligonucleotide. The final reaction contained 0.1 nM target NA and 0, 0.02, 0.05, 0.1, 0.2 0.5, and 1 nM of GsSir2/Ago:guide complex, according to guide concentration. For the displacement control lane (C_{g*}), GsSir2/Ago was pre-mixed with 5'-^{32}P-guide at a ratio of 1:1, diluted to 0.2 nM according to guide and mixed with 5'-phosphorylated target NA to final concentrations of 0.1 nM target, 0.1 nM GsSir2/Ago:guide complex, according to guide concentration. Three independent replicates were performed.

The binding reaction mixtures were analysed by electrophoretic mobility shift assay (EMSA) in a PAA gel (8% 29:1 acrylamide/bis-acrylamide in TAE). The electrophoresis TAE buffer was supplemented with 5 mM magnesium acetate. Radiolabelled substrates were detected and quantified using a phosphor imager. The results were analysed with OptiQuant and OriginPro software. The K_d was calculated from the following formula:

$$\mathbf{S_{NB}} = \frac{\mathbf{S_0 - E_0 - K_D + \sqrt{(S_0 + E_0 + K_D)^2 - 4S_0E_0}}}{\mathbf{2}}$$

where S_{NB} – unbound substrate, nM; S_0 – initial substrate concentration, nM; E_0 – initial protein complex concentration, nM; K_d – dissociation constant.

Preparation of *E. coli* cells for endogenous NAD$^+$ quantification. Overnight cultures of single colonies of *E. coli* DH10B strain harbouring a pBAD-His construct with either wt GsSir2/Ago or mutant system (GsSir2/Ago-HSH or GsSir2(D230A)/Ago), or empty vector (negative control) were diluted and grown in LB broth (BD) supplemented with respective antibiotics (50 µg/ml ampicillin and 25 µg/ml streptomycin) at 37°C until they reached OD_{600} of 0.4–0.5. Cell cultures were either induced to express the protein or not (control samples). L-Ara (0.1% final concentration) was added to induce protein expression. Induced and non-induced cultures were harvested 2 hours later. The cultures were normalized to OD_{600} of approximately 0.7 and the pellet from 1 mL of culture suspension was stored at -80°C until further analysis. All cell pellets were lysed by adding B-PER solution (Thermofisher Scientific) supplemented with 6 mg/ml lysozyme (62971, Fluka) for 20 min at room temperature while gently rocking (Multi Bio 3D Mini-Shaker, Biosan). Cell debris was removed by centrifugation and metabolites were isolated by phenol:chloroform:isoamyl alcohol (PCI) (25:24:1, v/v/v) extraction. Metabolites were stored at -20°C until MS-HPLC analysis. The endogenous cell NAD$^+$ amounts were estimated from four independent measurements.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

High-performance liquid chromatography-mass spectrometry (HPLC-MS) of NAD^+. To quantitate NAD^+ bound to GsSir2/Ago and CcSir2/Ago complexes HPLC-MS/MS analysis was used. First, purified pAgos complexes were diluted to 5 µM in a buffer containing 20 mM Tris-HCl (pH 8.0 at 25°C), 200 mM NaCl. Then 20 µl of the solution was incubated at 70°C for 20 min and centrifuged for 30 min (16,100 g at 4°C) to remove unfolded proteins. The supernatants and NAD^+ standards were analysed by Electrospray Ionization mass spectrometry (ESI-MS) using an integrated HPLC/ESI-MS system (1290 Infinity, Agilent Technologies / Triple Quadrupole 6410, Agilent Technologies), equipped with a Supelco Discovery®HS C18 column (7.5 cm × 2.1 mm, 3 µm), Agilent Technologies. HPLC/ESI-MS/MS was performed using two ion transitions to detect NAD^+ in the samples: 662.1→540.1 and 662.1→426.0. Ion transition 662.1→540.1, as it is the most abundant, was used for the quantitative analysis. Mobile phase A was 5 mM ammonium acetate in water, pH 7.0 and mobile phase B was 5 mM ammonium acetate in methanol, pH 7.0. The HPLC parameters were as follows: flow 0.25 ml/min; column temperature 30°C; 0-3 min, 0%B; 3-9 min, 0-40%B; 9-10 min, 40-100%B; 10-13 min, 100%B. The MS was operated using negative electrospray ionisation at 2500 V, the gas temperature was set to 300°C, the fragmentor voltage was 135V. Multiple reaction monitoring (MRM) was used with a collision energy of 15 V to measure ion m/z 540.1 (ion transition 662.1→540.1) and also with a collision energy of 20 V to measure ion m/z 426.0) (ion transition 662.1→426.0).

To quantitate endogenous NAD^+ HPLC-MS analysis was performed by Electrospray Ionization mass spectrometry (ESI-MS) using an integrated HPLC/ESI-MS system (1290 Infinity, Agilent Technologies / Q-TOF 6520, Agilent Technologies), equipped with a Supelco Discovery®HS C18 column (7.5 cm × 2.1 mm, 3 µm), Agilent Technologies. The samples were investigated in both negative and positive ionization modes. For negative ionization mode, solvents A (5 mM ammonium acetate in water pH 7.0) and B (5 mM ammonium acetate in methanol, pH 7.0) were used. For positive ionization mode, solvents C (0,02% formic acid in water) and D (0,02% formic acid in acetonitrile) were used. In both cases elution was performed with a linear gradient of solvents at a flow rate of 0.3 ml/min at 30°C as follows: 0–5 min, 0% B; 5–18 min, 20% B; 18–22 min, 100% B, 22–27 min 100% B. Ionization capillary voltage was set to 2500 V and fragmentor to 150V. A list of compounds that could be expected to be products of NAD^+ hydrolysis and relative m/z value is as follows: ADPR [M-H]- m/z=558.0644, cADPR [M-H]- m/z =540.0538, AMP [M-H]- m/z=346.0558, cAMP [M-H]- m/z=328.0452, ADP [M-H]- m/z=426.0221, cADP [M-H]- m/z=408.0116, nicotinamide [M+H]+ m/z= 123.0553, adenine [M+H]+ m/z=136.0618. Only traces of AMP and ADP were detected in all samples; other products from the list were absent.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Kuhn, C. D. & Joshua-Tor, L. Eukaryotic Argonautes come into focus. *Trends Biochem. Sci.* **38**, 263–271 (2013).

2. Pratt, A. J. & MacRae, I. J. The RNA-induced silencing complex: A versatile gene-silencing machine. *J. Biol. Chem.* **284**, 17897–17901 (2009).

3. Sheu-Gruttadauria, J. & MacRae, I. J. Structural Foundations of RNA Silencing by Argonaute. *J. Mol. Biol.* **429**, 2619–2639 (2017).

4. Olina, A. V., Kulbachinskiy, A. V., Aravin, A. A. & Esyunina, D. M. Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes. *Biochem.* **83**, 483–497 (2018).

5. Hutvagner, G. & Simard, M. J. Argonaute proteins: Key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32 (2008).

6. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21**, 743–753 (2014).

7. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat. Struct. Mol. Biol.* **19**, 145–151 (2012).

8. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The Expanded Universe of Prokaryotic Argonaute Proteins. *MBio* **9**, 1–20 (2018).

9. Hegge, J. W., Swarts, D. C. & Van Der Oost, J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nat. Rev. Microbiol.* **16**, 5–11 (2018).

10. Lisitskaya, L., Aravin, A. A. & Kulbachinskiy, A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nat. Commun.* **9**, 1–12 (2018).

11. Willkomm, S., Makarova, K. S. & Grohmann, D. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiol. Rev.* **42**, 376–387 (2018).

12. Kuzmenko, A. *et al.* DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* (2020) doi:10.1038/s41586-020-2605-1.

13. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).

14. Jolly, S. M. *et al.* Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. *Cell* **182**, 1545-1559.e18 (2020).

15. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Mol. Cell* **51**, 594–605

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2013).

16. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009).

17. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).

18. Ka, D., Oh, H., Park, E., Kim, J. H. & Bae, E. Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. *Nat. Commun.* **11**, 1–8 (2020).

19. Kim, S., Jung, Y. & Lim, D. Argonaute system of Kordia jejudonensis is a heterodimeric nucleic acid-guided nuclease. *Biochem. Biophys. Res. Commun.* (2020) doi:10.1016/j.bbrc.2020.02.145.

20. Dasgupta, S., Masukata, H. & Tomizawa, J. ichi. Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. *Cell* **51**, 1113–1122 (1987).

21. del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. Replication and Control of Circular Bacterial Plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).

22. Selzer, G., Som, T., Itoh, T. & Tomizawa, J. ichi. The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **32**, 119–129 (1983).

23. Skalka, A. M. DNA Replication---Bacteriophage Lambda. in *Current Topics in Microbiology and Immunology* (eds. Arber, W. et al.) 201–237 (Springer Berlin Heidelberg, 1977). doi:10.1007/978-3-642-66800-5_7.

24. Ofir, G. *et al.* Antiviral activity of bacterial TIR domains via signaling molecules that trigger cell death. *bioRxiv* (2021) doi:10.1101/2021.01.06.425286.

25. Tal, N. *et al.* Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728-5739.e16 (2021).

26. Zheng, L., Baumann, U. & Reymond, J. L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32**, (2004).

27. Ofir, G. *et al.* DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).

28. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (80-.).* **359**, eaar4120 (2018).

29. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of Bacteriophages Using the Small Drop Plaque Assay System. *Methods Mol. Biol.* 81–85 (2009) doi:10.1007/978-1-60327-164-6_9.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015).

31. Franke, D., Petoukhov, M. V, Konarev, P. V & Panjkovich, A. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 1212–1225 (2017) doi:10.1107/S1600576717007786.

32. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

33. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

34. Durand, D. *et al.* NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169**, 45–53 (2010).

35. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).

36. Fischer, H., De Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).

37. Koonin, E. V & Makarova, K. S. CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 Biol. Rep.* **1**, 1–6 (2009).

38. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

39. Burroughs, A. M., Ando, Y. & Aravind, L. New perspectives on the diversification of the RNA interference system: Insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip. Rev. RNA* **5**, 141–181 (2014).

40. Boutet, E. *et al.* UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol. Biol.* **1374**, 23–54 (2016).

41. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

42. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).

43. Frickey, T. & Lupas, A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704 (2004).

44. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

45. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

46. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

47. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

48. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

49. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).

50. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

51. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-5 (2016).

52. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-85 (2016).

53. Mirdita, M., Ovchinnikov, S. & Steinegger, M. ColabFold - Making protein folding accessible to all. *bioRxiv* 2021.08.15.456425 (2021).

54. Olechnovič, K. & Venclovas, Č. VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic Acids Res.* **47**, W437–W442 (2019).

55. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140 (2020).

56. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

57. Kaya, E. *et al.* A bacterial Argonaute with noncanonical guide RNA specificity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4057–4062 (2016).

58. Blattner, F. R. *et al.* The Complete Genome Sequence of Escherichia coli K-12. *Science (80-.)*. **277**, 1453–1462 (1997).

59. Bhawsinghka, N., Glenn, K. F., Schaaper, R. M. & Rasko, D. Complete Genome Sequence of Escherichia coli BL21-AI. *Microbiol. Resour. Announc.* **9**, e00009-20 (2020).

60. Grant, S. G. N., Jesseet, J., Bloomt, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. **87**, 4645–4649 (1990).

61. Kulczyk, A. W. & Richardson, C. C. *The Replication System of Bacteriophage T7*. *Enzymes* vol. 39 (Elsevier Inc., 2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

62. Casjens, S. R. & Hendrix, R. W. Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479**–**480**, 310–330 (2015).

63. Yap, M. L. & Rossmann, M. G. Structure and function of bacteriophage T4. *Future Microbiol.* **9**, 1319–1337 (2014).

64. Tolia, N. H. & Joshua-Tor, L. Strategies for protein coexpression in Escherichia coli. *Nat. Methods* **3**, 55–64 (2006).

65. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

Acknowledgements

We thank the Siksnys laboratory members for comments on the manuscript and fruitful discussion. This work was supported by the Research Council of Lithuania [09.3.3-LMT-K-712-01-0126 to V.S.], the Israel Science Foundation [grant ISF 296/21 to R.S.], and the Deutsche Forschungsgemeinschaft [SPP 2330, grant 464312965]. The authors thank Tomás de Garay of Refeyn Ltd. for conducting the mass photometry experiments. Funding for open access charge: Vilnius University.

Author Contributions

R.S., V.S. and M.Z. designed the study; K.T. and Č.V. performed bioinformatics and structural modelling; A.L. performed the phage restriction experiments; E.S., D.D., E.G., S.A. performed the plasmid transformation experiments under the supervision of M.Z.; A.S. purified the proteins and performed the SEC-(MALS) experiments under the supervision of M.Z.; R.G. performed the SEC experiments under the supervision of M.Z.; E.M. performed the SAXS experiments; E.G., D.T. and R.G. performed the EMSA experiments under the supervision of M.Z.; A.R. performed the mass spectrometry analysis; D.D. performed the NAD determination experiments *in vivo* under the supervision of M.Z.; R.S., V.S. and M.Z. analysed the data; M.Z. wrote the manuscript with input from E.G., R.S., V.S. and other authors. All authors approved the final version.

Competing Interests statement

R.S. is the scientific founder of BiomX and Ecophage. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https:

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Materials

Supplementary Figures

Supplementary Fig. 1. Bioinformatic analysis.

Supplementary Fig. 2. Additional and control *in vivo* experiments.

Supplementary Fig. 3. Purification and characterization of Sir2/Ago complexes.

Supplementary Fig. 4. SAXS data.

Supplementary Fig. 5. Structural analysis.

Supplementary Fig. 6. EMSA and nucleic acid cleavage experiments.

Supplementary Fig. 7. The GsSir2/Ago complex binds NAD^+ and causes its depletion.

Supplementary Tables

Supplementary Table 1. Oligonucleotides used in this work.

Supplementary Table 2. Strains, bacteriophages, plasmids, and proteins used in this work

Supplementary Table 3. Parameters of the SAXS data.

Supplementary text

Closer inspection of the genomic neighbourhood of the selected pAgos showed that no other conserved operons are formed with the pAgo genes (Supplementary Fig. 1B). However, the observed enrichment of (putative) restriction endonucleases, mobile genetic elements (e.g. transposases, integrases) and toxin-antitoxin systems in the neighbourhood of the GsAgo and CcAgo genes indicates that they could be a part of so-called bacterial defence islands[37]. All three systems have MID-PIWI domains, characteristic of Agos, yet in this case, the PIWI domains are catalytically inactive (Fig. 1A and Supplementary Fig. 1C). MID domains have conserved HK-motif residues, which are essential for binding the 5'-end of the guide strand, and are conserved in other short pAgos (Supplementary Fig. 1E). Unlike long pAgos, short pAgos lack the PAZ domain and are rather associated with an APAZ (analogue of PAZ) domain (Fig. 1A)[16]. By phylogeny of the APAZ domain region, PgSir2-Ago belongs to Ia, and GsSir2/Ago and CcSir2/Ago to Ib groups, respectively (Supplementary Fig. 1A)[8]. The sequences of Sir2 domains associated with the three studied pAgos are similar to canonical sirtuins[16], and also have an identifiable signature of a putative NAD^+-binding pocket (Supplementary Fig. 1D). Yet there are many differences from canonical sirtuins, and the identified conserved amino acid positions found only in pAgo-associated Sir2 domains suggest that these domains might have a function different from the typical deacetylase activity of sirtuins (Supplementary Fig. 1D). pAgos-associated Sir2 domains also show similarities to NADase domains of ThsA proteins from the Thoeris anti-phage systems (Supplementary Fig. 1D)[18,28].

GsSir2/Ago complex resembles long PAZ-free pAgos containing an additional effector domain. To get an insight into the 3D structure of GsSir2/Ago and the other two systems, we used AlphaFold[38] to generate corresponding structural models. The constructed models for all three systems showed close structural similarity with each other as could be expected based on their homology (Supplementary Fig. 5). Therefore, we will focus only on the GsSir2/Ago complex. In

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

this complex, the Sir2 and Ago subunits bind together to form a structure similar to that of a single-chain long pAgo protein such as RsAgo (Supplementary Fig. 5). Based just on a structural similarity search against PDB, GsSir2/Ago appears to be most similar to the TtAgo structure classified as a long-A pAgo[8]. However, another close structural match, RsAgo, a member of long-B pAgos[8], has a catalytically inactive PIWI domain. This is a shared feature with short pAgo proteins and therefore comparison of GsSir2/Ago with RsAgo might be more relevant.

The APAZ part of the GsSir2 chain is structurally similar to the combination of N, L1 and L2 domains of RsAgo (similarity to the N domain has been already proposed before[11,39]). However, the L2 linker domain in GsSir2/Ago corresponds to the two fragments, C-terminus of Sir2 and N-terminus of Ago. Importantly, the PAZ domain, required for the 3'-end recognition of the guide strand in long pAgos, is missing from the GsSir2/Ago structure altogether. Additionally, GsSir2/Ago has a smaller N domain than long pAgos, which could also alter nucleic acid binding. PIWI and MID domains are structurally similar to corresponding RsAgo domains. Notably, GsSir2/Ago has a longer loop (residues 268-272) compared to the corresponding region in RsAgo. This loop shows some steric overlap with the copied-in RNA/DNA duplex, but presumably, its conformation may adjust to accommodate RNA/DNA heteroduplex. The loop contains two positively charged residues (R269, K270) that might be involved in the binding of the nucleic acid backbone.

In the structural model, the N-terminal Sir2 domain of the GsSir2 protein is attached to the C-terminal domain corresponding to the N-terminal region of a long pAgo through a long linker. In effect, the two domains of GsSir2 are positioned at the opposite extremes of the complex (Sir2 domain is bound to the PIWI domain whereas the C-terminal domain is bound to the MID domain). Sir2 domain of GsSir2/Ago is structurally similar to canonical Sir2 proteins suggesting it could bind NAD^+ in a similar way. However, the overall structurally closest homolog is the Sir2 domain of the ThsA protein from the Thoeris defence system[18] (Supplementary Fig. 5). The ThsA residues, N112 and H152, shown to be essential for NAD^+ hydrolysis have their counterparts in GsSir2 (N142 and H186). Interestingly, the loop containing H186 in different GsSir2/Ago models displays some conformational heterogeneity hinting at possible flexibility, which might be relevant for the activity regulation. These two positions are conserved among both short pAgos and other Sir2 homologs (Supplementary Fig. 1D). On the other hand, the GsSir2 D230 residue shown here to be important for NAD^+ hydrolysis is conserved in ThsA Sir2 domain, but not in canonical Sir2 proteins (e.g., V193 in TmSir2).

Supplementary methods

Sequences analysis. Lists of pAgo homologues and associated APAZ domains were retrieved from supplementary data of the Ryazansky et al. article[8]. *Thermotoga maritima* Sir2 homologues were collected from the SwissProt database[40] using BLAST[41] (1e-5 e-value cutoff). Full-length sequences were retrieved from NCBI. Bacillus cereus ThsA (PDB ID 6LHX) homologues were collected using BLAST from UniRef50[42] database (1e-10 e-value cutoff). To remove other SIR2 homologues, these ThsA homologue sequences were clustered with CLANS[43]. Cluster separated at p=1e-70 was selected as the representative ThsA group. Fragmented ThsA sequences and sequences missing one of the domains were discarded. Multiple sequence alignments were generated using MAFFT (l-INS-i mode for high accuracy)[44]. Jalview[45] was used for multiple sequence alignment analysis, cutting and visualization. Sequence motif visualization was done using WebLogo 3 server[46]. To align the multiple sequence alignments of GsSir2 homologues TmSir2 homologues and

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ThsA homologues, GsSir2 sequence alignments to TmSir2 and BsThsA were obtained using the HHpred server[47].

Phylogenetic tree construction. The phylogenetic trees were constructed with FastTree[48] using the WAG model of amino acid substitution[49] and the gamma model of rate heterogeneity. Prior to phylogenetic analysis, multiple sequence alignment positions containing more than 50% gaps were removed using trimAl[50]. Phylogenetic tree visualization was done with iTOL[51].

Genomic neighbourhood analysis. For genomic neighbourhood analysis, the *Geobacter sulfurreducens*, *Caballeronia cordobensis*, and *Paraburkholderia graminis* genomes (respective GenBank accessions: GCA_000210155.1, GCA_001544575.2 and GCA_000172415.1) and all associated sequence and annotation data were obtained from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). Genes in the neighbourhood of each pAgo (10 upstream and 10 downstream) were identified based on available genome annotations. To refine available functional annotations of these genes, searches through Pfam[52] and PDB databases were performed using the HHpred server[47].

Expression and purification of CcSir2/Ago complex. Expression vector constructs of the CcSir2/Ago system were used to transform *E. coli* BL21(DE3) strain. Transformed bacteria were grown at 37°C in LB medium in the presence of 50 µg/ml ampicillin until OD_{600} = 0.7 was reached. Then, the medium was cooled to 16°C temperature and proteins were expressed for 16 h by adding 0.1% w/v L-arabinose. Harvested cells were disrupted by sonication in buffer A (20 mM Tris–HCl (pH 8.0 at 25°C), 1.0 M NaCl, 2 mM phenylmethylsulfonyl fluoride, 5 mM 2-mercaptoethanol), and cell debris was removed by centrifugation. TwinStrep-CcSir2/Ago complex was purified to > 90% homogeneity by chromatography through Strep-Tactin XT Superflow (Iba), HiLoad Superdex 200, HiTrap Heparin HP columns (GE Healthcare). Purified proteins were stored at −20 °C in a buffer containing 20 mM Tris-HCl (pH 8.0 at 25°C), 500 mM NaCl, 2 mM DTT and 50% v/v glycerol. The identity of the purified proteins was confirmed by mass spectrometry. Protein concentrations were determined from OD_{280} measurements using the theoretical extinction coefficients calculated with the ProtParam tool available at http://web.expasy.org/protparam/. TwinStrep-CcSir2/Ago complex concentrations are expressed in terms of heterodimer.

Construction and analysis of Sir2/Ago structural models. Structural models were generated using the AlphaFold method[38] implemented as ColabFold[53], an online Google Colaboratory notebook. For modelling, the 'Alphafold2_advanced' notebook (colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb) was used. GsSir2/Ago and CcSir2/Ago complexes were modelled as heterodimers, whereas the Pg sequence representing a fusion of Sir2 and a short pAgo (PgSir2-Ago) was modelled as a monomer. The modelling pipeline was run with default parameters except for the multiple sequence alignment (MSA) pairing, which was set to 'paired+unpaired'. The best-of-five model in each case was selected using VoroMQA[54]. Structure similarity searches of models/domains against PDB were performed using Dali[55]. Structure analysis and visualization was performed using UCSF Chimera[56]. Putative binding sites of Sir2/Ago complexes were investigated by simply copying the RNA/DNA duplex and NAD^+ from RsAgo (PDB ID: 5AWH) and TmSir2 (PDB ID: 2H4F) structures correspondingly after their superposition onto structural models. No attempts to remove possible clashes between protein models and either RNA/DNA or NAD^+ were made.

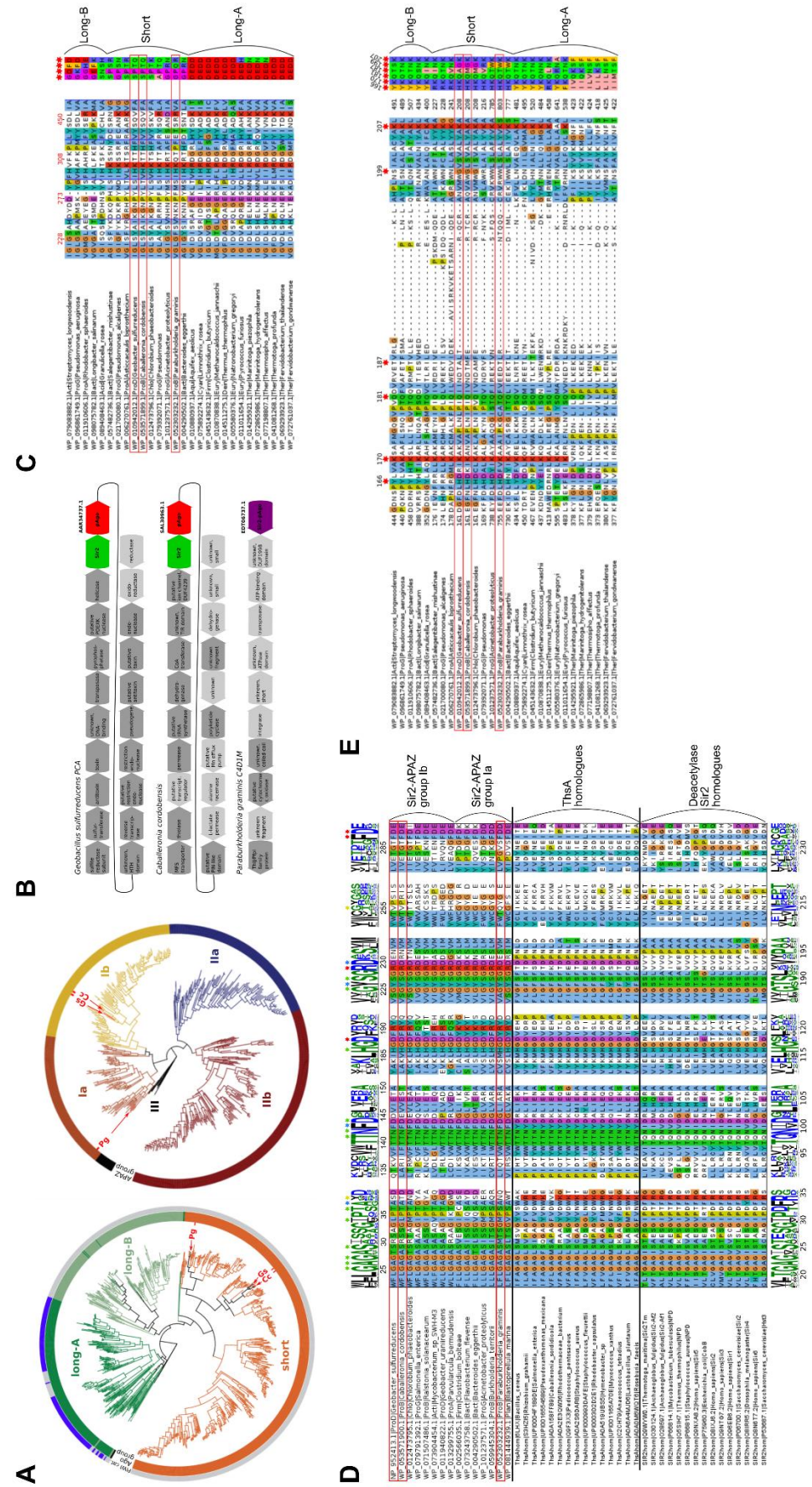
Nucleic acid cleavage assay. The same linear oligonucleotides used for EMSA (Supplementary Table 1) were given as substrates for nucleic acid cleavage assay. To raise the total substrate concentration, 5'-^{32}P radiolabelled oligonucleotides were mixed with appropriate cold 5'P-oligonucleotides at a ratio of 1:4 of hot:cold, and diluted to a working concentration of 100 nM in

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reaction buffer (33 mM Tris-acetate, pH 7.9, supplemented with 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/ml BSA, 5 mM DTT). Protein dilutions were made using the same reaction buffer to 2x final reaction concentration. Protein complexes and nucleic acids were mixed in final concentrations of 50 nM NAs and 0, 50 or 500 nM protein, incubated for 1 hour at 25 °C. The reaction was stopped with the addition of 2x 95% formamide dye and incubating for 5 min at 95 °C. Reaction products were resolved by denaturing PAA gel electrophoresis (21% 29:1 acrylamide/bis-acrylamide in TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA), supplemented with 8 M urea), visualised with a phosphor imager and analysed in OptiQuant software.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

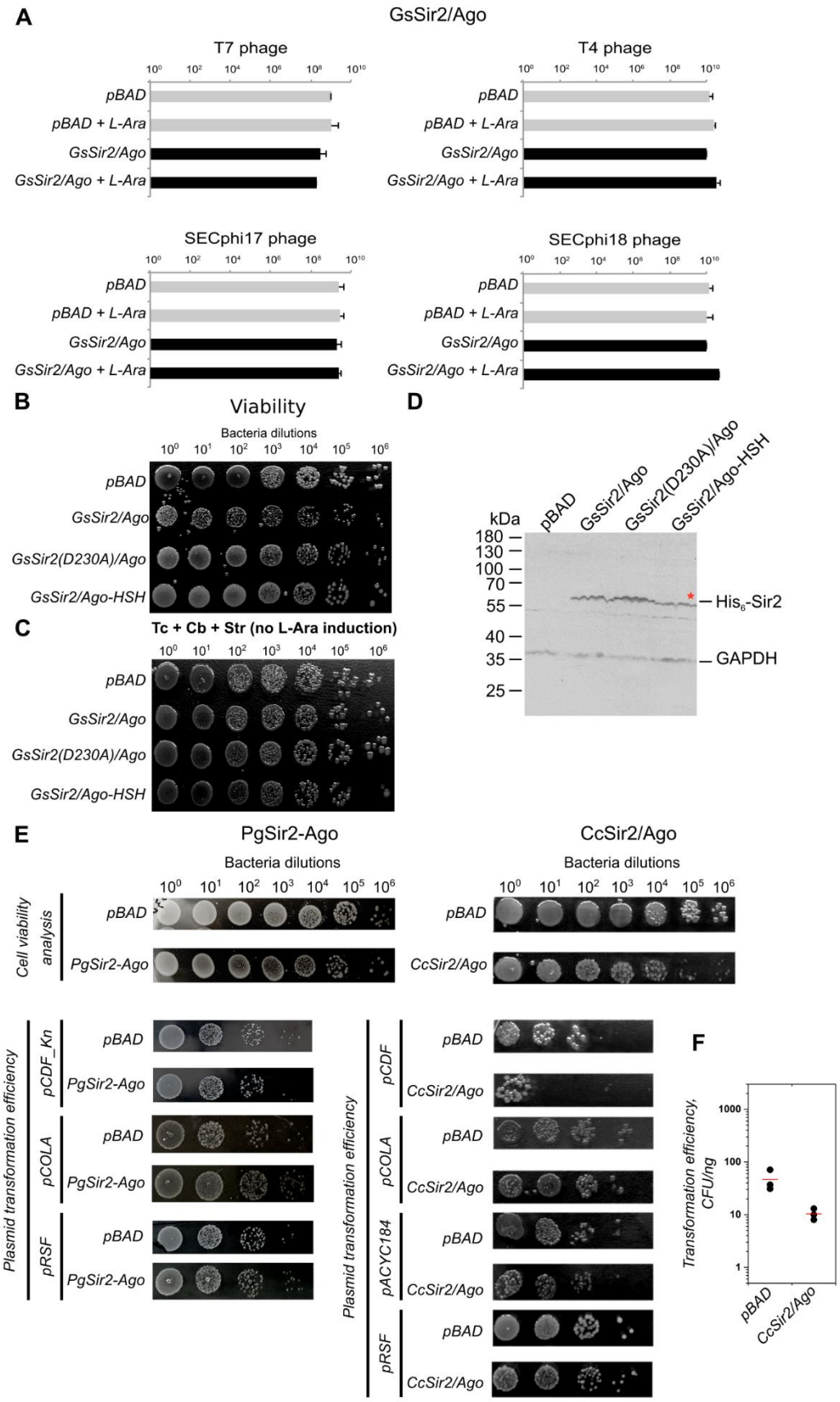
Supplementary figures



bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Fig. 1. Bioinformatic analysis. A, Left - a circular phylogenetic tree was generated according to supplementary data provided with Ryazansky et al.[8] Long-A pAgo variants are coloured in green (truncated variants without the PAZ domain, light green), long-B pAgo proteins are light green (truncated variants without PAZ, green), and short pAgo proteins are orange. pAgo proteins containing the catalytic tetrad DEDX in their PIWI domain are indicated in blue on the outer circle; pAgos with inactivated PIWI domain are indicated in light grey on the outer circle. pAgo proteins of the GsSir2/Ago, CcSir2/Ago and PgSir2-Ago systems are indicated by 'Gs', 'Cc' and 'Pg', respectively. Right - Circular phylogenetic tree of APAZ domains. The circular phylogenetic tree of the five groups of APAZ domains was generated using APAZ domain alignments from Ryazansky et al.[8] supplementary file 7. **B**, Genomic neighbourhood (±10 genes) of the pAgo sequences from *Geobacter sulfurreducens*, *Caballeronia cordobensis*, and *Paraburkholderia graminis*. Annotations are based on homologue search results and available sequence annotations. Co-directional sequences (that could form an operon with pAgo) are coloured darker grey, opposite direction – lighter. Sequences with no identified homologues for the larger part of the sequence are marked 'unknown' (any smaller identified domains are noted); less certain results are marked 'putative'. The *Paraburkholderia graminis* C4D1M genome is not complete and the PgSir2-Ago gene resides at the edge of a shotgun fragment, so only neighbouring genes from one side are shown for the PgSir2-Ago system. **C**, PIWI catalytic tetrad DEDX alignment. The 4 catalytic residues (red numbers indicate positions of corresponding GsAgo positions) are shown in 4 motifs of ±3 positions. The motifs are separated by vertical blue lines. Sequence names consist of the following: NCBI sequence ID, abbreviated phylum (e.g., 'ProG' – gamma-proteobacteria) and organism name. **D**, Combined alignment of Sir2 domains. Alignment consists of 3 parts, separated by horizontal black lines. In the top part, the Sir2 domain sequences of the GsSir2, CcSir2, PgSir2-Ago and homologues are shown. Logos above depict the conservation of Sir2 domains of Ia and Ib groups. The indicated position numbers correspond to the GsSir2 sequence. In the bottom part, homologues (sirtuins) of catalytically active Thermotoga maritima Sir2 (TmSir2) deacetylase are shown. Logos below indicate the conservation of these homologues. The position numbers correspond to the TmSir2 sequence. Sequences of six motifs that include all positions that form the NAD^+-binding pocket, as seen in the TmSir2 structure (PDB ID 2H4F) are shown. Sequence names for the top alignment consist of sequence ID, abbreviated phylum and organism name. Sequence names for bottom alignment all start with "Sir2hom" followed by sequence ID, organism name and short protein name (based on annotation). Stars above the logos indicate residues in the NAD^+-binding pocket of canonical sirtuins (e.g., TmSir2) that are also conserved. Star colours indicate conservation between the two groups: green – conserved in both canonical sirtuins and GsSir2-like; blue – conserved in both groups, but different; yellow – conserved only in canonical sirtuins; red – conserved only in GsSir2-like proteins. In the middle, alignment of ThsA homologues with Sir2 domains. **E**, MID domain alignment. Red stars indicate positions of amino acids involved in the binding of the 5'-P end of the guide nucleic acid. The numbering above corresponds to the GsAgo sequence. Additionally, concatenated alignment of just the 6 indicated positions is shown on the right. The three sequences of interest are indicated with red rectangles. Numbers on the left and right of the alignment indicate the first and last positions in the alignment for each sequence.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

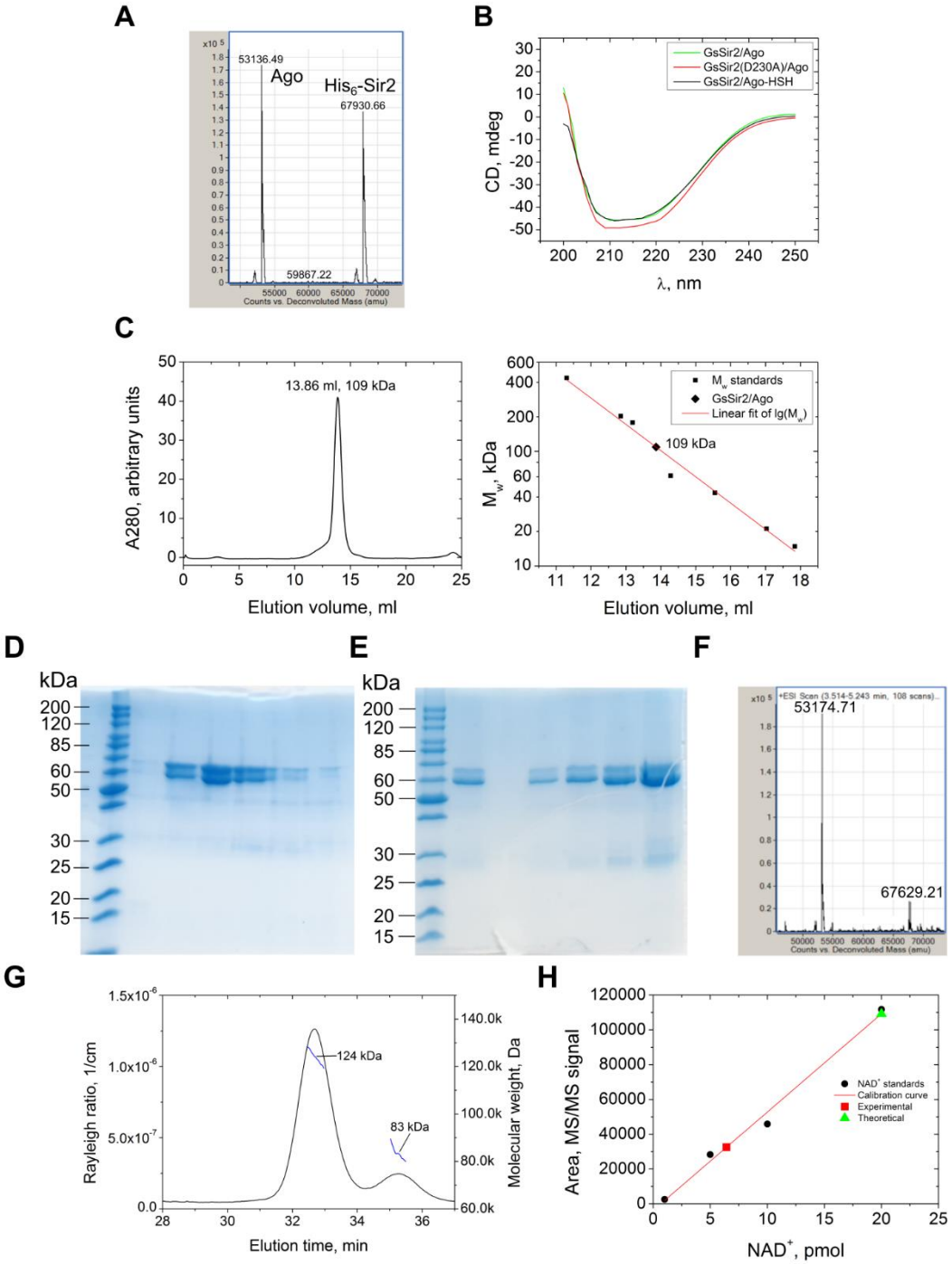


Supplementary Fig. 2. Additional and control *in vivo* experiments. A, Efficiency of plating (EOP) of 4 phages infecting *E. coli* cells with and without the GsSir2/Ago system from *Geobacter sulfurreducens*, where the GsSir2/Ago

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system exhibits no defence activity. The x-axis represents the number of p.f.u. Shown are the means of three replicates in the absence and in the presence of the inducer L-arabinose (L-Ara), and error bars are s.d. of the mean. Grey bars represent efficiency of plating (EOP) on pAgo-lacking cells and black bars are EOP in pAgo-containing cells. **B-D**, In vivo experimental controls for GsSir2/Ago plasmid restriction. **B**, Cell viability control of BL21-AI, expressing GsSir2/Ago and mutants. **C**, Control for Fig. 2E – cells contain the pCDF plasmid, however, expression of the GsSir2/Ago system is not induced. **D**, Expression of GsSir2/Ago assayed by Western blot. **E**, Left - qualitative characterization of plasmid restriction capabilities of PgSir2-Ago system in *E. coli* strain DH10B. Top: comparison of cell viability in the presence or absence of plasmid-borne PgSir2-Ago expression. Bottom: comparison of plasmid transformation efficiencies in the presence or absence of plasmid-borne PgSir2-Ago expression. Right - qualitative characterization of plasmid restriction capabilities of CcSir2/Ago system in *E. coli* strain BL20-AI: top - comparison of cell viability in the presence or absence of plasmid-borne CcSir2/Ago expression. Bottom - comparison of plasmid transformation efficiencies in the presence or absence of plasmid-borne CcSir2/Ago expression. **F**, Quantification of transformation efficiencies for pCDF plasmid with CcSir2/Ago system (three independent replicates, the red line represents average transformation efficiency).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

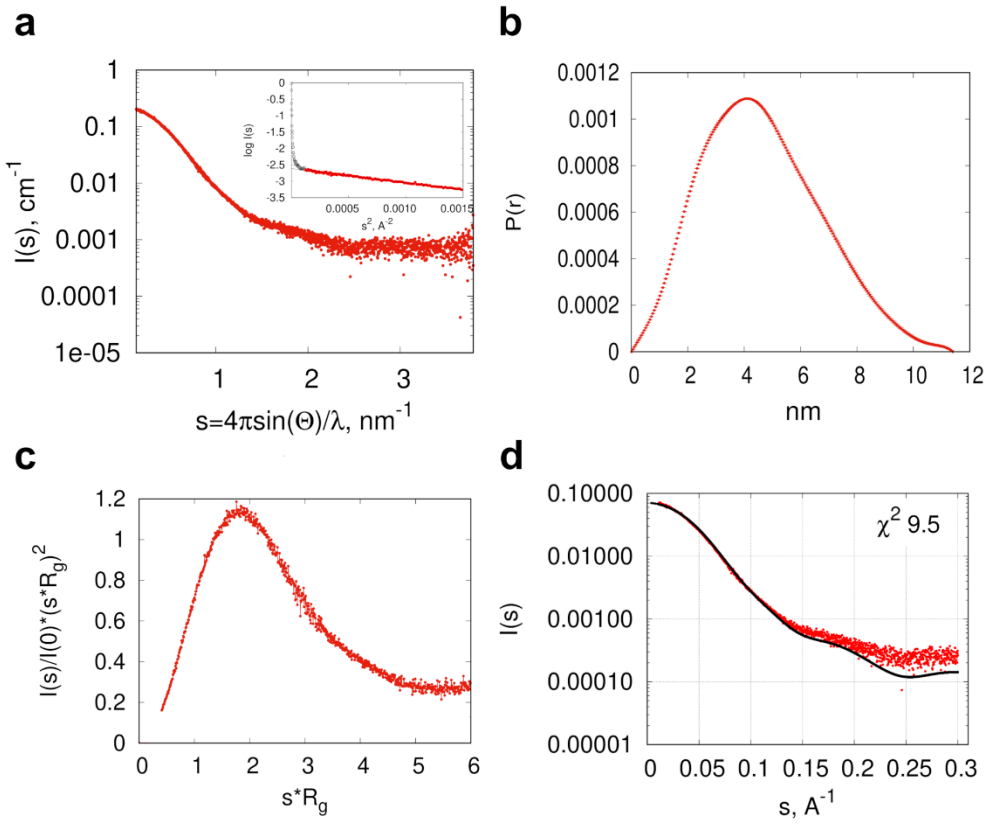


Supplementary Fig. 3. Purification and characterization of Sir2/Ago complexes. A, MS analysis of wt GsSir2/Ago complex. The theoretical Mw of the GsSir2 and GsAgo proteins (without 1st Met) are 67929.51 Da and 53135.61 Da, respectively. **B**, CD spectra of wt GsSir2/Ago and mutants. Mutant spectra are similar to that of a natively folded protein. **C**, Size-exclusion chromatography of wt GsSir2/Ago, showing elution volume and comparing to mass standards. According to mass spectrometry of the purified GsSir2/Ago complex, the molar mass of the GsSir2/Ago heterodimer is 121 kDa. **D**, SDS-PAGE analysis of fractions containing the CcSir2/Ago complex eluted from Heparin column. Densitometric inspection shows that Sir2 and Ago proteins are in the ratio ~1:1. **E**, SDS-PAGE analysis of the CcSir2/Ago stock after dialysis against a storage buffer. Various amounts of the stock solution were loaded on the gel. Densitometric inspection shows that Sir2 and Ago proteins are in the ratio ~0.3:1. **F**, MS analysis of the CcSir2/Ago complex. The experimental masses (53174.71 Da and 67629.21 Da) are close to the theoretical molecular masses of the Ago protein (53173.91 Da) and the Sir2 protein with the truncated tag at the N terminus (67626.24 Da). **G**, SEC-MALS analysis of the CcSir2/Ago complex. The experimental mass of 124 kDa is close to the theoretical molecular mass of the CcSir2/Ago heterodimer (121 kDa). **H**, MS/MS calibration curve of NAD^+ standard (marked in black) and the

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

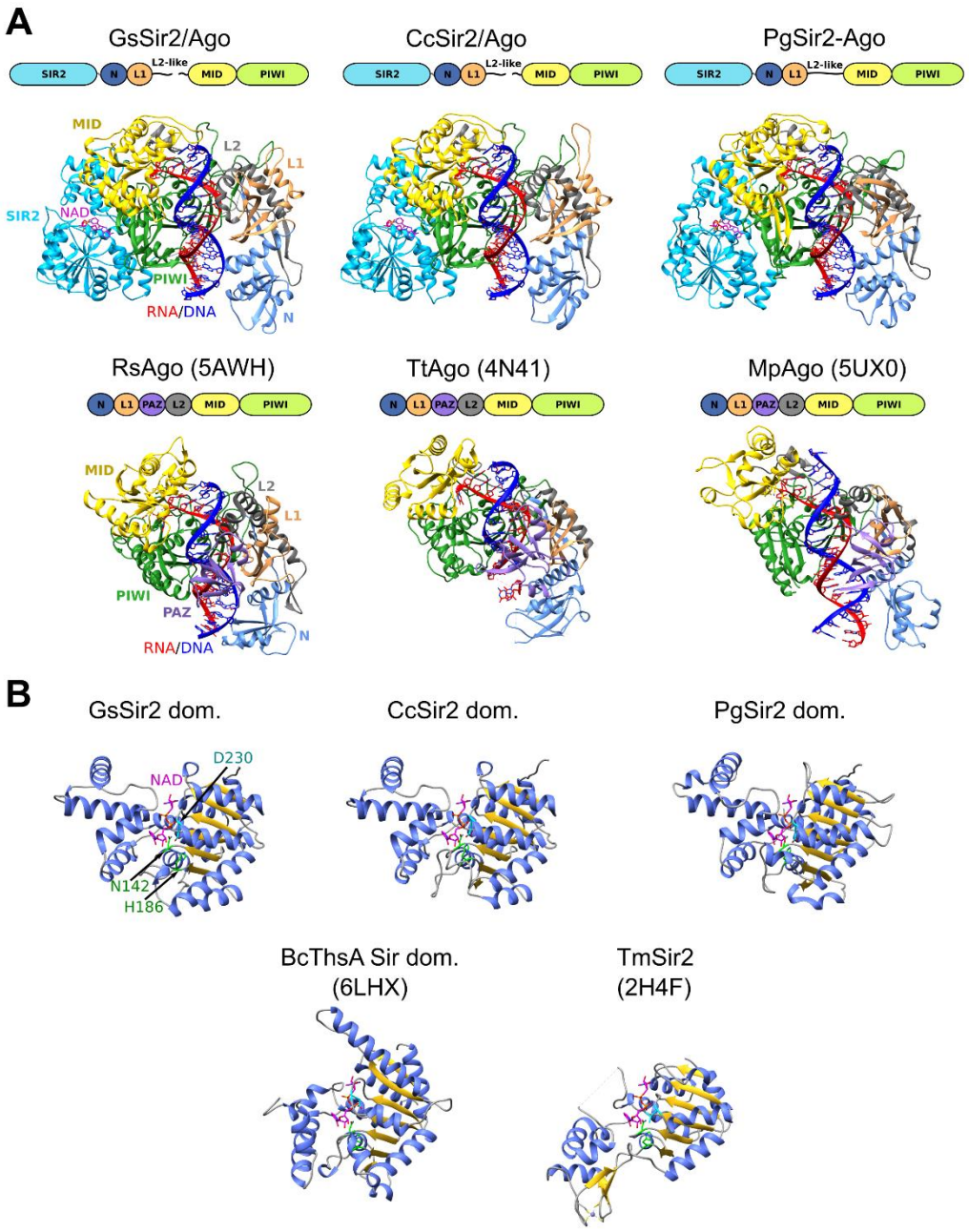
observed amount of NAD^+ (marked in red) in the CcSir2/Ago complex (20 pmol according to the Ago protein). The discrepancy between the expected amount of NAD^+ (20 pmol, marked in green) and the actual amount (6.45 pmol, marked in red) was due to the decrease of the Sir2 protein in the CcSir2/Ago preparation (see **E**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



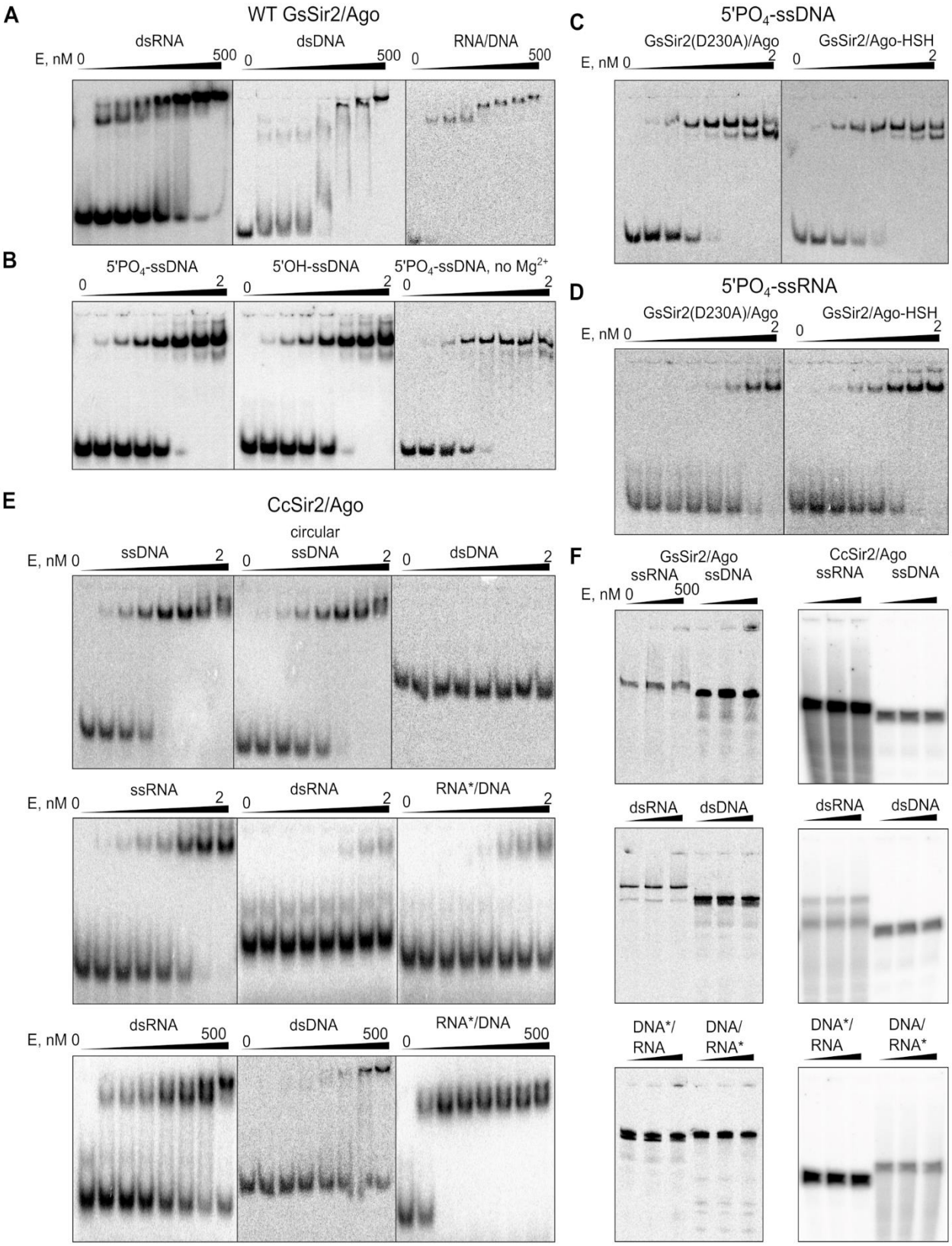
Supplementary Fig. 4. SAXS data. A, Scattering data on an absolute scale. Linear Guinier plot of the initial part of the scattering curve is in the insert. Points cut from the further processing are shown with empty black symbols. **B**, Kratky plot, normalized by Rg and I(0) parameters. **C**, Pair distance distribution function. **D**, CRYSOL Fit of the scattering curve calculated from GsSir2/Ago model (black curve) with SAXS data (red points).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Fig. 5. Structural analysis. A, Comparison of Sir2/Ago models to long pAgo structures. Structures are coloured by domains, schematic domain architecture is given above each structure. Guide RNA and target DNA strands are coloured red and blue, respectively. PDB ID codes for long pAgo structures are given in parentheses. RsAgo represents long-B group, TtAgo – longA group of long pAgos (based on Ryazansky et al. classification[8]). MpAgo has a distinct OH-type MID domain that is specific to 5'-OH instead of phosphate (MID domain classification – Ryazansky et al.[8], MpAgo MID domain biochemical assay - Kaya et al.[57]). In Sir2/Ago models the N, L1 and L2-like domains previously identified as the APAZ domain correspond to the analogous domains of long pAgos. **B**, Gs, Cc and Pg Sir2 structural models (cut from full-length models) compared to canonical Sir2 deacetylase TmSir2 and the Sir2 domain of Thoeris defence system protein ThsA. Structures are coloured based on secondary structure. Positions corresponding to ThsA Sir2 N112 and H152 are indicated in green. These residues have been shown to be critical for NAD^+ hydrolysis in ThsA[18]. GsSir2 D230 and corresponding positions in other structures are indicated in cyan. NAD^+ was also superimposed on the ThsA Sir2 structure from TmSir2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

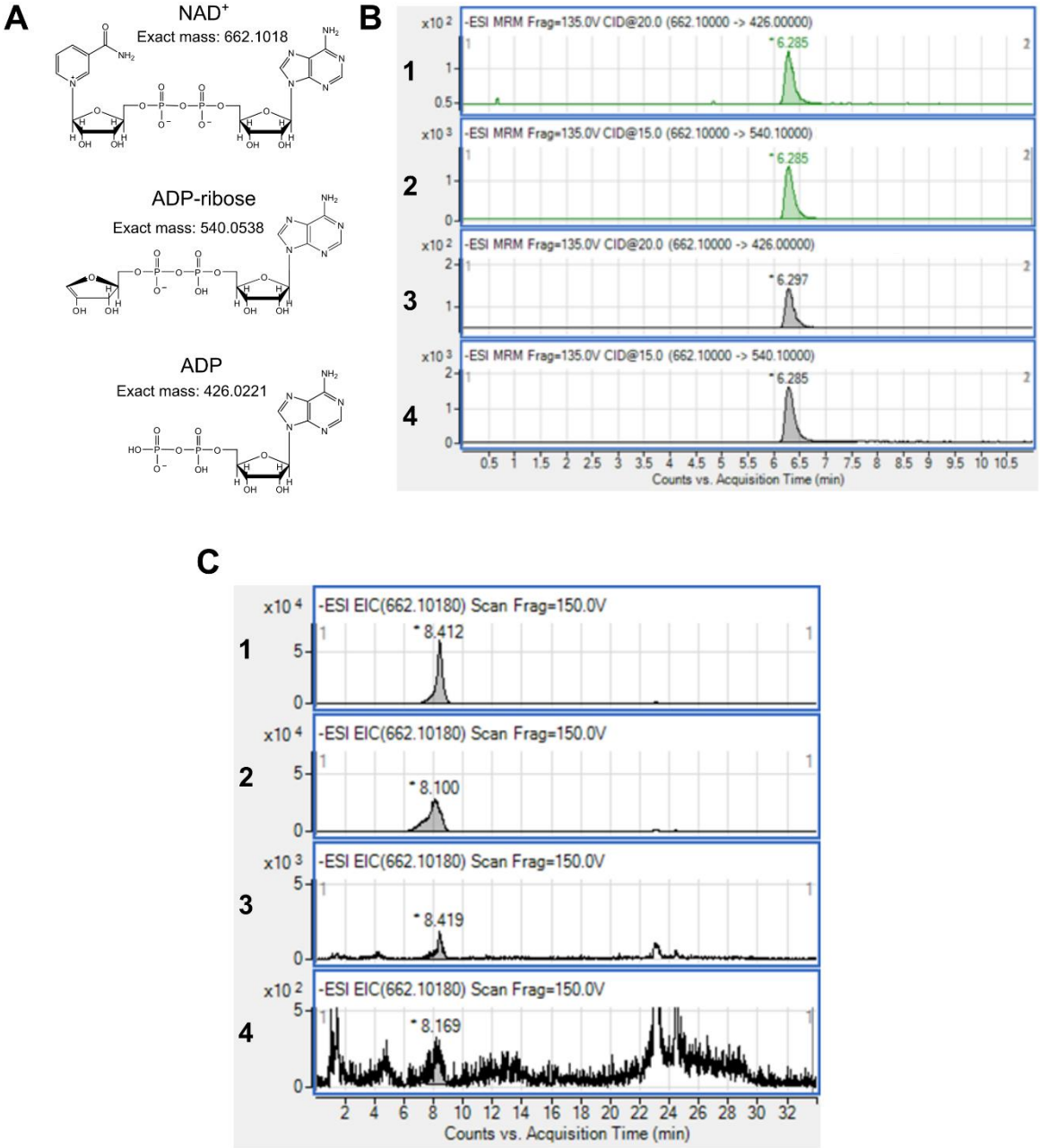


Supplementary Fig. 6. EMSA and nucleic acid cleavage experiments. A, Binding of double-stranded oligonucleotides by wt GsSir2/Ago. **B**, Influence of the presence of the terminal 5'-phosphate and Mg^{2+} ions on ssDNA binding by wt GsSir2/Ago. **C** and **D**, Binding of ssDNA and ssRNA, respectively, by the GsSir2(D230A)/Ago and GsSir2/Ago-HSH mutants. **E**, EMSA experiments of CcSir2/Ago with various DNA and RNA oligonucleotides. **F**, (No) cleavage activity of various DNA and RNA oligonucleotides by GsSir2/Ago and CcSir2/Ago. Reaction products

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were resolved on a 21% denaturing polyacrylamide gel. In heteroduplexes, the asterisk indicates the radiolabelled strand.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary Fig. 7. The GsSir2/Ago complex binds NAD$^+$ and causes its depletion. A, Two ion transitions were used to detect NAD$^+$ in the analysed samples: 662.1→540.1 and 662.1→426.0. **B**, Detection of NAD$^+$. Comparison of the extracted LC-MS/MS chromatograms: ion transition 662.1→540.1 of wt GsSir2/Ago sample (panel 1) and NAD$^+$ standard (panel 3); ion transition 662.1→426.0 of wt GsSir2/Ago sample (panel 2) and NAD$^+$ standard (panel 4). **C**, Mass chromatogram. For NAD$^+$ detection an extracted ion current (EIC) for [M-H]- m/z = 662.1018 was used. The comparison of EIC signals shows that the amount of NAD$^+$ in the samples of the non-induced GsSir2/Ago system in the absence (panel 1) and presence (panel 2) of pCDF plasmid is almost the same, while a significant decrease is observed in the sample of the induced GsSir2/Ago system in the absence of pCDF plasmid (panel 3) and only traces of NAD$^+$ are detected in the presence of pCDF plasmid (panel 4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary tables

Supplementary Table 1. Oligonucleotides used in this work.

Oligonucleotide	Sequence (5'->3' direction)	Description
Oligonucleotides used for cloning and mutagenesis		
MZ-239	GCTGATGGTCTCGCTAGCATGGATGTCTTAACTGACAATGAGTTTTAC	Forward primer for the amplification of the GsSir2/Ago operon.
MZ-240	CTGATTCTCGAGTTACATAAAAAACCGATAATCATATATTTCGTTAAC	Reverse primer for amplification of the GsSir2/Ago operon.
MZ-351	GTGGTCGTGCTGAAAATGTTATGACTATGCTTC	Mutagenic primer (top strand) for the D230A mutation within the GsSir2 protein.
MZ-352	GTCATAACATTTTCAGCACGACCACTATAGCC	Mutagenic primer (bottom strand) for the D230A mutation within the GsSir2 protein.
MZ-325	GCTGATGGTCTCGCATGGATGTCTTAACTGACAATGAGTTTTAC	Forward primer for the amplification of the GsSir2/Ago operon.
MZ-326	TGATGCTCGAGCATAAAAAACCGATAATCATATATTTCGTTAAC	Reverse primer for the amplification of the GsSir2/Ago operon.
MZ-915	TGATTACCTGCTTAGCATGCATAATCGCGCTGCTTTGCAAACG	Forward primer for the amplification of the PgSir2-Ago gene.
MZ-916	TGATTAAGCTTTACATAAACAGACGGTAATCGTACGAGTGGTC	Reverse primer for the amplification of the PgSir2-Ago gene.
MZ-1217	GAAATCTTAGAGTGATGGTGTCGGGAATCC	Forward primer for the amplification of the COLA *ori*.
MZ-1218	AGTGATGGTGTCGGGAATCC	Reverse primer for the amplification of the COLA *ori*.
MZ-1230	AAATAGCTAGCTCACTCGGTC	Forward primer for the amplification of the CDF *ori*.
MZ-1231	GAAATCTAGAGCGGTTCAGTAG	Reverse primer for the amplification of the CDF *ori*.
Oligonucleotides used for EMSA and cleavage experiments		
ssRNA (TF-A)	AUAAUGGUUUCUUAGACGUCGUUUUAGAGCUGUGUUGUUUCG	42 base RNA oligonucleotide
ssDNA (TK-49)	ATAATGGTTTCTTAGACGTCGTTTTAGAGCTGTGTTGTTTCG	42 base DNA oligonucleotide. Sequence identical to TF-A, only DNA.
ssDNA (MZ-949)	CGAAACAACACAGCTCTAAAACGACGTCTAAGAAACCATTAT	42 base DNA oligonucleotide, complementary to TF-A and TK-49
ssRNA (MZ-1185)	CGAAACAACACAGCUCUAAAACGACGUCUAAGAAACCAUUAU	42 base RNA oligonucleotide, complementary to TF-A and TK-49
dsRNA	5'-AUAAUGGUUUCUUAGACGUCGUUUUAGAGCUGUGUUGUUUCG-3' 3'-UAUUACCAAAGAAUCUGCAGCAAAAUCUCGACACAACAAAGC-5'	TF-A annealed to MZ-1185
dsDNA	5'-ATAATGGTTTCTTAGACGTCGTTTTAGAGCTGTGTTGTTTCG-3' 3'-TATTACCAAAGAATCTGCAGCAAAATCTCGACACAACAAAGC-5'	TK-49 annealed to MZ-949
RNA/DNA	5'-AUAAUGGUUUCUUAGACGUCGUUUUAGAGCUGUGUUGUUUCG-3' 3'-TATTACCAAAGAATCTGCAGCAAAATCTCGACACAACAAAGC-5'	TF-A annealed to MZ-949

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 2. Strains, bacteriophages, plasmids, and proteins used in this work

Bacterial strains (*Escherichia coli*)	**Details**	**Source or reference, links**
MG1655	F^- λ^- *ilvG*$^-$ *rfb*-50 *rph*-1	Blattner, F. *et al.*[58]
BL21-AI	F^- *omp*T *hsd*S_B (r_B m_B^-) *gal dcm ara*B::T7RNAP-*tet*A	Bhawsinghka, N. et al.[59]
DH10B	F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara*,*leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ^-	Grant, S.G.N. *et al.*[60]
Bacteriophages (*Escherichia coli*)		
T7	NC_001604.1	Udi Qimron; Kulczyk, A. & Richardson, C.C.[61]
lambda-vir	NC_001416.1	Udi Qimron; Casjens, S.R. & Hendrix, R.W.[62]
SECphi27	LT961732.1	Sorek lab collection; Doron, S. *et al.*[28]
SECphi18	LT960607.1	Sorek lab collection; Doron, S. *et al.*[28]
T4	AF158101.6	Udi Qimron; Yap, M.L. & Rossmann, M.G.[63]
SECphi17	LT960609.1	Sorek lab collection; Doron, S. *et al.*[28]
Plasmids		
pBAD/HisA	Bacterial expression vector, carrying a pBR322 replicon and ampicillin resistance, provides an N-terminal His_6 tag for expressed proteins, ~40 copies/cell.	ThermoFisher cat#V43001. Tolia, N.H. & Joshua-Tor, L.[64] https://benchling.com/s/seq-tfRnXVmxwmGJqYnXmsDw?m=slm-gnSb1LLWrQibTPl2ibbs
pBAD24	Bacterial expression vector, carrying pBR322/ColE1 and F1 replicons, and ampicillin resistance gene, ~40 copies/cell.	Guzman, L.M *et al.*[65] https://benchling.com/s/seq-zaoDsxfoymWz0CDoHSEI?m=slm-ALGNMJ5F3aprLKyOrHOK
pBAD24-HSH	pBAD24 vector with an HSH affinity tag before the rrnB terminator. Used to fuse the tag on target protein C-terminus, ~40 copies/cell.	Guzman, L.M *et al.*[65]
pCDF	Bacterial vector carrying a CloDF13 replicon, *lacI* gene and streptomycin/spectinomycin resistance gene, 20-40 copies/cell.	Tolia, N.H. & Joshua-Tor, L.[64] https://benchling.com/s/seq-B4oyRce17nWdhFpsWmGx?m=slm-SrF4zHJbfnBfIhrG49jP
pCDF_Kn	pCDF vector, carrying kanamycin resistance instead of the canonical streptomycin, 20-40 copies/cell.	In-house construct.[64] https://benchling.com/s/seq-7KKOFJRtBzzCbjA4MN3L?m=slm-Ez8cisdT8LtXDMSNSlur
pCOLA	Bacterial vector with a ColA origin of replication, harbours kanamycin resistance gene, 20-40 copies/cell.	Tolia, N.H. & Joshua-Tor, L.[64] https://benchling.com/s/seq-dtqZ232iXhuGqGiyZR0a?m=slm-hLBJqu72mbC0yEKkFKeu
pACYC184	Bacterial vector carrying p15A origin of replication and tetracycline resistance gene, 10-12 copies/cell.	Tolia, N.H. & Joshua-Tor, L.[64] https://benchling.com/s/seq-Rz0xd8JhIeZa1vwsFx7c?m=slm-5rwfx80s5z1sp04YeEbJ
pRSF	Bacterial vector carrying the RSF1030 replicon, *lacI* gene and kanamycin resistance gene, >100 copies/cell.	Tolia, N.H. & Joshua-Tor, L.[64] https://benchling.com/s/seq-7C0lDtWUXncuAQyKn19F?m=slm-wTVKOxRQQYAFlv1CXJET
pCDF_ColA	pCDF plasmid with its *ori* site replaced with the *ori* site from a pCOLA plasmid.	In-house construct.[64] https://benchling.com/s/seq-yXHSYIuEhIy5tKTDgHmb?m=slm-PQajuJX3EDmQqYnXOw9k
pCOLA_CloDF13	pCOLA plasmid with its *ori* site replaced with the *ori* site from a pCDF plasmid.	In-house construct.[64] https://benchling.com/s/seq-PGM8cj3vsbJ2foudD4gS?m=slm-CUFdjxao5BVO2rDYjI9w

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.14.472599; this version posted December 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 3. Parameters of the SAXS data.

Reciprocal space parameters	***Merged SAXS normalized against water scattering merged data***
s min (nm^{-1}) as in Guinier range	0.1116200
Guinier range (points) used for Guinier analysis	48 to 170
Rg, by Guinier analysis / AUTORG, nm	3.61 / 3.61±0.034
I(0) AUTORG, (cm^{-1})	0.0734±0.00022
Rg*s by Guinier analysis	1.33281
Molecular Mass estimations out of SAXS data, kDa	
PRIMUS Qp	107.9
PRIMUS Vc	99.7
PRIMUS Size&Shape	125.3
PRIMUS Absolute scale	101.1
PRIMUS Bayessian inference	101.1
SAXSMoW server, v2.1	124.8
Real-space parameters (GNOM)	
s range (nm^{-1}) used in GNOM	0.1116 to 3.2449
Dmax as parameter in GNOM / DATCLASS / SHANUM, nm	11.4 / 12.3 / 13.4
Rg real GNOM / DATPOROD, nm	3.579±0.005044 / 3.58
I(0) real GNOM / DATPOROD, (cm^{-1})	0.07277±0.0001016 / 0.0728
Porod volume DATPOROD (nm^3)	185.06
SASDB ID	3526 (draft)