Department of Chemistry
Cumulative Examinations

November 10, 2018

You may choose to answer any exam from any area covered in the examination booklet. Each exam may contain multiple parts. You may answer more than one exam but each exam is scored separately and is treated as an individual examination result. Thus, answering parts of two exams with a score of 50% would not yield a 100% grade for this cumulative exam. Instead you would receive 50% on each examination attempted.

This booklet contains five examinations.

1) Analytical Cumulative Examination, Pages 1-3
2) Biochemistry Cumulative Examination, Pages 4-12
3) Inorganic Cumulative Examination, Pages 13-14
4) Organic Cumulative Examination, Pages 15-25
5) Physical Cumulative Examination, Pages 26-27

On your examination booklet:

1) Print your student ID number.
2) Print the Exam Booklet number.
3) Print the question number you are answering.
4) Print the Exam Date.

Do not write your name anywhere on the examination booklet. Each exam will be scored anonymously. If you attempt more than one exam, you must use a separate examination booklet for each examination.

When you complete the examination, return the examination and your answer booklet to the proctor. Exam results will be posted on bulletin board #2B on the north side of the hall near BRWN 2124.
1) The paper age dates corn samples from central Mexico using $^{14}C$, which decays by first order kinetics
   a. Write a first order rate law for the decay of $^{14}C$ (10 pts)
   b. Derive an equation (show derivation) that predicts the corn's $^{14}C$ content at any time (15 pts.)
   c. Derive an equation for the $\frac{1}{2}$ life of $^{14}C$ (10 pts).
   d. What are the units of the decay constant (5 pts)?
   e. Given a decay constant of $3.9 \times 10^{-12}$, show the calculation and value for the half-life of $^{14}C$ (in years) (10 pts).
   f. If the initial $^{14}C$ content was $6 \times 10^5$ atoms/molC and the final was $3 \times 10^3$ atoms/molC what is the age sample in years (10 pts)?

2) $^{14}C$ is produced by neutron capture by N (nitrogen) and decays by beta decay.
   a. Use standard isotope notation and write the production and decay reaction (20 pts).
   b. Calculate the energy change during production and decay of $^{14}C$ (20 pts).

3) One $^{14}C$ measurement technique is liquid scintillation. Briefly describe how this technique works (30 pts).

4) Another $^{14}C$ measurement technique is by accelerator mass spectrometer (AMS), which is about 10,000 times more sensitive than liquid scintillation. During the initial phase of an AMS measuring negative $^{14}C$ ions are produced by sputtering carbon from a target using cesium ions. The negative ions produced are extracted (by electrostatic attraction) from the ion source and sent down the evacuated beam line towards the first magnet. At this point the beam is about 10 micro-amps (mostly the stable isotopes). What is this in ions per second (20 pts)?

5) These negatively ionized carbon atoms pass through focusing devices and an injection magnet before reaching the tandem accelerator where they are accelerated to the positive terminal by a voltage difference of two million volts. What is the velocity of the $^{14}C$ atom under this potential (30 pts)?

6) The $^{14}C$ strikes a gas or solid "stripper" that coverts it into a $^{14}C^3+$, what is the minimum # of electron volts required for this (15 pts)?

7) The $C^3+$ ions are separated by a magnetic sector analyzer and detectors count $^{12}C^3+$, $^{13}C^3+$, $^{14}C^3+$ ions. What mass resolving power is required for this separation (15 pts)?
8) The ion counts are reported as ratios $13R = \frac{13C^{3+}}{12C^{3+}}$, $14R = \frac{14C^{3+}}{12C^{3+}}$. If we are interest only in $^{14}C$ abundances, briefly explain the advantage of measuring all three $^{12}C^{3+}$, $^{13}C^{3+}$, $^{14}C^{3+}$ ions and reporting their ratio (30 pts).

9) The article notes that changes in $^{14}C$ are corrected for $^{13}C$. This is because during a process like photosynthesis, the plant can discriminate between isotopes due to mass differences.
   a. Write the chemical equation for photosynthesis (20 pts).
   b. Assuming the discrimination was a function the kinetic energy of a CO₂ molecule, what would be the discrimination of $^{13}C$ relative to $^{12}C$, and $^{14}C$ relative to $^{12}C$ (20 pts)?

10) Stable isotope abundance ratios $^{13}C/^{12}C$ are often reported in parts per 1000 or $\%\%$ and are reported as a difference ($\delta$) with respect to a reference ratio, for $^{13}C/^{12}C$ this reference ratio is 0.0112372.
   a. If a soy plant has a $^{13}C/^{12}C$ ratio of 0.010934 calculate the $\delta^{13}C(\%\%)$ value with respect to the reference (15 pts).
   b. Corn has a $^{13}C/^{12}C$ ratio of 0.0112483 calculate the $\delta^{13}C(\%\%)$ value with respect to the reference (15pts).
   c. If organic matter in the Wabash river was only from corn and soy residues and has a $\delta^{13}C = -18\%\%$, what fractions would be from corn and soy (15 pts)?

11) What is the acronym for the AMS facility at Purdue University? (Extra credit) (10 pts).
### Ionisation Energies of Carbon

<table>
<thead>
<tr>
<th>Ionisation energy number</th>
<th>Enthalpy / kJ mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>1086.45</td>
</tr>
<tr>
<td>2nd</td>
<td>2252.62</td>
</tr>
<tr>
<td>3rd</td>
<td>4620.47</td>
</tr>
<tr>
<td>4th</td>
<td>6222.69</td>
</tr>
<tr>
<td>6th</td>
<td>3783.0</td>
</tr>
<tr>
<td>8th</td>
<td>4727.1</td>
</tr>
</tbody>
</table>

### Fundamental Subatomic Particles

<table>
<thead>
<tr>
<th>Particle</th>
<th>Symbol</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>electron</td>
<td>e⁻</td>
<td>0.0005486 amu</td>
</tr>
<tr>
<td>proton</td>
<td>p⁺</td>
<td>1.007276 amu</td>
</tr>
<tr>
<td>neutron</td>
<td>n⁰</td>
<td>1.008665 amu</td>
</tr>
</tbody>
</table>

### Constants
- **Avogadro's Number** \( N_A = 6.022 \times 10^{23} \text{ mol}^{-1} \)
- **Faraday Constant** \( F = 96485.33 \text{ C mol}^{-1} \)
- **Atomic Mass Constant** \( 1 \text{ amu} = 1.660538 \times 10^{-27} \text{ kg} \)
- **Molar Gas Constant** \( R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1} \)
- **Molar Gas Constant** \( R = 0.08205746 \text{ L atm K}^{-1} \text{ mol}^{-1} \)
- **Coulomb's Constant** \( k_C = 6.897551 \times 10^9 \text{ N m}^2 \text{ C}^{-2} \)
- **Speed of Light (Vacuum)** \( c = 299792458 \text{ m s}^{-1} \)
- **Boltzmann Constant** \( k_B = 1.38065 \times 10^{-23} \text{ J K}^{-1} \)
- **Charge on a Proton/Electron** \( e = 1.602176 \times 10^{-19} \text{ C} \)
- **Standard acceleration of gravity** \( g = 9.806 \text{ m s}^{-2} \)
- **Rydberg constant** \( R_n = 1.0973731586 \times 10^7 \text{ m}^{-1} \)
- **Planck's Constant** \( h = 6.62607004 \times 10^{-34} \text{ J s} \)

### Thermodynamics
#### Equilibrium
- **pH**: \( \text{pH} = -\log[H^+] \)
- **pOH**: \( \text{pOH} = -\log[OH^-] \)

### Heat Transfer
- **Heat Transfer**: \( q = mc\Delta T \)
- **Enthalpy**: \( \Delta H = H_{\text{products}} - H_{\text{reactants}} \)
- **Entropy**: \( \Delta S = S_{\text{products}} - S_{\text{reactants}} \)
- **Free Energy**: \( \Delta G = \Delta H - T\Delta S \)

1 eV = 1.60218e⁻¹⁹ joules

\[ E = \frac{1}{2}mv^2 \]

\[ E = mc^2 \]

MRP = \( \Delta m/m \)

\( KE = \frac{1}{2}mV^2 \)

\( A = \frac{C}{s} \)
Biochemistry Cumulative Exam
11/10/18 {100 Points Total}

Refer to Cox et al. (2017) Science 358:1019 and your knowledge of biochemistry to answer the following questions. Short answers must be limited to 1-3 sentences.

1. (20 pts) a) Draw the structure of adenosine.

   b) Draw the catalytic mechanism of ADAR. Use “R” to represent the sugar group and draw only the base structure and its chemical transformations. Hint: There are glutamate residues present in the active site to act as proton donors and acceptors, and water is activated by a Zn²⁺ cofactor.

2. (20 pts) a) Draw diagrams that show the difference between NHJ and HDR for gene knockout.

   b) Explain one advantage of HDR over NHJ in nucleic acid editing.

   c) Explain on disadvantage of HDR over NHJ in nucleic acid editing.

   d) Explain one advantage of Crispr systems over either TALEN or ZFN gene editing.

3. (15 pts) a) ADARs can also efficiently edit DNA. Explain whether or not you would expect REPAIRv2 to efficiently edit genomic DNA.

   b) Explain one advantage of editing RNA rather than DNA.

4. (15 pts) a) Explain how PspCas13b mediates RNA knockdown.

   b) Explain the evidence that PspCas13b would be useful for knocking down a broad range of targets.

5. (15 pts) a) Why is the PspCas13b(H133A/H1058A) mutant important to the function of the REPAIR systems?

   b) Why is the ADAR(E487Q) mutant important to the function of the REPAIR systems?

   c) Why is the ADAR(T375G) mutation to the overall usefulness of the REPAIRv2 system?

6. (15 pts) Explain one scenario in which REPAIRv2 would not be an effective treatment strategy for the correction of a disease-related G>A mutation.
Precise nucleic acid-editing technologies are valuable for studying cellular function and as novel therapeutics. Current editing tools, based on programmable nucleases such as the prokaryotic CRISPR-associated nucleases Cas9 (1–4) or Cpf1 (5), have been widely adopted for mediating targeted DNA cleavage, which in turn drives targeted gene disruption through nontargeting end joining (NHEJ) or precise gene editing through template-dependent homology-directed repair (HDR) (6). NHEJ uses host machineries that are active in both dividing and post-mitotic cells and provides efficient gene disruption by generating a mixture of insertion or deletion (indel) mutations that can lead to frame shifts in protein-coding genes. HDR, in contrast, is mediated by host machineries whose expression is largely limited to replicating cells. Accordingly, the development of gene-editing capabilities for post-mitotic cells remains a major challenge. DNA base editors, consisting of a fusion between Cas9 nuclease and cytidine deaminase, can mediate efficient cytidine-to-uridine conversions within a target window and substantially reduce the formation of double-strand break–induced indels (7, 8). However, the potential targeting sites of DNA base editors are limited by the requirement of Cas9 for a protospacer adjacent motif (PAM) at the editing site (9). Here, we describe the development of a precise and flexible RNA base editing technology using the type VI CRISPR-associated RNA-guided ribonuclease (Rnase) Cas3 (10–13).

Cas3 enzymes have two high eukaryotic and prokaryotic nucleotide-binding (HEPN) endoribonuclease domains that mediate precise RNA cleavage with a preference for targets with protospacer flanking sites (PFSs) observed biochemically and in bacteria (10, 11). Three Cas3 protein families have been identified to date: Cas3a (previously known as C2c2), Cas3b, and Cas3c (12, 13). We recently reported that Cas3a enzymes can be adapted as tools for nucleic acid detection (14) as well as mammalian and plant cell RNA knockdown and transcript tracking (15), and observed that the biochemical PFS was not required for RNA interference with Cas3a (15). The programmable nature of Cas3 enzymes makes them an attractive starting point to develop tools for RNA binding and perturbation applications.

The adenosine deaminase acting on RNA (ADAR) family of enzymes mediates endogenous editing of transcripts via hydrolytic deamination of adenosine to inosine, a nucleobase that is functionally equivalent to guanosine in translation and splicing (16, 17). There are two functional ADAR orthologs in mammals, ADAR1 and ADAR2, which consist of N-terminal double-stranded RNA–binding domains and a C-terminal catalytic deamination domain. Endogenous target sites of ADAR1 and ADAR2 contain substantial double-stranded identity, and the catalytic domains require duplexed regions for efficient editing in vitro and in vivo (18, 19). The ADAR catalytic domain is capable of deaminating target adenosines without any protein cofactors in vitro (20). ADAR1 has been found to target mainly repetitive regions, whereas ADAR2 mainly targets nonrepetitive coding regions (17). Although ADAR proteins have preferred motifs for editing that could restrict the potential flexibility of targeting, hyperactive mutants, such as ADAR2(E488Q) (21), relax sequence constraints and increase adenosine-to-inosine editing rates. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, E488Q indicates that glutamic acid at position 488 was replaced by glutamine.) ADARs preferentially deaminate adenosines mispaired with cytidine bases in RNA duplexes (22), providing a promising opportunity for precise base editing. Although previous approaches have engineered targeted ADAR fusions via RNA guides (23–26), the specificity of these approaches has not been reported, and their respective targeting mechanisms rely on RNA-RNA hybridization without the assistance of protein partners that may enhance target recognition and stringency.

We assayed a subset of the family of Cas3 enzymes for RNA knockdown activity in mammalian cells and identified the Cas3b ortholog from Prevotella sp. PS-125 (PopsCas3b) as the most efficient and specific for mammalian cell applications. We then fused the ADAR2 deaminase domain with the E488Q mutation (ADAR2pops) to catalytically inactive PopsCas3b and demonstrated RNA editing for programmable A to I (G) replacement (REPAIR) of reporter and endogenous transcripts as well as disease-relevant mutations. Last, we used a natural mutagenesis scheme to improve the specificity of dCas3b-ADAR2pops fusions in order to generate REPAIR2R with more than 910-fold higher specificity.

**Comprehensive characterization of Cas3 family members in mammalian cells**

We previously developed Leptotrichia wadei Cas3a (LwaCas3a) for mammalian knockdown applications, but it required a monomeric superfolder green fluorescent protein (mSFGFP) stabilization domain for efficient knockdown, and although the specificity was high, knockdown levels were not consistently below 50% (25). We sought to identify a more robust RNA-targeting CRISPR system by characterizing a genetically diverse set of Cas3 family members in order to assess their RNA knockdown activity in mammalian cells (Fig. 1A). We generated mammalian codon-optimized versions of multiple Cas3 proteins, including 21 orthologs of Cas3a, 15 of Cas3b, and seven of Cas3c and cloned them into an expression vector with N- and C-terminal nuclear localization signal (NLS) sequences and a C-terminal mSFGFP to enhance protein stability (table S1). To assay interference in mammalian cells, we designed a dual-reporter construct expressing the independent Gausiasia (Gluo) and Cupriavidus (Chao) luciferases under separate promoters, allowing one luciferase to function as a measure of Cas3 interference activity and the other to serve as an internal control. For each Cas3 ortholog, we designed PPF-compatible guide RNAs, using the Cas3b PPF motifs derived from an ampicillin interference assay (fig. S1, table S2, and supplementary text).
Fig. 1. Characterization of a highly active Cas13b ortholog for RNA knockdown. (A) Schematic of stereotypical Cas13 loci and corresponding CRISPR RNA (crRNA) structure. (B) Evaluation of 19 Cas13a, 15 Cas13b, and seven Cas13c orthologs for luciferase knockdown by use of two different guides. Orthologs with efficient knockdown by using both guides are labeled with their host organism/strain. Values are normalized to nontargeting guide designed against the Escherichia coli LacZ transcript, with no homology to the human transcriptome. (C) PspCas13b and LwaCas13a knockdown activity (as measured by luciferase activity) by using tilting guides against Gluc. Values represent mean ± SEM. Nontargeting guide is the same as in (B). (D) PspCas13b and LwaCas13a knockdown activity (as measured by luciferase activity) by using tilting guides against Gluc. Values represent mean ± SEM. Nontargeting guide is the same as in (B). (E) Expression levels in log_{2}(transcripts per million (TPM)+1)] values of all genes detected in RNA-seq libraries of nontargeting control (x axis) compared with Gluc-targeting condition (y axis) for LwaCas13a (red) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Nontargeting guide is the same as in (B). (F) Expression levels in log_{2}(transcripts per million (TPM)+1)] values of all genes detected in RNA-seq libraries of nontargeting control (x axis) compared with Gluc-targeting condition (y axis) for PspCas13b (blue) and shRNA (black). Shown is the mean of three biological replicates. The Gluc transcript data point is labeled. Nontargeting guide is the same as in (B). (G) Number of significant off-targets from Gluc knockdown for LwaCas13a, PspCas13b, and shRNA from the transcriptome-wide analysis in (E) and (F).

and the 9° H (not C) PFS from previous reports of Cas13a activity (10).

We transfected human embryonic kidney (HEK) 293FT cells with Cas13-expression, guide RNA, and reporter plasmids and then quantified levels of Cas13 expression and the targeted Gluc 48 hours later (Fig. 1B and fig. S2A). Testing two guide RNAs for each Cas13 ortholog revealed a range of activity levels, including five Cas13b orthologs with similar or increased interference across both guide RNAs relative to the recently characterized LwaCas13a (Fig. 1B), and we observed only a weak correlation between Cas13 expression and interference activity (fig. S2, B to D). We selected the top five Cas13b orthologs and the top two Cas13a orthologs for further engineering.

We next tested Cas13-mediated knockdown of Gluc without msfGFP to select orthologs that do not require stabilization domains for robust activity. We hypothesized that Cas13 activity could be affected by subcellular localization, as we previously reported for optimization of LwaCas13a (19). Therefore, we tested the interference activity of the seven selected Cas13 orthologs C-terminally fused to one of six different localization tags without msfGFP. Using the luciferase reporter assay, we identified the top three Cas13b designs with the highest level of interference activity: Cas13b from Prevotella sp. P5-125 (PspCas13b) and S. gordonii C-terminal fused to the HIV Rev nuclear export sequence (NES), and Cas13b from Ruminococcus nautilophilus (RnaCas13b) C-terminally fused to the mitogen-activated protein kinase NES (fig. S3A). To further distinguish activity levels of the top orthologs, we compared the three optimized Cas13b constructs with the optimal LwaCas13a-msfGFP fusion and to short hairpin-mediated RNA (shRNA) for their ability to knock down the endogenous KRAS (V-Ki-ras Kirsten rat sarcoma viral oncoprotein homolog) transcript by using position-matched guides (fig. S3B). We observed the highest levels of interference for PspCas13b (average knockdown, 62.9%) and thus selected this for further comparison with LwaCas13a.

To more rigorously define the activity of PspCas13b and LwaCas13a, we designed position-matched guides tilting along both Gluc and shRNA transcripts and assayed their activity using our luciferase reporter assay. We tested 9 and 20 position-matched guides targeting Gluc and shRNA, respectively, and found that PspCas13b had consistently increased levels of knockdown relative to LwaCas13a (average of 92.3% for PspCas13b versus 40.1% knockdown for LwaCas13a) (Fig. 1, C and D).

Specificity of Cas13 mammalian interference activity

To characterize the interference specificities of PspCas13b and LwaCas13a, we designed a plasmid library of luciferase targets containing single mismatches and double mismatches throughout the target sequence and the three flanking 5' and 3' base pairs (fig. S3C). We transfected HEK293FT cells with either LwaCas13a or PspCas13b, a fixed guide RNA targeting the unmodified target sequence, and the mismatched target library
corresponding to the appropriate system. We then performed targeted RNA sequencing (RNA-seq) of uncleaved transcripts in order to quantify depletion of mismatched target sequences. We found that LwaCas3a and PspCas3b had a central region that was relatively intolerant to single mismatches, extending from base pairs 12 to 26 for the PspCas3b target and 13 to 24 for the LwaCas3a target (fig. S3D). Double mismatches were even less tolerated than single mutations, with little knockdown activity observed over a larger window, extending from base pairs 12 to 29 for PspCas3b and 12 to 27 for LwaCas3a in their respective targets (fig. S3E). Additionally, because there are mismatches included in the three nucleotides flanking the 5' and 3' ends of the target sequence, we could assess FFS constraints on Cas3a knockdown activity. Sequencing showed that almost all FFS combinations allowed robust knockdown, indicating that a FFS constraint for interference in mammalian cells likely does not exist for either enzyme tested. These results indicate that Cas3a and Cas3b display similar sequence constraints and sensitivities against mismatches.

We next characterized the interference specificity of PspCas3b and LwaCas3a across the mRNA fraction of the transcriptome. We performed transcriptome-wide mRNA sequencing to detect significant differentially expressed genes. LwaCas3a and PspCas3b demonstrated robust knockdown of Gluc (Fig. 1E and F) and were highly specific compared with a position-matched shRNA, which showed hundreds of off-targets (Fig. 1G), a finding consistent with our previous characterization of LwaCas3a specificity in mammalian cells (15).

Cas3a-ADAR fusions enable targeted RNA editing

Given that PspCas3b achieved consistent, robust, and specific knockdown of mRNA in mammalian cells, we envisioned that it could be adapted as an RNA-binding platform to recruit RNA-modifying domains, such as ADAR 

To engineer a PspCas3b lacking nuclease activity (dPspCas3b, referred to as dCas3b hereafter), we mutated conserved catalytic residues in the HEPN domains and observed loss of luciferase RNA knockdown (fig. S4A). We hypothesized that the dCas3b-ADAR 

To enhance target adenosine deamination rates, we introduced two additional modifications to our initial RNA editing design: We introduced a mismatched cytidine opposite the target adenosine, which has been previously reported to increase deamination frequency, and fused dCas3b with the deaminase domains of human ADAR1 or ADAR2 containing hyperactivating mutations in order to enhance catalytic activity (ADAR1 

Fig. 2. Engineering dCas3b-ADAR fusions for RNA editing.

(A) Schematic of RNA editing by dCas3b-ADAR 

The crRNA specifies the target site by hybridizing to the bases surrounding the target adenosine, creating a dsRNA structure for editing and recruiting the dCas3b-ADAR 

(B) Schematic of Cypirdina luciferase W85X target and targeting guide design. Deamination of the target adenosine restores the stop codon to the wild-type tryptophan. Spacer length is the region of the guide that contains homology to the target sequence. Mismatch distance is the number of bases between the 3' end of the spacer and the mismatched cytidine. The cytidine mismatched base is included as part of the mismatch distance calculation. (C) Quantification of luciferase activity restoration for (left) dCas3b-ADAR1 

Values are background-subtracted relative to a 30 nt nontargeting guide that is randomized with no sequence homology to the human transcriptome. (D) Schematic of the sequencing window in which A-to-I edits were assessed for Cypirdina luciferase W85X. (E) Sequencing quantification of A-to-I editing for 50 nt guides targeting Cypirdina luciferase W85X. Blue triangle indicates the targeted adenosine. For each guide, the region of duplex RNA is outlined in red. Values represent mean ± SEM. Nontargeting guide is the same as in (C).
To test the activity of dCas13b-ADAR3p, we generated an RNA-editing reporter on Chuc by introducing a nonsense mutation [W86X (UGG → UAG)], which could functionally be repaired to the wild-type codon through A-to-I editing (Fig. 2B) and then be detected as restoration of Cluc lumenase. We evenly tilled guides with spacers 30, 50, 70, or 94 nucleotides (nt) long across the target adenosine so as to determine the optimal guide placement and design (Fig. 2C). We found that dCas13b-ADAR3p-(E488Q) required longer guides to repair the Cluc reporter, whereas dCas13b-ADAR2p-(E468Q) was functional with all guide lengths tested (Fig. 2C). We also found that the hyperactive E488Q mutation improved editing efficiency as wild-type ADAR2p displayed reduced luciferase restoration (Fig. S4B). From this demonstration of activity, we chose dCas13b-ADAR2p-(E488Q) for further characterization and designated this system RNA Editing for Programmable A to I Replacement with 1 (REPAIRv1).

To validate that restoration of luciferase activity was due to bona fide editing events, we directly measured REPAIRv1-mediated editing of Cluc transcripts via reverse transcription and targeted next-generation sequencing. We tested 30- and 50-nt spacers among the target site and found that both guide lengths resulted in the expected A-to-I edit, with 50-nt spacers achieving higher editing percentages (Fig. 2, D and E, and fig. S4C). We also observed that 50-nt spacers had an increased propensity for editing at non-targeted adenosines within the sequencing window, likely because of increased regions of duplexed RNA (Fig. 2E and fig. S4C).

We next targeted an endogenous gene, PPIB. We designed 50-nt spacers tilling PPIB and found that we could edit the PPIB transcript with up to 28% editing efficiency (fig. S4D). To test whether REPAIR could be further optimized, we modified the linker between dCas13b and ADAR2p-(E488Q) (fig. S4E and table S3) and found that linker choice modestly affected luciferase activity restoration. Additionally, we tested the ability of dCas13b and guide alone to mediate editing events, finding that the ADAR3p is required for editing (fig. S5, A to D).

Defining the sequence parameters for RNA editing

Given that we could achieve precise RNA editing at a test site, we wanted to characterize the sequence constraints for programming the system against any RNA target in the transcriptome. Sequence constraints could arise from dCas13b-targeting limitations, such as the PFS, or from ADAR sequence preferences (29). To investigate PFS constraints on REPAIRv1, we designed a plasmid library that carries a series of four randomized nucleotides at the 5’ end of a target site on the Cluc transcript (Fig. 3A). We targeted the center adenosine within either a UAG or AAC motif and found that for both motifs, all PFSs demonstrated detectable levels of RNA editing, with a majority of the PFSs having >50% editing at the target site (Fig. 3B). Next, we sought to determine whether the ADAR2p-(E488Q) in REPAIRv1 had any sequence constraints immediately flanking the targeted base, as has been reported previously for ADAR2p-(E488Q). We tested every possible combination of 5’ and 3’ flanking nucleotides directly surrounding the target adenosine (Fig. 3C) and found that REPAIRv1 was capable of editing all motifs (Fig. 3D). Lastly, we analyzed whether the identity of the base opposite the target A in the spacer sequence affected editing efficiency and found that an A-C mismatch had the highest luciferase restoration, in agreement with previous reports of ADAR2 activity, with A-G, A-U, and A-A having drastically reduced REPAIRv1 activity (fig. S5E).
truncations tested were still functional and able to restore luciferase signal (Fig. S7), and the largest truncation, C-terminal A894–1090 (total size of the fusion protein, 4192 bp) was small enough to fit within the packaging limit of AAV vectors.

Transcriptome-wide specificity of REPAIRv1

Although RNA knockdown with PspCas33b was highly specific in our luciferase titling experiments, we observed off-target adenosine editing within the guide-target duplex (Fig. 2B). To see whether this was a widespread phenomenon, we tiled an endogenous transcript, KRAS, and measured the degree of off-target editing near the target adenine (Fig. S5A). We found that for KRAS, although the on-target editing rate was 23%, there were many sites around the target site that also had detectable A-to-I edits (Fig. S5B).

Because of the observed off-target editing within the guidetarget duplex, we initially evaluated transcriptome-wide off-targets by performing RNA-seq on all human T-cells with 12x coverage. Of all the editing sites across the transcriptome, the on-target editing site had the highest editing rate, with 89% A-to-I conversion. We also found that there was a substantial number of A-to-I off-target events, with 1732 off-targets in the targeting guide condition and 925 off-targets in the nontargeting guide condition, with 828 off-targets shared between the targeting and nontargeting guide conditions (Fig. 5, C and D). Given the high number of overlapping off-targets between the targeting and nontargeting guide conditions, we reasoned that the off-targets may arise from ADAR2. To test this hypothesis, we repeated the Cas9-targeting experiment, this time comparing transcriptome changes for REPAIRv1 with a targeting guide, REPAIRv1 with a nontargeting guide, and REPAIRv1 alone, or ADAR2(E488Q) alone (Fig. S8). We found differentially expressed genes and off-target editing events in each condition (Fig. S8, B and C). There was a high degree of overlap in the off-target editing events between ADAR2(E488Q) and all REPAIRv1 off-target
edits, supporting the hypothesis that REPAIR off-target edits are driven by dCas13b-independent ADAR2p(E488Q) editing events (fig. S8D).

Next, we sought to compare two RNA-guided ADAR systems that have been described previously (fig. S8A). The first uses a fusion of ADAR2p to the small viral protein lambda N (32N), which binds to the Box-B RNA hairpin (24). A guide RNA with double BoxB hairpins guides ADAR2p(E488Q) to edit sites encoded in the guide RNA (25). The second design uses full-length ADAR2 (ADAR2) and a guide RNA with a hairpin that the double-strand RNA (dsRNA)-binding domains (dsRBDs) of ADAR2 recognize (23, 24). We analyzed the editing efficiency of these two systems compared with REPAIRv1 and found that the BoxB-ADAR2 and full-length ADAR2 systems demonstrated 50 and 34.5% editing rates, respectively, compared with the 80% editing rate achieved by REPAIRv1 (fig. S9). Additionally, the BoxB and full-length ADAR2 systems created 3814 and 66 observed off-targets, respectively, in the targeting guide conditions, compared with the 2111 off-targets in the REPAIRv1 targeting guide condition. All the conditions with the two ADAR2p-based systems (REPAIRv1 and BoxB) showed a high percentage of overlap in their off-targets, whereas the full-length ADAR2 system had a largely distinct set of off-targets (fig. S9F). The overlap in off-targets between the targeting and nontargeting conditions and between REPAIRv1 and BoxB suggests that ADAR2p drives off-target independent of dCas13 targeting (fig. S9F).

Improving specificity of REPAIR through rational protein engineering

To improve the specificity of REPAIRv1, we used structure-guided protein engineering of ADAR2p(E488Q). Because of the guide-independent nature of the off-targets, we hypothesized that destabilizing ADAR2p(E488Q)-RNA binding would selectively decrease off-target editing, but maintain on-target editing because of increased local concentration from dCas13b tethering of ADAR2p(E488Q) to the target site. We mutated residues in ADAR2p(E488Q) previously determined to contact the duplex region of the target RNA (fig. 6A) (19). To assess efficiency and specificity, we tested 17 single mutants with both targeting and nontargeting guides, under the assumption that background luciferase restoration on the nontargeting condition would be indicative of broader off-target activity. We found that mutations at the selected residues had substantial effects on the luciferase activity for targeting and nontargeting guides (fig. 6, A and B, and fig. S10A). A majority of mutants either significantly improved the luciferase activity for the targeting guide or increased the ratio of targeting to nontargeting guide activity, which we termed the specificity score (fig. 6, A and B).

We selected a subset of these mutants (fig. 6B) for transcriptome-wide specificity profiling by next-generation sequencing. As expected, off-targets measured from transcriptome-wide sequencing correlated with our specificity score (fig. S10B) for mutants. We found that with the exception of ADAR2p(E488Q/R456E), all sequenced REPAIRv1 mutants could effectively edit the reporter transcript (fig. 6C), with many mutants showing reduction in the number of off-targets (fig. 6C and figs. S10C and S11). We further explored motifs surrounding off-targets for the various specificity mutants and found that REPAIRv1 and most of the engineered variants exhibited a strong 3′ G preference for their off-target edits, which is in agreement with the characterized ADAR2p motif (fig. S11A) (26).

We focused on the mutant ADAR2p(E488Q/T375C) because it had the highest percent editing of the four mutants with the lowest numbers of transcriptome-wide off-targets—and termed it REPAIRv2. Compared with REPAIRv1, REPAIRv2 exhibited increased specificity, with a reduction from 18,385 to 20 transcriptome-wide off-targets with high-coverage sequencing (12x coverage) and 10 ng of REPAIR vector transfected (fig. 6D). In the region surrounding the targeted adenosine in Cluc, REPAIRv2 also had reduced off-target editing, visible in sequencing traces (fig. 6E). In motifs derived from the off-target sites, REPAIRv1 presented a strong preference toward 3′ G but showed off-target editing for all motifs (fig. S12B); by contrast, REPAIRv2 only edited the strongest off-target motifs (fig. S12C). The distribution of edits on transcripts was heavily skewed for REPAIRv1, with highly edited genes having more than 60 edits (fig. S13A), whereas REPAIRv2 only edited one transcript (EF1AT) multiple times (fig. S13B). REPAIRv1 off-target edits were predicted to result in numerous variants, including 1000 missense base changes (fig. S13C), with 93 events in genes related to...
cancer processes (Fig. S3D). In contrast, REPAIRv2 only had six predicted nonsense changes (Fig. S3E), none of which were in cancer-related genes (Fig. S3F). Analysis of the sequence surrounding off-target edits for REPAIRv1 or v-2 did not reveal homology to guide sequences, suggesting that off-targets are likely dCas13b-independent (Fig. S4), which is consistent with the high overlap of off-targets between REPAIRv1 and the ADAR2 deaminase domain (Fig. S8D). To directly compare REPAIRv2 with other programmable ADAR systems, we repeated our Cluc-targeting experiments with all systems at two different dosages of ADAR vector, finding that REPAIRv2 had comparable on-target editing with that of 8X3 and ADAR2.

**Fig. 6. Rational mutagenesis of ADAR2 to improve the specificity of REPAIRv1.**

(A) Quantification of luciferase signal restoration (on-target score, red boxes) by various dCas13-ADAR2 mutants as well as their specificity score (blue boxes) plotted along a schematic of the contacts between key ADAR2 deaminase residues and the dsRNA target (the target strand is shown in gray; the nontarget strand is shown in red). All deaminase mutations were made on the dCas13-ADAR2(ED488Q) background. The specificity score is defined as the ratio of the luciferase signal between targeting guide and nontargeting guide conditions. [Schematic of ADAR2 deaminase domain contacts with dsRNA is adapted from (20).] (B) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants versus their specificity score. Nontargeting guide is the same as in Fig. 2C. (C) Quantification of on-target editing and the number of significant off-targets for each dCas13-ADAR2(ED488Q) mutant by transcriptome-wide sequencing of mRNAs. Values represent mean ± SEM. Nontargeting guide is the same as in Fig. 2C. (D) Transcriptome-wide sites of significant RNA editing with (top) REPAIRv1 and (bottom) REPAIRv2, with a guide targeting a preterminator site in Cluc. The on-target Cluc site (254 A→I) is highlighted in orange. Ten nanograms of REPAIR vector was transfected for each condition. (E) Representative RNA-seq reads surrounding the on-target Cluc-editing site (254 A→I; blue triangle) highlighting the differences in off-target editing between (top) REPAIRv1 and (bottom) REPAIRv2. A-to-I edits are highlighted in red; sequencing errors are highlighted in blue. Gaps reflect spaces between aligned reads. Nontargeting guide is the same as in Fig. 2C. (F) RNA editing with REPAIRv1 and REPAIRv2, with guides targeting an out-of-frame UAG site in the endogenous KRAS and PP1B transcripts. The on-target editing fraction is shown as a sideways bar chart on the right for each condition row. For each guide, the region of duplex RNA is outlined in red. Values represent mean ± SEM. Nontargeting guide is the same as in Fig. 2C.
but with substantially fewer off-target editing events at both dosages (fig S15). REPAIRv2 had enhanced specificity compared with REPAIRv1 at both dosages (fig. S15B), a finding that also extended to two guides targeting distinct sites on PPIB (fig. S16, A to D). It is also worth noting that in general, the lower-dosage condition (10 ng REPAIR vector) had fewer off-targets than that of the higher dosage condition (150 ng REPAIR vector) (fig. S8).

To assess editing specificity with greater sensitivity, we sequenced the low-dosage condition (10 ng of transfected DNA) of REPAIRv1 and v2 at much higher sequencing depth (22x coverage of the transcriptome). Increased numbers of off-targets were found at higher sequencing depths corresponding to detection of rarer off-target events (fig. S17). Furthermore, we speculated that different transcriptome states could also potentially alter the number of off-targeting events. Therefore, we tested REPAIRv2 activity in the osteosarcoma U2OS cell line, observing six and seven off-targets for the targeting and non-targeting guide, respectively (fig. S18).

We targeted REPAIRv2 to endogenous genes to test whether the specificity-enhancing mutations reduced nearby edits in target transcripts while maintaining high efficiency on-target editing. For guides targeting either KRAS or PPIB, we found that REPAIRv2 had no detectable off-target edits, unlike REPAIRv1, and could effectively edit the on-target adenine at efficiencies of 27% (KRAS) or 18% (PPIB) (fig. 6F). This specificity extended to additional target sites, including regions that demonstrate high levels of background in nontargeting conditions for REPAIRv1, such as other KRAS or PPIB target sites (fig. S19).

Overall, REPAIRv2 eliminated off-targets in duplexed regions around the edited adenosine and showed dramatically enhanced transcriptome-wide specificity.

Discussion

We show here that the RNA-guided RNA-targeting type VI-B CRISPR effector Cas13b is capable of highly efficient and specific RNA knockdown, providing the basis for improved tools for integrating essential genes and noncoding RNA as well as controlling cellular processes at the transcript level. Catalytically inactive Cas13b (dCas13b) retains programmable RNA-binding capability, which we leveraged here by fusing dCas13b to the adenosine deaminase domain of ADAR2 to achieve precise A-to-I edits, a system we term REPAIRv1. Further engineering of the system produced REPAIRv2, which has dramatically higher specificity than previously described RNA-editing platforms (25, 29) while maintaining high levels of on-target efficacy.

Although Cas13b exhibits high fidelity, our initial results with dCas13b-ADAR2(39-488Q) fusions revealed a substantial number of off-target RNA editing events. To address this, we used a rational mutagenesis strategy to vary the ADAR2(39-488Q) residues that contact the RNA duplex, identifying a variant, ADAR2(39-488Q)T375G, that is capable of precise, efficient, and highly specific editing when fused to dCas13b. Editing efficiency with this variant was comparable with or better than that achieved with two currently available systems, Bso3f-ADAR2(39-488Q) or ADAR2 editing. Moreover, the REPAIRv2 system created only 20 observable off-targets in the whole transcriptome, which is at least an order of magnitude better than both alternative editing technologies. Although it is possible that ADAR could deaminate adenosine bases on the DNA strand in RNA-DNA heteroduplexes (20), it is unlikely to do so in this case because Cas13b does not bind DNA efficiently and because REPAIR is cytoplasmically localized. Additionally, the lack of homology of off-target sites to the guide sequence and the strong overlap of off-targets with the ADAR2(39-488Q)-only condition suggest that off-targets are not mediated by off-target guide binding. Deeper sequencing and novel insolation enrichment methods could further refine our understanding of REPAIR specificity in the future.

The REPAIR system offers many advantages compared with other nucleic acid-editing tools. First, the exact target site can be encoded in the guide by placing a cytidine within the guide across from the desired adenosine to create a favorable A-C mismatch ideal for ADAR-editing activity. Second, Cas3 has no targeting sequence constraints, such as a PFS or PAM, and no motif preference surrounding the target adenine, allowing any adenosine in the transcriptome to be potentially targeted with the REPAIR system. The lack of motif for ADAR editing, in contrast with previous literature, is likely due to the increased local concentration of REPAIR at the target site owing to Cas13b binding. DNA base editors can target either the sense or antisense strand, whereas the REPAIR system is limited to transcriptioned sequences, constraining the total number of possible editing sites. However, because of the less constrained nature of targeting with REPAIR, this system can effect more edits within ClinVar (fig. 4C) than Cas9-DNA base editors. Third, the REPAIR system directly deaminates target adenosines to inosines and does not rely on endogenous repair pathways to generate desired editing outcomes. Therefore, REPAIR should be able to mediate efficient RNA editing even in post-mitotic cells such as neurons. Fourth, in contrast to DNA editing, RNA editing is transient and can be more easily reversed, allowing the potential for temporal control over editing outcomes. The transient nature of REPAIR-mediated edits will likely be useful for treating diseases caused by temporary changes in cell state, such as local inflammation, and could also be used to treat disease by modifying the function of proteins involved in disease-related signal transduction. For instance, REPAIR editing would allow the reading of some serine, threonine, and tyrosine residues that are the targets of kinases (fig. S20). Phosphorylation of these residues in disease-relevant proteins affects disease progression for many disorders, including Alzheimer’s disease and multiple neurodegenerative conditions (30). REPAIR might also be used to transiently or even chronically change the sequence of expressed, risk-modifying G-to-A variants so as to decrease the chance of entering a disease state for patients. For instance, REPAIR could be used to functionally mimic A-to-G alleles of IFIH1 that protect against autoimmune disorders such as type 1 diabetes, immunoglobulin A deficiency, psoriasis, and systemic lupus erythematosus (31, 32).

The REPAIR system provides multiple opportunities for additional engineering. Cas13b possesses pre-CRISPR-RNA (crRNA) processing activity (19), allowing for multiplex editing of multiple variants—any one of which alone may not affect disease, but together might have additive effects and disease-modifying potential. Extension of our rational design approach, such as combining promising mutations and directed evolution, could further increase the specificity and efficiency of the system, while unbiased screening approaches could identify additional residues for improving REPAIR activity and specificity.

Currently, the base conversions achievable by REPAIR are limited to generating inosine from adenosine; additional fusions of Cas9 with other catalytic RNA editing domains, such as APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like), could enable cytidine-to-uridine editing. Additionally, mutagenesis of ADAR could relax the substrate preference to target cytidine, allowing for the enhanced specificity conferred by the duplexed RNA substrate requirement to be exploited by C-to-U editors. Adenine-to-inosine editing on DNA substrates may also be possible with catalytically inactive DNA-targeting CRISPR effectors, such as Cas9 or dCpf1, either through formation of DNA-RNA heteroduplex targets (20) or mutagenesis of the ADAR domain.

We have demonstrated the use of the PspCas13b enzyme as both an RNA-knockdown and RNA-editing tool. The dCas13b platform for programmable RNA binding has many applications, including live transcript imaging, splicing modification, targeted localization of transcripts, pulldown of RNA-binding proteins, and epitranscriptomic modifications. We used dCas13b to create REPAIR, adding to the existing suite of nucleic acid-editing technologies. REPAIR provides a new approach for treating genetic disease or mimicking protective alleles and establishes RNA editing as a useful tool for modifying genetic function.

REFERENCES AND NOTES

   (a, 10pt) Define the first order Jahn–Teller effect and provide an example.
   (b, 15 pt) Define the second order Jahn–Teller effect based on your reading.
   (c, 35 pt) Use group theory to develop an orbital correlation diagram between the planar and *trans*-pyramided alkene analogue of heavy group 14 elements, $R_2E=ER_2$ ($E = Ge$, Sn and Pb), and explain why the second order Jahn–Teller effect is observed for these molecules (and not for $R_2C=CR_2$). *Hint:* you need to account for all the valence electrons & orbitals involved in E-E bonding. Feel free to use any of the character tables provided below.

2. (40 pt) Recently, Wu and coworkers reported isolation and spectroscopic characterization of carbonyl compounds of alkaline earth metals $M(CO)_n (M = Ca, Sr and Ba)$, which, surprisingly, obey the 18 electron rule.
   a. (10 pt) what is $n$ and why?
   b. (10 pt) What are the possible geometries of these compounds and the corresponding symmetry group for each of them?
   c. (15 pt) A single CO stretch was observed for each compound with the frequency of 1987, 1995 and 2014 cm$^{-1}$ for M as Ca, Sr and Ba, respectively. Provide a concise rationale based on the nature of M-CO bonding. *Note:* free CO stretch frequency is 2143 cm$^{-1}$.
   d. (5 pt) Which of the possible geometries you suggested is the most probable on the basis of single CO stretch being observed?
<table>
<thead>
<tr>
<th>$C_{2v}$</th>
<th>E</th>
<th>$C_2$</th>
<th>$\sigma(xz)$</th>
<th>$\sigma(yz)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_1$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_2$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$C_{2h}$</th>
<th>E</th>
<th>$C_2(z)$</th>
<th>i</th>
<th>$\sigma_h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_g$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$B_g$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_u$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_u$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D_4$</th>
<th>E</th>
<th>$2C_1(z)$</th>
<th>$C_2(z)$</th>
<th>$2C'_2$</th>
<th>$2C''_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_1$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_2$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D_{2h}$</th>
<th>E</th>
<th>$C_2(z)$</th>
<th>$C_2(y)$</th>
<th>$C_2(x)$</th>
<th>i</th>
<th>$\sigma(xy)$</th>
<th>$\sigma(xz)$</th>
<th>$\sigma(yz)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_h$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$B_{1h}$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_{2h}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_{3h}$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$A_u$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_{1u}$</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$B_{2u}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$B_{3u}$</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$D_{2d}$</th>
<th>E</th>
<th>$2S_4$</th>
<th>$C_2(z)$</th>
<th>$2C'_2$</th>
<th>$2S_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$A_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_1$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$B_2$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Organic Cumulative Examination

November 10th, 2018

The following questions are based on the paper from Zakarian, et al. (JACS 2018, 140, 6027).

i) Briefly describe what is isotope dilution mass spectrometry and how it works (10 points).

ii) Provide a plausible reaction mechanism to account for the transformation of $^{15}\text{NH}_3$ to $[^{15}\text{N}_2] \cdot 4$ (see Scheme 1; 10 points).

iii) Provide a plausible reaction mechanism to account for the transformation of $[^{15}\text{N}_2] \cdot 4$ to $[^{15}\text{N}_2] \cdot 5$ (see Scheme 1; 10 points).

iv) Provide a plausible reaction mechanism for the Sonogashira coupling (from $[^{15}\text{N}_2] \cdot 5$ to $[^{15}\text{N}_2] \cdot 7$; see Scheme 2; 10 points).

v) Provide a plausible reaction mechanism to account for the transformation of $[^{15}\text{N}_2] \cdot 11$ to $[^{15}\text{N}_2] \cdot 12$ (see Scheme 2; 10 points).

vi) Provide a plausible reaction mechanism to account for the transformation of $[^{15}\text{N}_2] \cdot 12$ to $[^{15}\text{N}_2] \cdot 13$ (see Scheme 2; 10 points).

vii) Provide a plausible reaction mechanism to account for the transformation of $[^{15}\text{N}_2] \cdot 13$ to $[^{15}\text{N}_2] \cdot 16$ (see Scheme 2; 10 points).

The following questions are based on the paper from Magauer, et al. (JACS 2018, 140, 8444).

viii) Provide a plausible reaction mechanism to account for the transformation of 11 to 7 (see Scheme 2; 10 points).

ix) Provide a plausible reaction mechanism to account for the transformation of 7 to 12 and 13 (see Scheme 2; 10 points).

x) Provide a plausible reaction mechanism to account for the transformation of 16 to 6. Note, you need to provide the chemical structure of the Dess-Martin periodinane and explain the Z/E selectivity of the Wittig olefination (see Scheme 3; 10 points).
Short Total Synthesis of $[^{15}\text{N}_5]$-Cylindrospermopsins from $^{15}\text{NH}_4\text{Cl}$ Enables Precise Quantification of Freshwater Cyanobacterial Contamination

Artur K. Mailyan, † Joanna L. Chen, † Weiwei Li, † Arturo A. Keller, ‡ Shawn M. Sternisha, § Brian G. Miller, § and Armen Zakarian* †‡

† Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93110, United States
‡ Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, United States
§ Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32310, United States

Supporting Information

ABSTRACT: Fresh water cyanobacterial algal blooms represent a major health risk because these organisms produce cylindrospermopsin, a toxic, structurally complex, zwitterionic uracil-guanidine alkaloid recognized by the EPA as a dangerous drinking water contaminant. At present, the ability to detect and quantify the presence of cylindrospermopsin in water samples is severely hampered by the lack of an isotopically labeled standard for analytical mass spectrometry. Herein, we present a concise, scaled total synthesis of $^{15}\text{N}$ cylindrospermopsin from $^{15}\text{N}$ ammonium chloride, which leverages a unique stereoselective intramolecular double conjugate addition step to assemble the tricyclic guanidine core. In addition to providing the first pure isotopically labeled probe for precise quantification of this potent biotoxin in fresh water sources, our results demonstrate how unique constraints associated with isotope incorporation compel novel solutions to synthesis design.

INTRODUCTION

In recent years, cyanobacterial algal blooms have occurred with increasing frequency near population centers worldwide, causing significant socioeconomic impact and risk to human health. 1 Cylindrospermopsin (CYN)2 and its naturally occurring 7-epi-3 and 7-deoxy-4 congeners are potent biotoxins released into fresh water sources during such blooms5 (Figure 1a,b). In June 2015, the US Environmental Protection Agency released a drinking water health advisory on cylindrospermopsins detailing this emerging environmental threat.6 In addition to the acute toxicity risks, exposure to cylindrospermopsin and its derivatives is linked to hepatotoxicity, gastrointestinal inflammation, hemorrhages, anorexia, and liver failure.7

High-performance liquid chromatography coupled with mass spectrometry is the preferred method for cylindrospermopsin detection in environmental samples.8–13 In particular, isotope-dilution mass spectrometry (IDMS)14 provides an attractive quantification method, as it uses an isotope of the analyte as an internal reference standard to ensure high sensitivity and precision in the measurement. The application of IDMS to cylindrospermopsin analysis, however, is hampered by the lack of isotopically labeled cylindrospermopsins. To solve this important problem, we coupled recent advances in synthetic chemistry and enzymatic bio-oxidation to prepare 7-epi- and 7-deoxy-cylindrospermopsin from $[^{14}\text{N}]$-ammonium chloride, achieving >99.5% isotope incorporation into all five nitrogen atoms of these complex alkaloid natural products.

To facilitate mass spectrometry data interpretation, we sought an isotope of CYN with a mass difference of at least 5 atomic units relative to the most abundant natural isotope. We considered isotopes of hydrogen, carbon, nitrogen, and oxygen for this purpose. To ensure structural integrity of the analytical standard, isotope insertion at nonexchangeable positions throughout the synthesis is essential. Based on these two key requirements, we chose to pursue the total synthesis of $[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5]$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5}$-$[^{15}\text{N}_5$
Several principal considerations factored into the synthesis planning. Clearly, the effective introduction of the $^{15}$N isotope becomes a major factor. Compared to a typical total synthesis effort, the payoff of a convergent versus linear, atom-by-atom assembly is now substantially enhanced because all but the most elemental sources of the $^{15}$N isotope ($^{15}$NH$_4$Cl or aqueous $^{15}$NH$_3$) become impractical. Thus, a convergent design with preassembled intermediates containing several $^{15}$N isotopes is essential. The common metric of synthetic efficiency, namely, minimizing the number of individual steps based on robust, reproducible chemical transformations, becomes especially critical given the costs of the starting isotopes. Finally, all interceded transformations must ensure retention of the isotopic label.

**RESULTS AND DISCUSSION**

**Synthesis Design.** Following the logic outlined above, we formulated a synthesis plan depicted in Figure 1c. $[^{15}N]_2$-Guandine and $[^{15}N]_2$-urea, each derived from $^{15}$NH$_4$Cl, serve as precursors of the tricyclic guanidine and uracil subunits of cylindropermopsin, respectively. The synthesis plan requires a late stage conversion of the electron-withdrawing group $Z$ in intermediate ii to uracil and C$\equiv$H oxidation at C7. The electron-withdrawing group $Z$ enables a key simplification to diene ii by disconnection of two C$\equiv$N bonds in tricyclic guanidine i at C8 and C10. Stereoselective intramolecular addition of guanidine to diene, although lacking precedent, is expected to ensure rapid assembly of the tricyclic guanidine subunit of cylindropermopsin. Alkynyl iii was intended as a precursor to diene ii via a transition-metal-mediated rearrangement. As an essential element of the plan, homoallylic alcohol iv serves as the substrate for a hydroxyl-directed stereoselective delivery of free $[^{15}N]_2$-guanidine that we previously developed for this application.

The success of this plan rests on identifying a suitable substituent $Z$, which must first activate the diene for addition of guanidine, then facilitate C7 hydroxylation, and eventually be converted to uracil. Although several options can be considered, heterocyclic substituents such as 2,4-dimethoxy-2-pyrimid-6-yl appeared to be especially enticing as precursors of uracil. Various electron-deficient heterocyclic groups are known to stabilize carbanions, and activation of double bonds by substitution with heterocycles for nucleophilic addition has been reported. What remained uncertain was whether diene activation for double addition of guanidine, a centerpiece

---

**Scheme 1.** Preparation of $[^{15}N]_2$-2,4-Dimethoxy-6-bromopyrimidine and $[^{15}N]_2$-Guandine Hydrochloride from $^{15}$NH$_4$Cl, the Key Precursors Constituting the Source of All $^{15}$N Isotopes in Cylindropermopsins
Scheme 2. Incorporation of All Five $^{15}$N Isotopes into Advanced Intermediate $[^{15}\text{N}]_{5}$-16 from $[^{15}\text{N}]_{2}$-2,4-Dimethoxy-6-bromopyrimidine $[^{15}\text{N}]_{5}$-S and $[^{15}\text{N}]_{2}$-Guanidine Hydrochloride

of our strategy, was feasible. After a careful analysis supported by experimentation, we identified 2,4-dimethoxyprismid-6-y1 as group $Z$ because of its potential to meet these criteria most effectively.

Preparation of $[^{15}\text{N}]$-Labeled Precursors. While the initial validation of the entire total synthesis was carried out with the natural isotope, we performed the scaled synthesis starting with $^{15}\text{NH}_{4}\text{Cl}$ as depicted in Scheme 1. $[^{15}\text{N}]$-Ammonium chloride proved to be the most economical source of dry liquid $[^{15}\text{N}]$-ammonia on the laboratory scale, affording >95% yields of the reagent reproducibly upon treatment with NaOH in solid phase as long as the scale exceeds 0.5 g. $[^{15}\text{N}]_{2}$-2,4-Dimethoxy-6-bromopyrimidine $[^{15}\text{N}]_{5}$-S was obtained by the initial conversion of $^{15}\text{NH}_{3}$ to urea, its condensation with diethyl malonate to $[^{15}\text{N}]_{2}$-barbituric acid $[^{15}\text{N}]_{2}$-4, treatment with POBr$_3$, and methanolation with 2 equiv of NaOMe. In contrast to the synthesis of urea, the reaction of $^{15}\text{NH}_{3}$ with diphenyl thiocarbonate gave ammonium thiocyanate, pyrolysis of which was required to access $[^{15}\text{N}]$-thiourea. The preparation of $[^{15}\text{N}]$-guanidine hydrochloride was accomplished in four steps by $S$-methylation, treatment with Boc$_2$O, displacement of the MeS group with $^{15}\text{NH}_{3}$ in chloroform, and complete hydrolysis with 12 M HCl at reflux. The resulting $[^{15}\text{N}]_{2}$-guanidine hydrochloride and $[^{15}\text{N}]_{5}$-S constitute the source of all $^{15}$N isotopes in the target alkaloid.

Total Synthesis of $[^{15}\text{N}]_{2}$-Cylindropermospins. Advancing the synthesis of $[^{15}\text{N}]_{2}$-cylindropermospin, bromopyrimidine $[^{15}\text{N}]_{5}$-S was subjected to Sonogashira coupling with 5-heynyl-1-ol (Scheme 2). After Swern oxidation of $[^{15}\text{N}]_{2}$-7 to $[^{15}\text{N}]_{2}$-8, efficient enantioselective crotylation was achieved using Leighton's reagent 95 (99% yield, dr 10:1, er 97:3). To shorten the synthesis, we explored direct conversion of $[^{15}\text{N}]_{5}$-7 to $[^{15}\text{N}]_{2}$-10 using a powerful method developed by Kröse and co-workers. Although the enantiocontrol was excellent (er 98:2), lower diastereoselectivity (4:1) and yield (21%) thwarted its application for the scaled synthesis of $[^{15}\text{N}]_{2}$-10 (see the Supporting Information for additional details). The next objective was the appellation of guanidine unit to the unactivated terminal alkene in $[^{15}\text{N}]_{2}$-10. This was achieved by a two-stage approach described previously. First, free $[^{15}\text{N}]_{2}$-guanidine was attached to the hydroxyl group by a carbonyl linker in good yield using 1.2 equiv of $^{15}$N-labeled reagent. Second, electrophilic cycloguanidation was performed in the presence of N-iodosuccinimide, delivering $[^{15}\text{N}]_{2}$-12 after treatment with 4-methoxybenzenesulfonyl chloride (MscI) in 71% yield and 2.5:1 dr favoring the requisite diastereomer. Allylone $[^{15}\text{N}]_{2}$-12 served as an effective precursor of the 1,3-diene required for the key double addition of guanidine. We identified (Ph$_3$P)$_2$Pd and benzoic acid at reflux in dioxane as the optimal reaction system for the challenging
Scheme 3. Total Synthesis of $[^{15}\text{N}]$-CYN, $[^{15}\text{N}]$-7-epi-CYN, and 7-Deoxy-CYN by (a) C7 Bio-hydroxylation with CyrI and (b) Stereoselective Chemical Oxidation via Lithiation at C7

a. Chemoenzymatic oxidation [most direct]

1. 12 M aq. HCl
2. SO$_2$-Py, DMF

95% yield

FeSO$_4$, ascorbic acid
2-ketoglutarate, 28 °C, 2 h

95% yield, dr 25:1
8.6 mg scale

b. Chemical oxidation [amenable to gram scale]

[Image of chemical structures and reactions]

substrate containing two heterocyclic substituents. Diene $[^{15}\text{N}]$-13, obtained in 75% yield as a 6:1 mixture of E and Z isomers, was now poised to implement the key cyclization to the tricyclic guanidine ring system of CYNs. Simple heating of $[^{15}\text{N}]$-13 in 14% methanol in acetonitrile (v/v) in the presence of Pr$_2$NEt at reflux resulted in rapid methanolation to $[^{15}\text{N}]$-14. Remarkably, the key addition of guanidine to the diene took place concurrently with methanolation under these mild reaction conditions, and subsequent removal of methyl carbonate in $[^{15}\text{N}]$-14 was accomplished with NaOMe in situ, affording $[^{15}\text{N}]$-16. The success of this process was essential to ensure the brevity of the total synthesis and demonstrated the activating properties of the 2,6-dimethoxy-6-pyrimidinyl substituent for addition of nitrogen nucleophiles to alkenes.

To address one of the most challenging phases of the synthesis, the stereospecific hydroxylation of the C7 methylene group of $[^{15}\text{N}]$-7-deoxy-CYN ([$^{15}\text{N}]$-3), we explored enzymatic biooxidation. This conversion represents the final step of CYN biosynthesis and is catalyzed by CyrI, an α-ketoglutarate-dependent nonheme iron oxygenase. We produced pure, recombinant CyrI (Oscillatoria strain PCC7926) in E. coli and applied the enzyme in the direct oxidation of ($[^{15}\text{N}]$)-3, which was readily generated from $[^{15}\text{N}]$-16 by acidic hydrolysis (12 M HCl, reflux) and O-sulfonation (SO$_2$-Py, DMF, 95% overall yield) (Scheme 3a). The enzyme showed high selectivity and did not accept a substrate lacking the O12 sulfate, suggesting that O-sulfonation must precede C7 hydroxylation in the biosynthesis of CYN. Importantly, CyrI enabled a catalytic, protecting-group-free oxidation to afford the final natural product in 93% yield with >90:1 diastereoselectivity achievable with high enzyme loading. The biocatalytic synthesis produced uniformly $^{15}$N labeled CYN in 10 total steps with 15% overall yield, demonstrating the advantages of combining traditional organic chemistry with enzymatic transformations.

As a comparison to the enzymatic approach, and to gain access to the 7-epi congeners of cytodrospermopsin, we also investigated the chemical oxidation of $[^{15}\text{N}]$-16 (Scheme 3b). Silylation of $[^{15}\text{N}]$-16 with TMSiCl (TESCI) afforded tricyclic guanidine $[^{15}\text{N}]$-17 in 93% overall yield from diene $[^{15}\text{N}]$-13 as a separable 3:2 mixture of diastereomers at C8, favoring the required R configuration. Replacement of methyl carbonate with the TES group in $[^{15}\text{N}]$-17 facilitated two practical objectives: (1) enabling a simple chromatographic separation of C8 diastereomers and (2) compatibility with basic conditions required for the ensuing C7 metalation and hydroxylation. Although the substrate was sensitive to the nature and stoichiometry of the base, we found that rapid lithiation with LDA (5 min, –78 °C) followed by a quench with an oxaziridine afforded the diastereomeric C7 hydroxylation products in good yields along with recovered starting material. The diastereomer
ratio is dependent on the oxaziridine: rac-19 afforded a 6:1 dr, while camphor based reagent 20 or ent-20 gave 25:1 and 20:1 dr in favor of $^{15}$N$_2$-1, respectively. The optimized hydroxylation was scaled to 1.50 g of $^{15}$N$_2$-17, providing $^{15}$N$_2$-21 as a pure isomer in 46% yield along with 41% of starting material. Complete hydrolysis to uracil along with the removal of the Mbs group was achieved on 0.350 g scale of $^{15}$N$_2$-21 by reflux in 12 M HCl for 3 h, resulting in a minor epimerization at C7 (10.1%). The mixture of epimers $^{15}$N$_2$-22 and $^{15}$N$_2$-23 was submitted to O-sulfonxylation with the SO$_3$-Py complex in DMF. After optimizing the reagent loading, we observed highly selective monosulfonation of the diol at the C12 OH group. The resulting mixture was readily separated by reversed-phase HPLC, affording 75 mg of $^{15}$N$_2$-CYN (77% yield) and 6.8 mg of $^{15}$N$_2$-7-epi-CYN (7% yield) in a single attempt. By comparison with biooxidation using Cyrl, the chemical oxidation shows improved scalability, but introduces one additional step in the total synthesis with slightly diminished overall yield and lower stereoselectivity.

With the supply of $^{15}$N$_2$-CYN secured by total synthesis, its development as an analytical standard for the precise quantification of CYN by LC/ESI-MS/MS was undertaken. Based on reported concentrations of the cyanotoxin in environmental samples, we developed an assay for the detection of cylindrospermopsin in a broad range from 10 ng/L to 100 µg/L. While the EPA limit for cylindrospermopsin in drinking water has been proposed to be 1 µg/L, concentrations below this level still represent an environmental hazard due to bioaccumulation and uncertain health risks of chronic sublethal exposure. The limit of detection (LOD, Signal/Noise = 3) and the limit of quantification (LOQ, Signal/Noise = 10) for the method were determined to be 2.5 and 8.25 ng/L, respectively, the highest sensitivity reported to date. The evaluation of matrix effect for analyte quantification was performed using environmental samples free of native CYN that were spiked with 100 µg/L. In all cases the observed recoveries were nearly 100%. This method was then used to quantify cylindrospermopsin in environmental water samples in 14 locations in Southern California where potential contamination or water resources was suspected. Notably, availability of $^{15}$N$_2$-CYN afforded the reliable and precise detection of cylindrospermopsin in Lake Casitas, a previously undescribed location of cyanotoxin contamination, where we detected the concentration of 157 ng/L of the toxin.

CONCLUDING REMARKS

In summary, we developed a strategy for the concise total synthesis of cylindrospermopsin that is directly applicable for $^{15}$N isotope incorporation. We demonstrated the scalability of the synthesis by preparing 0.125 g of $^{15}$N$_2$-CYN in 14 steps (longest linear sequence) and 9% overall yield from $^{15}$N$_2$-ammonium chloride, along with 7.5 mg of $^{15}$N$_2$-7-epi-CYN and 20 mg of $^{15}$N$_2$-7-deoxy-CYN. In the natural isotope series, the total synthesis of CYN was completed from commercially available 2,4-dimethoxy-6-bromopyrimidine (5) in only 10 steps via the enzymatic oxidation route (15% yield from 5), and 11 steps by chemical C7 hydroxylation (11% yield), representing a substantial improvement in brevity and scalability compared to the previous efforts. Furthermore, these efforts provided a high-quality isotopically labeled probe for the reliable detection and quantification of cyanotoxins in fresh water sources with the highest limits of detection reported to date. Our results demonstrate how atypical constraints associated with isotope incorporation empower the development of novel approaches to organic synthesis design.

ASSOCIATED CONTENT

Supporting information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b03071.

Full experimental procedures and copies of $^{13}$C, $^{15}$N NMR spectra (PDF)
Crystallographic data for 3 (CIF)

AUTHOR INFORMATION

Corresponding Author
zakarian@chem.ucsb.edu

ORCID
ARTURO A. KELLER: 0000-0002-7638-662X
BRIAN G. MILLER: 0000-0002-6641-0344
ARMEN ZAKARIAN: 0000-0002-9120-1332

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the NIH (NIGMS, R01-077379, R01-115388 and NIHES, R03 ES025345-01). Dr. Hongjun Zhou is acknowledged for assistance with NMR spectroscopy. Dr. Dmitriy Uchenik and the UCSB mass spectroscopy facility are thanked for assistance with mass spectral analysis. Dr. Guang Wu and the UCSB X-ray analytical facility are acknowledged for assistance with the X-ray crystallography. We are especially grateful to Anastasiia S. Minakova for assistance with the enzymatic oxidation experiment. Dr. Ben Lopez and the NRI-MCDB Microscopy Facility at UCSB are acknowledged for obtaining microphotographs of Cylindrospermopsis raciborskii.

REFERENCES

(22) Looper, R. E.; Runnegar, M. T. C.; Williams, R. M. *Tetrahedron* 2006, 62, 4549−4562.
(24) Cambridge Isotope Laboratories, Inc., listed prices: NHCl, $490/10g; aqueous 15NH4, $168/5g (33N solution), 15NH3, $454/1L of gas; (15NH3)2CO, $305/1g; [15N]guanidine, no longer available.
Total Synthesis of Salimabromide: A Tetracyclic Polyketide from a Marine Myxobacterium

Matthias Schmid,†,‡ Adriana S. Grossmann,† Klaus Wurst,§ and Thomas Magauer†,‡,○

†Department of Chemistry, Ludwig-Maximilians-University Munich, Butenandtstrasse 5-13, 81377 Munich, Germany
‡Institute of Organic Chemistry and Center for Molecular Biosciences, University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria
§Institute of General, Inorganic & Theoretical Chemistry, University of Innsbruck, Innrain 80-82, 6020 Innsbruck, Austria

Supporting Information

ABSTRACT: Salimabromide is an antibiotic polyketide that was previously isolated from the obligate marine myxobacterium Enhydrobacter salina, and its densely functionalized and conformationally rigid tetracyclic framework is unprecedented in nature. Herein we report the first chemical synthesis of the target structure by employing a series of well-orchestrated, robust transformations, highlighted by an acid-promoted, powerful Wagner–Meerwein rearrangement/Friedel–Crafts cyclization sequence to forge the two adjacent quaternary carbon centers embedded in the tetrahydrophthalene. A high-yielding ketiminium mediated [2+2]–cycloaddition was also utilized for the simultaneous construction of the remaining three stereocenters.

Myxobacteria of terrestrial origin produce an abundance of structurally complex secondary metabolites with notable biological activities. Prominent examples include epothilone, corallopyronin and soraphen.† Marine myxobacteria, on the other hand, have eluded their cultivation and isolation on many occasions and constitute a largely unexplored treasure trove of bioactive molecules.‡ Halimangin (2), which was reported in seminal work by Fudou in 2001,§ represents the first natural product isolated from a strictly marine myxobacterium (Scheme 1a). Following this early report, only a handful of new marine myxobacterium molecules have appeared in the literature, and in 2013, König disclosed the isolation of salimabromide (1) in minute quantities (0.5 mg from 64 L of culture) from the obligate marine bacterium Enhydrobacter salina.¶ Although a broad biological screening campaign was impossible at this stage, 1 was shown to possess inhibitory activity against Arthrobacter crysophakites (16 μg mL⁻¹).

From a structural point of view, salimabromide (1) contains an unprecedented tetracyclic ring-architecture that contains four consecutive stereocenters, one of which is quaternary. Additionally, the brominated tetrahydrophthalene core is bridged at C12 and C15 to form a highly substituted and synthetically challenging seven-membered carbocycle.¶ The conformational flexibility of this subunit is further reduced by fusion to a five-membered lactone.

The structure for 1 was exclusively deduced from extensive NMR measurements. Though this only confirmed the connectivity, the absolute stereochemistry was determined by comparison of calculated and measured CD spectra.

Scheme 1. (a) Structural Diversity Produced by Marine Myxobacteria and (b) Key Retrosynthetic Bond Disconnections for Salimabromide (1)

Received: June 13, 2018
Published: June 28, 2018
synthetic derivatives thereof. Herein we describe a streamlined synthesis of 1 employing a series of robust chemical transformations. The successful realization of this route allowed us to produce 1.9 g of a highly advanced intermediate in a single batch from which salmabromide (1) was prepared in only three steps.

Our retrosynthetic bond disconnections were guided by our desire to rapidly generate molecular complexity and to install the individual stereocenters via a minimum number of synthetic operations (Scheme 1B). We planned to first install the pivotal C12 quaternary carbon center and utilize this handle for the subsequent one-step construction of the stereotriad along C11, C13 and C15. For this purpose and in analogy to the logic of two-phase terpene (bio)synthesis, salmabromide (1) was first reduced to tetracycle 5. This intermediate was designed to eliminate steric hindrance and cross-reactivity of the bromine substituents en route to the tetracyclic carbon framework of 1. Compound 5 contains the retin for a powerful ketimine mediated [2+2]-cycloaddition to produce the simplified dicyclogeraphalene 6. Further C−C bond disconnections involving truncation of the side chain and a retro-Friedel−Crafts cyclization/Wagner−Meerwein rearrangement sequence provided epoxide 7.

We commenced our synthesis with the Claissen−Schmidt condensation of commercially available 3-methoxybenzaldehyde 8 with pinacolone 9 (Scheme 2). Though slow reactions provided 11 in good yield (76−83% over two steps, 32 g). Exposure of 11 to standard Corey−Chaykovsky conditions (NaOMe, Me2Si, 0 to 23 °C)10 effected clean conversion to epoxide 7. With 7 in hand, we investigated the crucial rearrangement-cyclization sequence. We found that exposure of 7 to hexanes/sulfuric acid or dichloromethane/titanium(IV) chloride was low-yielding for 12 (~30%) and afforded substantial amounts of the undesired ortho-product 13 (~10%). Further screening revealed 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, 0.1 M), a strong hydrogen-bond donor and weak nucleophile,11 as the solvent of choice and concentrated sulfuric acid as the optimal catalyst (10 mol %). Under these conditions, the para-product 12 was formed with very good regioselectivity (12:13 = 8:1). The remainder of the mass balance corresponds to a complex mixture of byproducts.15 Similar results were obtained when the reaction was conducted in a hexameric resorcinarene capsule.15 In agreement with related Lewis-acid catalyzed semipinacol and Wagner−Meerwein rearrangements,16 a high level of stereoselectivity should be possible for the initial rearrangement step. Evidence was obtained when a solution of enantiomERICALLY enriched 7 (82% ee) in dichloromethane was exposed to titanium(IV) chloride at −78 °C. Under these conditions, an enantiomERIC excess of 70% was obtained for 12 (see Supporting Information for details). This result provides an opportunity to access 1 in an asymmetric fashion.

Having successfully installed the crucial C12 quaternary stereocenter in only four steps, we turned our attention to the remaining functionalization of 12 (Scheme 3). To begin, the primary alcohol was protected as its tert-butylidiphenylsilyl (TBDDS) ether, and stereo 14 was then formed via exposure to 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ, 2 equiv.).15 The use of substoichiometric amounts of DDQ (30 mol %) in combination with manganese(IV) oxide (6 equiv) or alternative oxidation agents such as chloranil, Pd/C or electrochemical methods (4 V, chloranil, Pt/Pt, 0.1 M LiClO4, MeCN) were ineffective.

The TBDDS group was crucial for the stability of the silyl ether under the reaction conditions, and the presence of the electron-donating methoxy substituent was beneficial for the oxidation.16 For the installation of the missing ethyl substituent, we initially investigated the direct coupling of 14 using nickel (e.g., Ni(acac)2, dicye, EtMgBr, MgBr2, PhMe, 100 °C; Ni(cod)2, dicye, EtMgI, MgBr2, PhMe, 100 °C).17 Because our substrate proved to be remarkably unreactive under these conditions, we decided to replace the methoxy substituent with a more reactive triflate. For the removal of the methyl ether, freshly prepared lithium diethylphosphorodithioate (Ph2P, n-BuLi) was found to be optimal.18 The free phenol was then converted to the triflate upon exposure to the Hendrickson−McMurray reagent (PhNTf2)19 to afford 15 in quantitative yield. Standard Negishi cross-coupling (Pd(dpdpf)Cl2, ZnEt2, dioxane, 70 °C)20 enabled clean installation of the missing ethyl substituent (92%), and cleavage of the silyl ether (TBAF, THF, 23 °C) gave alcohol 16 (99%). The remaining carbon-chain was introduced in a three-step sequence beginning with the oxidation of 16 using Dess−Martin periodinane13 (96%), followed, by a high-yielding, Z-selective Wittig olefination and amide formation (pyrrolidine, 100 °C) to provide 6 (85%).

With robust access to the dihydroxyphenanthrene 6, we proceeded to investigate the key step of our synthesis: a ketimine ion mediated [2+2]-cycloaddition to construct

---

Scheme 2. Synthesis of the Dicyclogeraphalene Framework via Consecutive Wagner−Meerwein Rearrangement and Friedel−Crafts Cyclization

![Scheme 2](image-url)
Scheme 3. Total Synthesis of Salimabromide

Reagents and conditions: (a) TBDPSCI, imidazole, DMAP, DMF, 23 °C; (b) DDQ, dioxane, 93 °C, 91% over two steps; (c) LiPPh₂, THF, 60 °C; (d) Pd(dppf)Cl₂, dioxane, 70 °C, 92%; (f) TBAF, THF, 23 °C, 99%; (g) DMP, H₂O, CH₂Cl₂, 0 to 23 °C, 85%; (l) TfO, 2,4,6-collidine, CH₂Cl₂, CH₃CO₂H, 80 °C, 89%; (k) SeO₂, SiO₂, dioxane, 120 °C, 47% over five cycles; (l) t-BuCHO, Cu(OAc)₂, O₂ (1 atm), CICH₂CH₂Cl, 23 °C, then DMP, NaHCO₃, Br₂, CH₂Cl₂, 79%, 23:24 = 1:3.2; (m) AgTFPA, Br₂, CF₃COOH, 50%.

When S (500 mg) was treated with selenium dioxide (dioxane, 120 °C, 6 h) in the presence of silicon dioxide (>230 mesh) to prevent agglomeration, the diastereomerically pure allylic alcohol 22 was formed together with unreacted S. Extended reaction times were detrimental as overoxidation and decomposition started to prevail. Subjection of recovered S (76%) to the reaction conditions enabled us to prepare 250 mg of 22 after five cycles (47%, ~15% for the first cycle). Subsequent Baeyer–Villiger oxidation using standard conditions (m-CPBA, NaHCO₃, CH₂Cl₂) gave two region-specific lactones, which were directly oxidized (DMP, NaHCO₃, CH₂Cl₂) to afford 23 and 24 in a ratio of 1:4:1 in 84% combined yield. Separation of 23 and 24 was readily accomplished by flash column chromatography. To improve this undesired outcome, further optimization of the oxidant was performed. Interestingly, exposure of 22 to t-BuCHO (5 equiv) in the presence of molecular oxygen (1 atm) and copper(II) acetate (1 equiv) gave 24 as the major product.

the seven-membered carbocycle and complete the tetracyclic carbon framework. Under optimized conditions, a solution of amide 6 and syn-collidine (1.2 equiv) in dichloroethane was slowly added to a refluxing solution of freshly distilled trifluoromethansulfonic anhydride (1.2 equiv) in dichloroethane (0.1 M). The cycloaddition produced the tetracycle 5 with almost perfect regio- and diastereoselectivity in excellent yield (89%, 1.9 g). The exact mechanism, concerted (synchronous/asynchronous) or stepwise, has been a matter of debate for many decades. Depending on the substitution pattern on both the alkene and the ketimine salt, either of the two pathways might be operational. The greater resonance stabilization of the benzylic cation 21 versus 19 was envisioned to govern the regioselectivity favoring formation of 5.

Having installed the crucial stereocenters, we were poised to tackle the remaining challenges: regioselective oxidation of the carbon-framework and bromination of the aren: subunit.
(79%, 23:24 = 1:3.2). The directing effect of the free hydroxy group was crucial as the corresponding methyl ether led to lower regioselectivity (1:1.1) only slightly favoring the desired regioisomer (compare 24). It is also noteworthy that replacement of t-BuCHO with m-CPBA in the Baeyer–Villiger step under otherwise identical conditions was even less efficient and only poor regioselectivity (23:24 = 1:2.1) was obtained. For the introduction of the missing bromine substituents, 24 was exposed to a panel of brominating agents (e.g., Br₂, CHCl₃, NBS, HOAc; NaMe₂NBr, ZnBr₂, HOAc). Under these conditions, formation of 1 was only observed in trace amounts if at all. Finally, we found that treating a solution of 24 in trifluoroacetic acid (0.1 M) with silver(I) trifluoroacetate (3 equiv) and elemental bromine (3 equiv) enabled the desired bromination (50%) and thus completed the synthesis of salmabromide (1, 50 mg). The analytical data for 1 (¹H NMR, ¹³C NMR, HRMS) fully matched those reported for the natural compound. Additionally, the structure of 1 was unambiguously validated by single-crystal X-ray diffraction analysis.

In summary, we have completed the first total synthesis of salmabromide, a unique tetracyclic polyketide. The highlights of the developed route are (1) a powerful Wagner–Meerwein rearrangement/Friedel–Crafts cyclization sequence to forge the tetrahydronaphthalene skeleton and (2) a high-yielding ketimination mediated (2+2)-cycloaddition to set the remaining three stereocenters. The overall sequence benefits from a series of practical transformations that can be also conducted on large scale. The robustness of the developed synthesis is evident from the fact that more than 1.9 g of a highly advanced intermediate were prepared in a single batch.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b06228.

Experimental details and spectroscopic data (PDF)

X-ray crystallographic data for 1, 5, 23 and 24 (CIF)

AUTHOR INFORMATION

Corresponding Author

*thomas.magauer@uibk.ac.at

ORCID

Thomas Magauer: 0000-0003-1290-9556

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

T.M. acknowledges the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement no. 714049) and the Center for Molecular Biosciences (CMBI). M.S. gratefully acknowledges financial support from the German National Academic Foundation. We thank Dr. Kevin Mellem (Revolution Medicines) for helpful discussions, Prof. Alfred Berkessel (University of Cologne) for a screen of epoxidation catalysts and Prof. Konrad Tiefenbacher (University of Basel/ETH) for the screen of supramolecular catalysts.

REFERENCES


(12) The mixture of byproducts could not be separated; however, careful analysis of the ¹H NMR revealed minor traces of the Meinwald rearrangement product. Meinwald, J.; Labana, S. S.; Chadha, M. S. J. Am. Chem. Soc. 1963, 85, 582. An Indane, resulting from the direct Friedel–Crafts reaction, was not detected.


(18) (b) Replacement of the methoxy by a trifluoromethyl group gave low yields in the subsequent oxidations.


Physical Chemistry Cume: Maxwell Thermodynamics Relations

*Physical significance:* Maxwell relations are helpful in replacing unmeasurable thermodynamic quantities by measurable quantities for which experiments can be designed to derive properties of any complex real systems (molecules, chemical processes, chemical/biological reactions etc.). The real systems do not necessarily follow simple equations of state, e.g. equation of an ideal gas.

As a physical chemist, your mission, should you choose to accept, is to reduce the following partial derivatives to an expression containing only $\alpha_P$, $\kappa_T$, $C_P$, $C_V$ and/or other thermodynamic variables (e.g. $T$, $P$, $V$ etc.), AND show that the units match for your final answer. Express your results in a general form that pertains to ANY system with no change in number of molecules, i.e. $dN_i = 0$.

For full credit, show your work.

1. **10 points** Fill in the blanks in the following expressions:

   \[
   (a) \ V = \left( \frac{\partial S}{\partial P} \right)_S \quad (b) \ \left( \frac{\partial S}{\partial V} \right)_T = \left( \frac{\partial P}{\partial T} \right)_V \quad (c) \ \left( \frac{\partial V}{\partial P} \right)_T = -\frac{\left( \frac{\partial S}{\partial P} \right)_V}{\left( \frac{\partial S}{\partial V} \right)_P} \\
   (d) \ \left( \frac{\partial U}{\partial T} \right)_G = \left( \frac{\partial V}{\partial T} \right)_T \left( \frac{\partial P}{\partial T} \right)_G + \left( \frac{\partial V}{\partial T} \right)_T \left( \frac{\partial P}{\partial T} \right)_G \quad (e) \ \left( \frac{\partial U}{\partial P} \right)_T = T \left( \frac{\partial S}{\partial P} \right)_T - \left( \frac{\partial V}{\partial P} \right)_T 
   \]

2. **20 points** The Joule-Thompson coefficient, $\mu_J$, is a measure of the temperature change during a throttling process given by $\left( \frac{\partial T}{\partial P} \right)_H$. A similar measure of temperature change produced by an isentropic change of pressure is provided by the coefficient, $\mu_S = \left( \frac{\partial T}{\partial P} \right)_S$. Reduce these expressions and show how they are related to each other in terms of measurable quantities? Show that units match on either side.

3. **15 points** Derive a relationship between $C_P$ and $C_V$ for any system with no change in number of molecules and in terms of measurable thermodynamic quantities. Show that the units on either side match.

4. **30 points** Reduce $\left( \frac{\partial T}{\partial U} \right)_H$ in terms of measurable thermodynamic quantities and show that the units on either side match from your answer.

5. **10 points** Reduce $\left( \frac{\partial A}{\partial G} \right)_T$ in terms of measurable thermodynamic quantities and show that the units on either side match from your answer.

6. **15 points** Consider a two-component chemical process where $n_1$ moles of a solute is added to $n_0$ moles of the solvent at constant volume and temperature. We want to measure the change in pressure, $\left( \frac{\partial P}{\partial n_1} \right)_{T,V,n_0}$, which is not easy to measure experimentally. Reduce the partial derivative in terms of the solute partial molar quantity (that can be measured experimentally), as well as, $C_P$, $C_V$, $\alpha_P$, $\kappa_T$, and/or other thermodynamic variables to do this experiment. Show the units match on either side.
Fundamental Equations and Derivative Relations

SI units:  \( 1 \ J = 1 \ N \ m = 1 \ kg \ (m/s)^2 \)  \( 1 \ N = 1 \ J/m = 1 \ kg \ m/s^2 \)  
\( 1 \ Pa = 1 \ J/m^3 = 1 \ N/m^2 \)  \( 1 \ W = 1 \ J/s \)

\( \beta = 1/k_B T \) or \( 1/RT \),  \( RT = N_A k_B T \)  \( (RT \sim 2.5 \ kJ/mol \ at \ 300 \ K) \)

\( 1 \ MPa = 10^6 \ Pa = 1 \ kJ/L = 1 \ J/cm^3 \)  \( 1 \ m^3 = 10^3 \ L = 10^6 \ cm^3 \)

\[
U = TS - PV + \sum_i \mu_i N_i  \quad H = U + PV  \quad A = U - TS  \quad G = U - TS + PV
\]

\[
dU = TdS - PdV + \sum_i \mu_i dN_i  \quad dA = -SdT - PdV + \sum_i \mu_i dN_i
\]

\[
dH = TdS + VdP + \sum_i \mu_i dN_i  \quad dG = -SdT + VdP + \sum_i \mu_i dN_i
\]

\[
\alpha_p = \frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T  \quad \kappa_T = -\frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_P  \quad C_V = T \left( \frac{\partial S}{\partial T} \right)_V = \left( \frac{\partial U}{\partial T} \right)_V  \quad C_P = T \left( \frac{\partial S}{\partial T} \right)_P = \left( \frac{\partial H}{\partial T} \right)_P
\]

\[
f = c - p + 2  \quad \dot{\chi}_i = \left( \frac{\partial X}{\partial n_i} \right)_{T,P,n_j \neq n_i}  \quad S = k_B \ln\Omega  \quad dS = \frac{\delta Q}{T}  \quad dS_{surr} = \frac{\delta Q_{surr}}{T_{surr}}
\]

For any function \( F(x,y,z) \):

\[
\left[ \frac{\partial}{\partial x} \left( \frac{\partial F}{\partial y} \right) \right]_{x,y} = \left[ \frac{\partial}{\partial y} \left( \frac{\partial F}{\partial x} \right) \right]_{y,x}
\]

\[
\left( \frac{\partial F}{\partial x} \right)_y = \left( \frac{\partial F}{\partial y} \right)_y \left( \frac{\partial x}{\partial y} \right)_y
\]

\[
\left( \frac{\partial y}{\partial x} \right)_F = \left( \frac{\partial F}{\partial y} \right)_y
\]

\[
\left( \frac{\partial F}{\partial x} \right)_y = \left( \frac{\partial F}{\partial x} \right)_z \left( \frac{\partial z}{\partial y} \right)_y
\]

Derivatives and integrals:

\[
d \frac{d}{dx} [\ln(f)] = \frac{1}{f} \frac{d}{dx} [f]  \quad d \frac{d}{dx} [e^f] = \frac{d}{dx} [f] e^f  \quad d \frac{1}{f} = \frac{d}{f^2} \frac{d}{dx} [f]
\]

\[
\int a x^n dx = a \frac{x^{n+1}}{n+1} + C  \quad \int a \frac{dx}{x} = a \ln x + C  \quad \int e^{ax} dx = \frac{e^{ax}}{a} + C
\]

27