Revealing ocean-scale biochemical structure with a deep-diving vertical profiling autonomous vehicle

John A. Breier1,*, Michael V. Jakuba2, Mak A. Saito2, Gregory J. Dick3,4,5, Sharon L. Grim3,†, Eric W. Chan1, Matthew R. McIlvin2, Dawn M. Moran2, Brianna A. Alanis1, Andrew E. Allen6,7, Chris L. Dupont6, Rod Johnson8

Vast and diverse microbial communities exist within the ocean. To better understand the global influence of these microorganisms on Earth’s climate, we developed a robot capable of sampling dissolved and particulate seawater biochemistry across ocean basins while still capturing the fine-scale biogeochemical processes therein. Carbon and other nutrients are acquired and released by marine microorganisms as they build and break down organic matter. The scale of the ocean makes these processes globally relevant and, at the same time, challenging to fully characterize. Microbial community composition and ocean biochemistry vary across multiple physical scales up to that of the ocean basins. Other autonomous underwater vehicles are optimized for moving continuously and, primarily, horizontally through the ocean. In contrast, Clio, the robot that we describe, is designed to efficiently and precisely move vertically through the ocean, drift laterally in a Lagrangian manner to better observe water masses, and integrate with research vessel operations to map large horizontal scales to a depth of 6000 meters. We present results that show how Clio conducts high-resolution sensor surveys and sample return missions, including a mapping of 1144 kilometers of the Sargasso Sea to a depth of 1000 meters. We further show how the samples obtain filtered biomass from seawater that enable genomic and proteomic measurements not possible through in situ sensing. These results demonstrate a robotic oceanography approach for global-scale surveys of ocean biochemistry.

INTRODUCTION

On Earth, the cycling and availability of carbon, nitrogen, phosphorus, and oxygen, as well as other essential nutrients and cofactors, are strongly influenced by processes that occur within the ocean over basin scales (1–3). Planktonic marine microorganisms, which collectively account for close to half of planetary primary production, drive these processes (4). Redfield (5) described the reciprocal, and chemical, relationships between ocean biology and the environment many years ago, suggesting that this relationship was the result of plankton communities building their cellular biochemical constituents. Geochemists and ecologists have incorporated these relationships into large-scale conceptual models of nutrient flows, generally referred to as biogeochemical cycles. However, it has always been challenging to identify the essential processes to model and validate their representation and predictions. While our understanding of the processes connecting planktonic microbiology to ocean chemistry has grown substantially over the years (6–8), our appreciation of the complexities in these relationships has grown, too (9–11). However, the recent inclusion of genomic (12, 13), transcriptomic (14), proteomic (15), and metabolomic (16) measurements in ocean studies is creating great potential to understand the specific biochemical mechanisms coupling ocean life and environmental chemistry.

Currently, there is keen interest in better understanding these natural processes (17) and urgency to do so because anthropogenic processes are altering the physical and chemical environment of the ocean, with the potential to reduce the ocean’s ability to buffer large-scale climatic shifts (18). To improve our understanding of Earth’s biogeochemical cycles, we developed a robotic vehicle to facilitate water column ocean mapping of microbial and biochemical data (Fig. 1). This robot, called Clio, collects dissolved and particulate samples and measurements from the water column for microbial community proteomics, genomics, transcriptomics, organic and inorganic chemistry, and related physical data. These data types include biochemicals, sparsely mapped at present, and supporting parameters that must be measured synoptically. Sensors can measure only a subset of the necessary data; thus, sample return for shore-based analyses is essential. Clio collects samples and sensor data in the form of (i) high-resolution ocean depth profiles, (ii) ocean basin scale transects of the same, and (iii) targeted data collection from biochemical hotspots hidden within the ocean. Although we can make some biochemical and chemical measurements with small sample volumes (milliliters to liters), we require large volumes for other measurements (e.g., proteomics, organic carbon, and trace metals). Therefore, Clio is capable of processing several hundred liters of seawater per sample to return material for subsequent high-throughput laboratory analyses. Clio is designed for diving to depths of 6000 m and can, therefore, access most of the ocean. Clio drifts freely in the lateral plane to better observe ocean water masses. Clio operates autonomously without a tethered connection to a surface vessel; this frees the robot from the restricted motion that a tether imposes, and it frees the surface vessel to carry out synoptic science activities (fig. S1). Before Clio, an ensemble of sequential techniques, including the repeated lowering and recovery of cabled equipment from research vessels, was required to collect similar full ocean depth datasets.
unique and directed at facilitating the dual goals of global ocean mapping and providing samples for biochemistry.

Fig. 1. Cléo biochemical AUV. Cléo vertically profiles the ocean water column to 6000 m and drifts horizontally in a Lagrangian manner. The (A) H-shape hull protects the (B) thrusters and creates payload bays for sensors and samplers, the inlets of which extend through the (C) vehicle sides. The (D) payload bays are arranged at the corners of the H-shape. The (E) floatation is near the vehicle’s top, and the heaviest components are near the bottom. Cléo is (F) comparable in size to other deep-diving vehicles and sampling platforms. However, its design and purpose are unique and directed at facilitating the dual goals of global ocean mapping and providing samples for biochemistry.

Existing vehicle designs are not well suited to diving deep, systematically crossing an ocean in short order, and processing large volumes of water for sample return (Fig. 1F). We addressed this challenge by developing a new robot for the most unfulfilled and currently limiting aspect of these objectives while making use of existing ocean surface vessels where they were most capable: quickly transiting between stations. Specifically, we optimized Cléo to execute single vertical ocean data collection dives as effectively as possible because Lagrangian motion is better for sampling. Thus, the Cléo design foregoes any lateral control. During deployment, Cléo is launched from a surface vessel and conducts an autonomous sample and data collection mission over a predetermined depth range. Cléo is recovered by the surface vessel at the end of each dive, typically lasting 7 to 14 hours, and transported to the next dive location where this process is repeated (Fig. 2 and fig. S1). In this way, Cléo can acquire data and samples within weeks that span ocean basins.

Global-scale ocean studies

In 1934, Redfield (19) hypothesized that the ratios of inorganic nutrients in seawater were in a reciprocal self-regulating relationship with marine microorganisms based on the analysis of global datasets. This was not obvious and only exists in the oceans because phytoplankton dominate marine biomass and oceans are constantly mixing, but this only became apparent through the examination of global datasets. Subsequent global-scale studies have also transformed our understanding of ocean processes. The Geochemical Ocean Sections Study (GEOSECS) improved our understanding of ocean circulation (20). The Tropical Ocean Global Atmosphere (TOGA) program allowed us to predict better El Niño and Southern Oscillation events (21). The World Ocean Circulation Experiment (WOCE) and the Joint Global Ocean Flux Study (JGOFS) informed our initial understanding of ocean circulations’ influence on Earth’s climate and the ocean carbon cycle, respectively (22, 23).

Current global ocean studies, including GEOTRACES (24), build on these previous efforts; more details are in the Supplementary Materials. An important part of these studies has been the systematic collection of data for the creation of ocean sectional maps of the water column. The field methods for creating ocean sectional maps have generally been consistent: (i) collect water and particulate samples from multiple depths in a vertical profile from the ocean surface to the deepest depth, (ii) repeat this process at multiple stations spread along a transect crossing an ocean, and (iii) complete multiple transects over as many research cruises to create a network of sectional maps that cover Earth. To date, ocean sectional studies rely on wire-based water column sampling technology, including water sampling rosettes and wire-deployed pumps (fig. S1B). The time to collect samples this way is determined by winch wire speed (typically 0.5 to 0.7 m s⁻¹), sample depths, pumping time, processing time, and the number of sampling deployments, which must be sequential to avoid entangling wires. It can take 1 to 2 days per station, sometimes more, for modern sectional studies such as GEOTRACES to acquire the necessary samples for all measurements (more details are in the “Sectional cruise methodology” section in the Supplementary Materials). This results in sectional cruises that are typically 30 to 65 days long, a duration that can be shortened or made more productive with robotic tools like Cléo.

The biochemistry of the oceans, including microbial community genomics, transcriptomics, proteomics, and metabolomics, has yet to be widely measured. Global genomics surveys to date, such as the Global Ocean Sampling program (25) and the Tara Expedition (26), have focused almost entirely on the sunlit eutrophic zone. The biochemistry of the rest of the ocean is largely unmapped. This fact and the results of previous ocean basin studies are inspiring efforts to launch new global microbiology and biochemistry studies (27), such as the proposed Biogeosciences effort.

Cléo is designed to enable global studies of this kind by speeding the collection of the necessary data. Cléo collects a full suite of samples...
These platforms typically have small payload and battery capacities that cannot accommodate large sample-processing systems. Both platforms are optimized for travel over great distances but over long time scales that would (i) limit the sample collection rate even if payload constraints were not limiting and (ii) raise greater concern regarding sample preservation. Floats and gliders are not currently well suited to the type of sample return necessary for biochemical sectional studies.

ROVs have been used as ocean-observing platforms since the 1980s (37–39). Among ocean vehicles, they have the highest payload capacities and are the most reconfigurable. The precision navigation capabilities of ROVs have enabled targeted sampling in and around seafloor seeps and hydrothermal biological “hotspots” (40–43). At these sites, the concentration of microbial life and chemical nutrients can be high; thus, processing large samples is unnecessary in many cases. However, collecting material emitted from these systems up into the water column as it mixes and dilutes with seawater does require processing large samples; the sampling systems developed for ROVs inherit design principles from instruments developed to return samples from deep-sea hydrothermal plumes (43, 44). Nevertheless, ROVs require cables, cables require winches, and cables and winches typically constrain the supporting research vessel to one activity at a time. For ocean sectional studies, ROVs provide little benefit over existing cabled sampling technologies.

AUVs have been used as ocean observing platforms since the 1990s (45–47). Optimized for longer-distance surveys and more task-specific observations than ROVs, AUVs trade smaller payload capacities for lower drag and increased range. Fully autonomous, they can operate independently from surface ships. Most AUV designs have been multipurpose. They are used widely for mapping operations, particularly at the seafloor (47) but also under ice (48). AUVs have been adapted to the study of surface plankton processes (49, 50) and equipped with dedicated sample return systems to support these studies (51, 52). In the coastal surface ocean, phytoplankton concentrations can be high; as a result, processing large samples is unnecessary. Therefore, AUV sample return systems designed for these surface ocean studies typically collect and process water volumes on the order of 1 to 2 liters per sample. A few AUV applications have made use of deep-sea sampling system designs and are capable of collecting larger volumes for a broader range of biochemical analyses (53, 54).

Moreover, most AUV designs are based on the familiar torpedo shape and optimized for long-range lateral motion. For example, the Tethys Long Range AUV’s have a range of at least 1800 km (55), and the long-range Autosub version is designed for a range of 6000 km (56). These AUVs and others like them can also dive deep, but they are not well suited to repeatedly profiling to full ocean depth nor doing so in a single deployment.
Clio is a reimagination of AUV design applied to vertical profiling. It builds upon sample return instrumentation and techniques developed for the deep sea, where biomass concentrations are low, and applies them to the full-water column. It is designed and optimized for a single purpose and a mode of science operations orthogonal to previous AUV applications. Instead of collecting core datasets at global scales or intensively observing regional processes, we designed Clio to complete an intensive study of global-scale biochemistry.

Experimental setting
We tested Clio over several years, six cruises, and 26 dives. Initial sea trials were in 2017 near the shelf break of the eastern Atlantic coast of North America, just north of the Gulf Stream and the Sargasso Sea. During 2018 and 2019, we deployed Clio repeatedly, on a semiquarterly basis, at the Bermuda Atlantic Time-series Study (BATS) station (31.66667°N, −64.16667°W) within the Sargasso Sea. The Sargasso Sea is an oligotrophic region of the North Atlantic Ocean where Redfield ([19]) made some of his initial observations in 1933.

Contributions
Clio is uniquely designed to facilitate global-scale studies of ocean biochemistry, move vertically through the water column with high precision, and specifically return samples from large swaths of the ocean ranging in depths from the surface to 6000 m. Clio’s range encompasses full ocean depths everywhere except for hadal trenches. Clio is capable of flexible, precise vertical motion that few other ocean robots can perform, and none to our knowledge over this depth range. We demonstrate that Clio can move through the water column at a controlled rate as slow as desired or up to 0.8 m s⁻¹ and settle at desired depths to within 50 cm in less than a minute. It can hold a vertical station within a depth window of less than 5 cm for multiple hours while sampling and sensing. Clio is isolated from the influence of ship heave, unlike for tethered instruments. Unlike other AUVs, Clio can hold station vertically while drifting in a Lagrangian manner laterally. This allows high-resolution, vertical profiling using both sensors and samplers that are ideal for targeted observations of fine chemical or biological gradients that exist within the ocean water column. Importantly for its mission, Clio can acquire and return a broader range of sample types, from larger water volumes, than other existing ocean robots. From samples collected in the North Atlantic Ocean and the Sargasso Sea, we demonstrated that Clio can be used to make metagenomic and proteomic measurements and collect sensor data and samples for more routine nutrient measurements. In June 2019, Clio completed its first sectional cruise in a series of nine dives spanning a 1144-km transect between Bermuda and Woods Hole, Massachusetts, USA, through the Sargasso Sea and the Gulf Stream. This transect focused on the upper 1000 m of the water column and also included Clio’s deepest dive to date, to 4100 m, demonstrating Clio’s ability to conduct ocean-crossing sectional studies.

RESULTS
In this section, we describe our AUV design for biochemical ocean mapping. We also describe the design of this robot’s sample return systems, which it uses to manipulate seawater to extract samples for return. We then report on the results of field trials where the robot was used to conduct high-precision sampling within the water column.
systems. Samples collected in this way produce a radially uniform pattern across the filter surfaces, which is essential to achieve because the filters are subdivided on the basis of this assumption; this is part of the methodology used with existing wire-based systems, and we match this with Clio. Clio is also constructed of materials selected to maximize sample integrity and cleanliness. In particular, ferrous alloys were all but eliminated. Iron is a trace and limiting nutrient in seawater. Even minute concentrations of anthropogenic iron introduced from ferrous alloys on the vehicle could influence the trace metal measurements or the biochemical measurements made with this system. In place of ferrous alloys, aluminum, titanium, and nickel alloys were used. For similar reasons, the use of oil-filled electrical junction boxes, standard on some ROVs and AUVs, was minimized, and where possible liquid fluorocarbons were used in their place. When the mineral oils commonly used for these applications inevitably leak out of their reservoirs, they coat the outer surfaces of vehicle components with a residue that can remain on many dives and become entrained into the sampling systems, potentially biasing them for some organic analyses. The fluorocarbon substitutes are more volatile than mineral oils, and residues quickly evaporate, minimizing this concern (59). The location of the thrusters and the shape of the vehicle are designed to minimize the influence of propeller wash on sensor data or the collected samples. The thrusters are located within vertical channels on the sides of the vehicles. The sensors are arranged at the bottom of the vehicle, which permits the collection of undisturbed sensor profile data during descent. The intakes for the sample return system are colocated with the sensors on sides orthogonal to the thrusters.

**Manipulating seawater**

Clio’s primary task is to extract material from seawater and return it for analyses that cannot be measured by in situ sensors. Therefore, Clio has four payload bays (each 25 cm by 25 cm by 115 cm), each capable of containing independent, custom, high-volume in situ water filtration sampling systems and whole-water and filtrate subsampling systems (Fig. 3A). Each sampler can collect up to 19 samples depending on media size. For operations to date, we outfitted Clio with sampling systems in two payload bays and configured them to collect nine sample sets each consisting of (i) particulate material filtered onto 142-mm-diameter, 0.2-μm—pore size media, (ii) 200 ml of filtrate, and (iii) 200 ml of unfiltered water (“Biochemical methods” section in Supplementary Materials). Particulate samples can be preserved in situ by preservative administration. Depending on biomass concentration, this configuration typically filters 40 to 300 liters of water per sample and 1500 to 3500 liters of water total per dive; even so, Clio has substantial unused payload for future instrumentation (table S1).

The sample-processing and return systems are the most essential and unique tools that the robot has for interacting with the environment (see Materials and Methods). They are effectively this robot’s manipulators. Cleanliness was a design concern throughout the vehicle, but especially so in the sample return systems. Apart from electronics and motor parts contained within the liquid fluorocarbon–filled stepper housing, nearly all other parts are plastic: high-density polyethylene, acetal copolymer, polycarbonate, polytetrafluoroethylene, and stereolithography resin (Durable, Formlabs). Titanium and nickel alloys were used selectively where strength was essential. Each sample container unit has an independent intake to preclude cross-contamination. The custom sample container units were designed to interconnect between the sampling valve and all other individual sample units in the set without the need for tubing (Fig. 3B), which is time consuming to interconnect and difficult to clean (42). The sample containers themselves are designed for critical cleaning and to facilitate sample extraction in shipboard fabricated cleanrooms (Figs. 2, E and F, and 3C).

**Experimental setup**

Several dive types were conducted during the field trials described here. During initial sea trials, a series of five dives were executed to increasing depths, the last of which was to 2000 m in a water depth of 4100 m. During the dives at BATS station, we tested a number of vehicle configurations, diving to depths of 2000 m. The most frequent dive profiles to date were 1000-m dives completed at BATS and during the Bermuda–to–Woods Hole transect. Near the end of the Bermuda–to–Woods Hole transect, Clio completed a dive to 4100 m. The configuration of the sample return systems for these deployments is described in the “Biochemical methods” section in the Supplementary Materials.

**AUV performance, endurance, and power efficiency**

The typical 1000-m AUV profile took 16 hours to complete (e.g., Fig. 4). While diving, Clio was subject to Lagrangian horizontal transport dependent on the...
vehicle surfaced 34 km from the launch site. The supporting research typically surfaced 2 to 15 km from the launch site; in the most ex- mass being studied. As a result of Lagrangian transport, the AUV prevailing vertical currents. This is intentional; Lagrangian motion samples for different types of measurements. In this case, the filter numbers are odd numbers between 1 and 17 and creased depth resolution; see table S1. Parallel filtration with independent payload bays facilitates the collection of filled with nine filter units, but more often, two payload bays of filter units are deployed for parallel filtration or in-

Clio was deployed with one payload bay, but three filter units were used for parallel filtration at three different depths. This allowed for the collection of samples for different types of measurements. In this case, the filter numbers are odd numbers between 1 and 17 and reflect the port position of the valve.

prevailing vertical currents. This is intentional; Lagrangian motion results in better samples that are more representative of the water mass being studied. As a result of Lagrangian transport, the AUV typically surfaced 2 to 15 km from the launch site; in the most extreme case, during a profile at the edge of the Gulf Stream Current, the vehicle surfaced 34 km from the launch site. The supporting research vessel typically transits at 15 km hour\(^{-1}\) and can typically reach the AUV in <1 hour; recovering the AUV back on to deck was completed in <1 hour depending on sea state. This operational tempo is in keeping with our design objectives. Meeting these objectives required careful attention to understanding and optimizing the vehicle’s ballasting and power demands.

Energy expended during dives consists of (i) baseline power consumption including profiling sensors and sampling pumps and (ii) propulsion power spent to overcome buoyancy while sampling and to overcome drag while transiting. As with all vehicles, this represents a trade-off, and every vehicle has an optimal speed at which range is maximized (60). Slow transits reduce the energy needed to overcome drag but increase the energy needed to sustain base load and overcome buoyancy. The energy consumed overcoming buoyancy during the descent comes close to, but does not entirely cancel, the energy negated by buoyancy on the ascent, because of asymmetry in the propellers. Fast transits reduce the energy needed to sustain base load and overcome buoyancy but increase the energy needed to overcome drag. **Clio**’s mission calls for relatively rapid transits where transits are required to sustain base load and overcome buoyancy. The energy consumed overcoming buoyancy during the descent comes close to, but does not entirely cancel, the energy negated by buoyancy on the ascent, because of asymmetry in the propellers. Fast transits reduce the energy needed to sustain base load and overcome buoyancy but increase the energy needed to overcome drag. **Clio**’s mission calls for relatively rapid transits where

**Clio**’s buoyancy changes with depth as the density of seawater, and all of the materials comprising the vehicle respond to differences in temperature and pressure. **Clio**’s deepest dive, to 4100 m, allowed us to test our understanding of the vehicle’s net coefficient of thermal expansion and net bulk modulus. Analysis of calibrated thruster data from **Clio**’s dive to 4100 m indicates a best-fit coefficient of thermal expansion of 3.65 × 10\(^{-5}\) m m\(^{-1}\) K\(^{-1}\) and a net bulk modulus of 3.3 GPa. This is 1.5 times less compressible than typical seawater (2.3 GPa) and favorable compared with most deep-diving AUVs that are typically two to five times less compressible than seawater (61). For a vehicle of **Clio**’s size, the typical range corresponds to 90 to 150 N in added buoyancy at 6000 m, i.e., a notable increase in parasitic propulsion load. The value for **Clio** is 50 N.

Figure 5 illustrates the power consumed while holding depth to collect samples. Background vehicle processes and sensors consume 1.2% of battery capacity hour\(^{-1}\). Each additional pump used while sampling consumes another 0.4% of battery capacity hour\(^{-1}\). The energy consumed by the thrusters varies with depth and depends on the surface ballast condition. For conservative ballasting, i.e., 100 N
buoyant or 1.4% of vehicle displacement, energy consumption is 1.5 to 3% of battery capacity hour$^{-1}$. For a more aggressive ballasting state, i.e., 50 N buoyant or 0.7% of vehicle displacement, energy consumption is reduced to 1 to 2% of battery capacity hour$^{-1}$. Propulsion energy consumption is lowest at intermediate depths of 100 to 1000 m where the air bladder volume has been compressed and vehicle buoyancy has yet to be strongly influenced by pressure or increasing background density. Figure 5 shows that with aggressive ballasting, the energy expended by propulsion while holding depth is commensurate with the energy required for sampling throughout Clio’s 6000-m depth range, thus validating a key assumption during the design process that chose thrusters as the sole means of depth control.

Baseline cleanliness and genomic resolution

Baseline cleanliness was assessed using genomic analyses of samples returned from a Clio profile. Complementary DNA (cDNA) sequences were scanned for putative contaminant taxa identified by Salter et al. (62) and the interpretation of Sheik et al. (63) for low-biomass sequenced samples from the Census of Deep Life. Nodes that match in taxonomy to these were flagged as putative contaminants (fig. S2 and table S2). In this profile, we assessed background cDNA sequence contamination to be between 0.2 and 4.5%, and 1% on average.

Genomic analyses also show the microbial community structure and genomic resolution achieved by Clio (Fig. 6 and table S2). Figure 6 shows the results from nine samples collected by Clio from depths of 30 to 2000 m and analyzed for microbial DNA and RNA measured as cDNA (“Biochemical methods” section in the Supplementary Materials). Up to 38,280 DNA sequences and 24,493 cDNA sequences were obtained from filtered volumes of 18 to 57 liters, using a one-sixteenth split portion of the filtered biomass. In two cases, insufficient biomass or extraction methods, or both, resulted in low cDNA yields (table S2, samples D5 and D17); in all other cases, filter biomass was sufficient to generate sequencing libraries with good resolution that were representative of the water column community. These results demonstrate core functionality and baseline necessary volumes and illustrate, to some degree, the amount of data returned from these biochemical measurements and contained within just one profile.

The genomic data in Fig. 6 (A and B) are categorized at the phylum level into 27 groups, but this is just a fraction of the data. A deeper characterization is possible within each of these levels. The genomic data in Fig. 6 (A and B) are categorized at the phylum level into 27 groups, but this is just a fraction of the data. A deeper characterization is possible within each of these levels. Figure 6 (C and D) shows the classification of the Proteobacteria phylum into a further 37 clades. Moreover, the cDNA is itself representative of only a fraction of the RNA transcriptome that can be measured for these groups.
chemical methods” section in the Supplementary Materials) (tables at BATS station, and analyzed using metaproteomic methods (“Bio - from 10 samples of microbial biomass collected by Clio on 14 April 2018, metabolomic analysis as well. Figure 7 (A and B) shows the results analysis can only be performed on returned samples, as is true of can be made by this method (Figs. 7 and 8 and table S3). Proteomic collected at BATS station shows the biochemical measurements that The proteomic analysis of samples returned from a profile col - Clio Relative abundances of phyla in Fig. 6. DNA and cDNA samples. (Clio) DNA (Clio) cDNA. Samples are organized by depth on the x axis and relative abundance on the y axis. Two cDNA samples, AR20-5-D5 at 200 m and AR20-5-D17 at 2000 m, do not show on the graph because of either or both insufficient biomass and extraction methods resulting in low cDNA yields. Nor does this include the proteomic and metabolomic datasets that can be acquired from the same samples.

High-resolution proteomic measurements

The proteomic analysis of samples returned from a Clio profile collected at BATS station shows the biochemical measurements that can be made by this method (Figs. 7 and 8 and table S3). Proteomic analysis can only be performed on returned samples, as is true of metabolomic analysis as well. Figure 7 (A and B) shows the results from 10 samples of microbial biomass collected by Clio on 14 April 2018, at BATS station, and analyzed using metaproteomic methods (“Biochemical methods” section in the Supplementary Materials) (tables S3 and S4). From these samples, 1577 proteins were identified using high-resolution liquid chromatography–mass spectrometry from filtered volumes of 56 to 135 liters per depth (similar to Fig. 4), using one 50% split portion of the filtered biomass (15, 64). The vertical distributions of these proteins follow the niche depth range for the various microbial groups, as illustrated by a subset of proteins taxonomically attributed to Prochlorococcus, Synechococcus, and Pelagibacter (Fig. 7B).

Clio can also conduct higher-resolution sampling than existing wire-based techniques. To demonstrate this, on 24 October 2018, at BATS station, Clio collected five samples across a chlorophyll maximum at 5-m increments while holding depth to a precision of <5 cm (Fig. 7, C and D, and tables S5 to S8). In comparison, wire-deployed in situ pumps, which were collecting samples synoptically, varied in depth by more than 15 m because of high winds and waves. Metaproteomic analysis of these targeted, high-resolution Clio samples revealed distinct depth differences within the 5-m sampling increments, with numerous proteins at the top (Fig. 7E) of the chlorophyll feature being significantly more abundant than those at the center and vice versa, based on Fisher’s exact test (< 0.05) (Fig. 7F). These results are consistent with prior observations, including for several Prochlorococcus proteins: Urea and phosphate transporters were more abundant above the maximum where nitrogen is scarcest (Fig. 7, E and G), and the iron-free electron transport protein flavodoxin was most abundant at the chlorophyll maximum where iron scarcity is known to occur (Fig. 7, F and G). Nutrient stress biomarkers such as these are valued for their ability to diagnose the environmental stresses faced by each major species present and how their biochemical responses change in time and space (15). Also, different isoforms of key biochemical proteins, such as gliceraldehyde-3-phosphate dehydrogenase, were observed with depth corresponding to high- and low-light Prochlorococcus ecotypes. These results demonstrate how Clio’s unique depth keeping capabilities can lead to new observations that are otherwise very difficult to achieve.

Sectional transect

During June 2019, in a series of nine dives spanning the 1144-km transect between Bermuda and Woods Hole, Massachusetts, USA, we used Clio to collect (i) 158 sample sets for proteins, metals, and nutrients (table S9) and (ii) complementary sensor data (table S10). Clio’s optical beam transmission data indicate how low suspended material concentration was in this region (fig. S3). Consequently, Clio had to filter 20,878 liters of seawater to collect these samples. The biochemical data from this cruise are being analyzed and will be reported elsewhere. Figure 8A shows the level 3, monthly averaged, chlorophyll a estimates from the MODIS-Aqua spectroradiometer for June 2019 (see the “Remote-sensing methods” section in the Supplementary Materials), and the Clio transect stations. Figure 8 also shows the results for chlorophyll a, measured by Clio’s fluorometer, and dissolved inorganic nitrogen measured from samples collected by Clio. A comparison of the remotely sensed MODIS-Aqua chlorophyll a to the in situ Clio observations shows the presence of a deep chlorophyll maximum (DCM) layer at depths between 80 and 150 m throughout this region (Fig. 8B). This feature indicates a region of active primary productivity that cannot be observed by remote-sensing techniques and must be observed within the water column.

During this transect, we also implemented an adaptive sampling algorithm to better target discrete features like the DCM.
Clio dives 024 and 025, the depths of a subset of stations were left unspecified with instructions to sample the DCM. Clio autonomously determined the depth of the DCM on the basis of its chlorophyll fluorometer during descent. After completing several preprogrammed stations at deeper depths, Clio ascended to distribute the remaining stations at depths across the DCM. This adaptive sampling approach is similar to that taken by Zhang et al. (65) to sample thin phytoplankton layers using a conventional AUV for lateral surveys. This capability, coupled with high-resolution sampling, will enable even finer sampling resolution of discrete features like the DCM in the future.

The sectional cruise was completed in 14 days. If the same set of samples were collected without Clio, we estimate that the cruise would have taken twice as long. Moreover, this transect is a relatively short distance for ocean sectional studies; the time savings would typically be greater. During this section, Clio also completed a profile to a depth of 4100 m. These data, and Clio’s reliable performance, demonstrate that this robot can complete the sectional cruises that we designed it to perform.

Biochemical significance of trials

During the BATS and Sargasso Sea deployments, we used Clio to better understand the vertical structure of plankton communities and their relationship to available nutrients. The thousands of functional proteins observed represent a tremendous amount of information, but the characteristics of the ecosystem can be discerned from a few example proteins. Key proteins increase in response to extreme nutrient scarcity for nitrogen, phosphorus, and iron; for example, vertical distributions of a urea transporter, phosphate transporter, and flavodoxin from the cyanobacterium Prochlorococcus reflect adaptations to scarce N, P, and Fe (Fig. 7G). Similarly, proteins can be indicative of key metabolic processes such as photosynthetic carbon fixation (PsbO and CsoS2; Fig. 7, F and G) from Prochlorococcus and the use of the organic molecule taurine as a substrate for respiration in the bacterium Pelagibacter (Fig. 7B). Together, these signals can provide a powerful assessment of the metabolic capabilities of each major microbial taxon present in the community, indicating what is controlling their growth and their contribution to the marine carbon cycle at each point in space and time.

DISCUSSION

These results demonstrate that Clio can be used to observe ocean biochemistry across large sections of the ocean. We showed that the
vehicle can efficiently transit the water column while stopping to loiter precisely at depths as needed. We showed how the vehicle, its sample-processing systems, and the methods for using them can collect samples preserving genomic information with low background contamination levels. We also showed that Clio returns samples for metaproteomic analysis that can only be realized from the processing of large volumes of water, much larger than can be collected by other current AUVs. We demonstrated how Clio maps chemical and physical properties across sections of the ocean using (i) sensor data, e.g., chlorophyll, salinity, temperature, and optical backscatter, and (ii) data produced from sample analysis. We showed, as an example, how the sectional map of chlorophyll a, measured in situ by Clio’s sensor, reveals the DCM environment hidden from remote satellite sensors. We also demonstrated, using nitrate plus nitrite data, how the analysis of returned samples collected by this method can reveal the distribution of biochemically relevant ocean parameters that are difficult to measure with existing sensors. The genomic and proteomic measurements illustrate the richness of these datasets; however, the data that we showed are only a limited fraction of what can be measured in these analyses. Furthermore, the data that we showed represent only a partial set of the important biochemical compounds that could be measured with this method. Therefore, the datasets that Clio enables are in a class of “big data” that is challenging to represent. In the simplest case, if a distinct sectional map were created for every biochemical compound so measured, the result would be many 10,000 s of mapping products. The oceanographic community is now only at the beginning of the process of studying this wide range of compounds and understanding their roles in the cycling of nutrients between the environment and living systems. Our goal is that Clio, and robots like it, will contribute to a global-scale biochemical mapping of the ocean (66), allowing researchers to answer fundamental questions about the limiting constraints on ocean life and the role of ocean life in regulating Earth’s chemical environment. Moreover, our goal is that Clio will accelerate this research so that we can be better informed about how Earth’s ocean and climate are changing and what we can do to mitigate the negative consequences of these changes.

Limitations and future steps
A limitation of Clio is that it is not suited to under-ice applications. However, Clio’s sample return systems can be used on other ocean vehicles, such as the hybrid ROV NUI (67), suited to under-ice operations. Also, we intentionally designed this version of Clio with a maximum depth of 6000 m because most of the ocean is shallower than this. However, a deeper diving hadal version of Clio is possible. Notably, Clio still has substantial unused scientific payload capacity and underused computational, data storage, and battery capacity. Thus, there is strong potential to add new sensors and in situ analysis systems, develop new autonomous behaviors that can better navigate the water column biogeochemical environment, and even develop the capacity to carry out experimental treatments on marine microorganisms, or seawater chemistry, under ambient conditions.

MATERIALS AND METHODS
Vehicle design
Clio is similar in size to water sampling rosettes (Fig. 1F), which are common to oceanographic research vessels, and can be deployed, recovered, and shipped with similar equipment. On deck, the vehicle stands vertically on integral landing skids and can be launched and recovered from this position. The vehicle has an aluminum structural chassis that supports all components including the (i) two vertical thrusters, (ii) three pressure housings for electronics and batteries below, (iii) syntactic foam flotation above, (iv) four sample return payload bays around the perimeter of the midsection, and (v) sensors near the bottom (Fig. 1). The vehicle’s free-flooded internal volume, including the sample return payload bays, is enclosed in high-density polyethylene shells. These shells are removable, which provides operators easy access to the sample return systems and the power and communication cables (Fig. 2B). At the top of the vehicle is an acoustic modem that allows operators to acoustically range on the vehicle and receive regular vehicle status updates. Also, on top are three independently powered depth-activated beacons, a GPS/Iridium unit to provide an over-the-horizon location, a strobe for line-of-sight location, and a VHF (very high frequency)
radio beacon, which alert operators to the vehicle’s position when it surfaces. The computer systems, batteries, and thrusters leverage designs are inherited from previous deep-sea AUVs: ABE (68), Sentry (69), Nereus (70), and NUI (67). Clio carries a Pariscientific pressure sensor, a Seabird 49 Fastcat CTD sensor, and typically also a WET Labs 25-cm pathlength transmissometer, a WET Labs ECO FL chlorophyll a fluorometer, and an Aanderaa dissolved oxygen optode. Components are pressure rated to 6000 m or greater.

**Sampling system design**

The Clio sampling systems inherit design features and principles from previous deep-sea ROV and AUV sampling systems (43, 44, 53), but their overall design, and many features, have been uniquely developed for this open-ocean application. Specifically, they use custom high-flow rate valves to speed sample collection and minimize power consumption while pumping; the diameter of the fluidic path within the valves and the rest of the fluidic system is never smaller than 9.5 mm. Each valve has 19 ports and is actuated by a direct-drive pump (8 liters min⁻¹ pump head, Mclane Research Laboratories) and has a pressure rated to 6000 m or greater.

The flow rate through the system depends on the filter media used and amount of suspended particulate material in the sample, but typically, the average sampling flow rate achieved is 1 to 2 liters min⁻¹. The control and motor power electronics for the samplers are pressure-tolerant designs located within the fluorocarbon-filled stepper motor housings. The custom sample units that we developed to maximize sample capacity and cleanliness are described in the “Engineering materials and methods” section in the Supplementary Materials.

**SUPPLEMENTARY MATERIALS**

robotics.sciencemag.org/cgi/content/full/5/48/eabc7104/DC1

Supplementary Text: Background, materials, and methods.

Fig. S1. Clio robotic approach to ocean biochemical profiling.

Fig. S2. Abundances of putative contaminant sequences in DNA and cDNA samples.

Fig. S3. Ocean environmental structure revealed using Clio.

Table S1. Clio sampling configuration used in Sargasso Sea study.

Table S2. North Atlantic shelf break DNA and RNA.

Table S3. BATS station proteins.

Table S4. Spectral counts and annotations of proteins from Clio dive 007 and 008 BATS station. (Microsoft Excel format).

Table S5. High-resolution chlorophyll maximum sampling profile.

Table S6. Total exclusive spectral counts of selected proteins abundant at the top of the chlorophyll maximum from Clio-014 high-resolution sampling from 110 to 130 m.

Table S7. Total exclusive spectral counts of selected proteins abundant at the center of the chlorophyll maximum from Clio-014 high-resolution sampling from 110 to 130 m.

Table S8. Spectral counts and annotations of 595 identified proteins from Clio-014 high-resolution sampling 110 to 130 m at BATS station.

Table S9. Sargasso Sea dissolved inorganic nitrogen.

Table S10. Temperature, salinity, chlorophyll a, turbidity, and optical beam transmission.

References (77–95)


Acknowledgments: We thank the crews of the R/V Neil Armstrong and R/V Atlantic Explorer for assistance and support with the ocean deployments. We thank the NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group for the Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua chlorophyll data, NASA OB. DAAC, Greenbelt, MD, USA, https://oceancolor.gsfc.nasa.gov; accessed on 1 May 2020, and the ESA and developers of the SNAP tool used to illustrate these data. We thank B. Searle, J. Haskin, R. Chmiel, and M. Kellog for assistance at sea. Funding: The development of Clio was supported by U.S. National Science Foundation (NSF) grants OCE-1333212 to J.A.B., M.V.J., and M.A.S. and OCE-1334727 to G.J.D. The Sargasso Sea study was supported by U.S. NSF grants OCE-1658030 to M.A.S. and M.V.J., OCE-1657885 to R.J., and OCE-1658067 to J.A.B. U.S. NSF grants OCE-1924554 to M.A.S. and M.V.J., and OCE-1924508 to J.A.B. also supported this work. The Gordon and Betty Moore Foundation grant GBMF3782 provided support to M.A.S. U.S. NSF grants OCE-1756848 and OCE-1136477 and The Gordon and Betty Moore Foundation grant GBMF3828 provided support to A.E.A. The National Oceanic and Atmospheric Administration (NOAA) Educational Partnership Program with Minority-Serving Institutions Cooperative Agreement Award #NA16SEC4811009 provided support to J.A.B. and B.A.A.; the contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Commerce, NOAA. Author contributions: J.A.B. designed the sampling systems. M.V.J. designed the vehicle. The overall vehicle concept and use case scenarios were envisioned by J.A.B., M.A.S., M.V.J., and G.J.D. The proteomic methodology and associated metagenomic libraries were developed by M.A.S., M.M.R., D.M.M., A.E.A., and C.L.D., and they performed the analyses. G.J.D. and S.L.G. developed the DNA/RNA methodology and performed the analyses, R.J., M.V.J., M.A.S., M.M.R., S.L.G., E.W.C., B.A.A., and J.A.B. developed the field oceanographic methods and performed the fieldwork. J.A.B. wrote the manuscript with contributions from all authors. Competing interests: The authors declare that they have no competing interests. Data and materials availability: The hydrographic, nutrient, and protein data for this study have been deposited in Biological and Chemical Oceanography Data Management Office (www.bco-dmo.org) (project 763945), and raw mass spectrometry and DNA sequence datasets have been deposited at the EBI PRIDE (project PXD018067) and NCBI repositories (project PRJNA607796), respectively. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.
Revealing ocean-scale biochemical structure with a deep-diving vertical profiling autonomous vehicle

John A. Breier, Michael V. Jakuba, Mak A. Saito, Gregory J. Dick, Sharon L. Grim, Eric W. Chan, Matthew R. McIlvin, Dawn M. Moran, Brianna A. Alanis, Andrew E. Allen, Chris L. Dupont, Rod Johnson

Sci. Robot., 5 (48), eabc7104. • DOI: 10.1126/scirobotics.abc7104

View the article online
https://www.science.org/doi/10.1126/scirobotics.abc7104
Permissions
https://www.science.org/help/reprints-and-permissions