INTRODUCTION

The marine microbial carbon pump (MCP) [1] produces a concentration gradient of reactivity in dissolved organic carbon (DOC) from low concentrations of reactive fractions of DOC that have an average lifetime of less than 100 years ($\text{DOC}_{<100}$) to high concentrations of recalcitrant DOC (RDOC) fractions that persist for longer than 100 years ($\text{DOC}_{>100}$) [2]. This gradient is maintained against the continuous degradation of DOC by photochemical transformation in surface waters, elevated temperatures at deep hydrothermal vents and gradual microbial degradation throughout the water column [1–3]. The resultant pool of RDOC amounts to the sequestration of carbon in the ocean equivalent to the amount of carbon dioxide in the atmosphere. The magnitude of the MCP is determined as the rate of production of $\text{DOC}_{>100}$ in mol C m$^{-2}$ y$^{-1}$ [2]. Global warming, weakening thermohaline circulation, increasing UV radiation, ocean acidification, increasing oxygen minimum zones and increasing nutrient concentrations are all processes potentially affecting the magnitude of the MCP, leading to increased or decreased transformation of marine RDOC to atmospheric CO$_2$ with associated consequences for global climate [3,4].

In order to quantify the MCP, and understand its sensitivity to large-scale environmental change, the rates of production and loss of RDOC via each potential formation and loss pathway need to be determined and understood. This is a major challenge, likely beyond the ability of any individual team of researchers, given the complexity and diversity of the substances, processes and microbial organisms involved and the variety of controlling factors. However, the connectivity and biogeographical characterization of the ocean—a consequence of ocean circulation and mixing—allow a co-ordinated approach where results from multiple groups, working at different locations and times, could be compared and assembled in order to achieve progress.
Such a co-ordinated approach will require a multi-component programme of systematic experimentation linked with related conceptual and numerical modelling. The suggested programme components include: (i) determination of the chemical components of RDOC, including ‘fingerprint’ constituents to identify the presence of RDOC; (ii) a standardized and intercomparable programme of long-term microcosm incubation experiments conducted worldwide to determine loss of labile DOC, and net accumulation or loss of bulk RDOC, under specific environmental conditions; (iii) the concurrent determination and quantification of RDOC production and loss pathways; and (iv) quantification of the presence and activity of the responsible microbial organisms and their genetic potential.

In order to test our understanding as well as the efficacy of our experimental protocols, a linked ‘backbone’ programme of larger-scale, controlled manipulation experiments and related, long-term time-series studies should be established (v). This should involve controllable meso- and macro-cosm facilities as well as in-situ time-series studies where the determination of rate processes, RDOC composition change (and associated proxies), microbial community composition and genetic potential can be examined repeatedly, under varying conditions. The programme needs to be structured within (vi) a framework of conceptual and numerical models and, crucially, in order to assess the global context of RDOC production and loss, (vii) large spatial-scale and long temporal-scale surveys of, at minimum, the fingerprints or proxies of RDOC accumulation will be required. The surveys will allow experimental findings to be parameterized in global biogeochemical models to explore the sensitivity of the MCP and associated carbon storage to environmental change over long timescales.

For each of these programme components, we describe here the principle of the approach, current limitations and future research requirements. This perspective paper will therefore provide a route by which to progress from the conceptual framework of the MCP, to the quantification of the microbial processes that produce and utilize RDOC, and the environmental, trophic and physiological influences on them.

Identification of the components of RDOC

The very nature of RDOC, in terms of its complexity, diversity, low concentration and longevity, means that identifying representative components of RDOC is a major analytical challenge.

A variety of methods are now available to extract low-molecular-weight, \(^{14}\)C-depleted RDOC and nuclear magnetic resonance (NMR) spectrometry, gas chromatography–mass spectrometry (GC–MS) and Fourier transform ion cyclotron resonance mass spectrometry (FT–ICR–MS) can identify specific chemical fingerprints of this material, including carboxyl-rich alicyclic molecules (CRAM), material derived from linear terpenoids and a group of stable molecular formulae known as the ‘island of stability’.

However, the comprehensive analysis of DOC and quantification of RDOC are still far from complete. Current techniques are limited in that existing isolation methods including solid phase extraction (SPE) can only isolate representative fractions of DOC, some of which are further discriminated during mass spectrometry analysis. In addition, direct injection of the DOC sample into the mass spectrometer cannot differentiate the numerous possibilities of isomer configuration for an assigned molecular formula, which therefore precludes the identification of an unambiguous chemical structure [5].

Future approaches should include the use of combined methods for DOC isolation and the development of new techniques to isolate the DOC (Fig. 1) [6]. Chromatographic analysis prior to ultra-high-resolution mass spectrometry could help to differentiate isomers, and subsequent analyses by a combination of FT–ICR–MS and high-field NMR would allow progress towards the ultimate goal of revealing the structures and concentration of representative RDOC molecules.

Incubation experiments to determine loss of labile DOC and accumulation of RDOC

Operationally, the reactivity of DOC is determined from the loss of DOC over a given period of time [7], with a biologically labile fraction (LDOC) defined as that which is degraded over hours to days, a semi-labile fraction (SLDOC) that is removed in weeks to months and a semi-refractory fraction (SRDOC) with turnover times of years to decades. Studies of the degradation of LDOC involve the incubation of seawater samples under controlled laboratory conditions for days to years, allowing the calculation of the DOC decay constant \(k_C\) over time [8].

However, there is no standard method by which to undertake these experiments and, since both the concentration and decay rate of DOC are dependent on time and incubation conditions such as temperature and nutrients, collation and comparison of the available data are problematic. Methodological differences include the method by which a diluted seawater sample is obtained, such as the composition and pore size of the filters used, the ratio of filtrate to inoculum, the incubation temperature, light and nutrient conditions, and the volume and time period of the incubation.

A comparison of methods and recommendations for standardization are required to advance this aspect of DOC reactivity research, following the approach taken recently by the research community investigating the impact of ocean acidification on marine plankton activity [9]. Once standardized methods are defined, systematic laboratory-scale bioassay experiments can be undertaken to assess the influence of environmental conditions such as temperature, light and nutrients on DOC reactivity, and incorporation of measurements of RDOC components or proxies of RDOC will allow the accumulation of RDOC to be determined alongside the loss of DOC.
Three broad areas of research this persistence are still being explored. Ocean is well known, but the reasons for production of bacterial RDOC in the upper and bioassay experiments, the rate of production of ‘model’ RDOC and 50% of the nitrogen is of bacterial origin [10]. Based on biomarker data and bioassay experiments, the rate of production of bacterial RDOC in the upper 200 m of the ocean is estimated to be 8–23 Tg C year⁻¹ [11].

The persistence of RDOC in the ocean is well known, but the reasons for this persistence are still being explored [4,12]. Three broad areas of research can identify which organisms are responsible for any observed biodegradation pathway with high phylogenetic resolution. Thus, incubation experiments with additions of ¹³C-labelled labile DOC and subsequent ¹³C- and ¹²C-DNA analysis with terminal-restriction fragment length polymorphism (T-RFLP) and pyrosequencing can identify the microbial populations that incorporate the ¹³C-labelled substrate. The molecular characteristics of the produced ¹³C-labelled DOC (relatively more refractory) can be further determined using FT–ICR–MS, NMR or excitation emission matrix (EEM) fluorescence. Unfortunately, SIP techniques are not inherently quantitative. Substrate assimilation rates can be derived from changes in δ¹³POC (particulate organic carbon) during concurrent incubations to help link the DOC-specific degradation rates to the activity of the microbial population.

Future work should use laboratory-controlled phytoplankton, zooplankton and bacterial cultures to produce DOC from each potential RDOC formation pathway to identify potential chemical biomarkers for each. The recent connection between RDOC fractions and precursors such as phytoplankton carotenoids lays the foundation for hypothesis testing of these potential transformation pathways. Long-term degradation incubations incorporating DNA/RNA SIP could then be conducted to assess the bioreactivity of DOC from each of these formation pathways.

**Microbial community composition and genetic potential**

The microbial transformation of short-lived particulate and dissolved organic carbon (DOC<100 and DOC<100) to DOC>100 is ultimately driven by the activity of microbial functional genes [16]. For example, genes for the ATP-binding cassette (ABC) transporter, including importers and exporters, are thought to be indicators of the capabilities of microbes to use or generate corresponding DOC compounds [16]. Therefore, a strategy to quantify the various processes are being investigated: (i) intrinsic properties of molecules, such as chemical composition and structure; (ii) environmental conditions, such as microbial community structure and exposure to solar radiation; and (iii) concentrations of molecules that are below a critical threshold for microbial utilization. Multiple and independent approaches are needed to delve deeper into the mysteries of RDOC removal from the ocean. For example, recent bioassay experiments have provided evidence that some RDOC is too diluted for microbial utilization [13] and that intrinsic properties of RDOC and environmental conditions limit its microbial utilization [3]. A modelling approach addressing these same questions provided independent evidence indicating the dilution of labile substrates likely accounts for a relatively small fraction of the RDOC in the deep ocean [14]. It appears that most RDOC persists in the deep ocean due to intrinsic properties and environmental conditions. Ocean mixing moves RDOC into different environmental conditions, and exposure to solar radiation and different microbial communities plays a critical role in its removal [3].

Stable isotope probing (SIP) [15] can be used to ascertain which organisms are responsible for DOC degradation. SIP tracks the incorporation of isotopically enriched substrates into the DNA, RNA or phospholipid fatty acids of these organisms. Combined with high-throughput sequencing, DNA/RNA SIP
of the MCP could depend on the quantification of microbial functional genes.

The quantification methods (such as PCR, qPCR, omics) usually used in ecological studies are time-consuming and potentially incomplete when considering the extremely high diversity of microbial functional genes involved in MCP processes. Microarrays that contain numerous probes targeting hundreds or thousands of genes are a promising technique in this regard. The development of a functional gene microarray involves gene selection and sequence retrieval from a database, probe design and evaluation, and microarray fabrication.

GeoChip is one of the most comprehensive functional gene arrays for quantification of the functional diversity, metabolic potential/activity and dynamics of microbial communities [17]. GeoChip 4 contains ~82,000 probes covering >140,000 coding sequences from 410 functional gene families, many of which are encoding biogeochemically important functions, such as carbon fixation and carbon degradation. In addition, GeoChip 4 has the ability to analyse targeted functional gene families of microorganisms in all four domains (Bacteria, Archaea, eukaryotes and viruses). A recent case study performed in the East China Sea suggested a possible connection between GeoChip data and MCP processes such as the potential formation of a pycnocline [18]. The global ocean is experiencing a slowdown in overturning circulation, increasing temperature, decreasing pH, increasing UV radiation, increasing dissolved CO2, decreasing dissolved O2 and changing inorganic and organic nutrient availability [22], with as-yet unknown consequences for ocean carbon storage. A combination of laboratory microcosm- (10−3 to 1 m3), mesocosm- (1 to 103 m3) and macrocosm- (>103 m3) scale manipulation experiments are required to test hypotheses related to the individual and combined effects of these multiple drivers on the MCP [23]. Mesocosms deployed at sea and macrocosms on land have been used successfully to investigate the effects of ocean acidification and nutrient additions (via, for example, atmospheric dust deposition) on marine plankton activity [24]. The range of scale is required due to the trade-off between numerous replication and full factorial matrices available at the macrocosm scale and improved representation of the complexity and depth distribution of the microbial community, including the potential formation of a pycnocline, at the macrocosm scale [23,25].

Future studies of the MCP should use at-sea mesocosms and on-land macrocosms to bridge the gap between data derived from microcosm bioassay experiments and in-situ measurements. A project to build a facility with 9 × 50 m tall macrocosms to allow experiments to be performed with three replicate controls and three replicates of each of two treatments is currently underway in Qingdao, China [25]. This advanced experimental system will allow the concurrent determination of microbial diversity, the chemical composition of DOC and the rate processes affecting the production and loss of RDOC over relevant space scales of euphotic zone depths and time scales of years, under two environmental treatments in a previously unobtainable statistically robust manner [23]. In addition to advances in process understanding, experiments conducted in meso- and macrocosms can be used as a basis for systematic comparison of methods used elsewhere (components (i) through (iv) above). The design of experiments should be linked directly, when possible, with hypotheses and process-related questions that emerge from field studies, including time-series

**Figure 2.** An experimental microarray procedure to link gene analysis with the quantification of MCP processes.
studies of naturally variable systems (component (vii) above).

Modelling framework
The dynamics underpinning the MCP are rarely considered in marine ecosystem models and this limits our capacity to predict carbon sequestration in future oceans [26]. While exploratory modelling studies on the MCP have been published recently [26,27], we need to embed these studies in a coherent framework if we want to have reliable MCP simulations. We propose that, in future projects, each aspect of the model development (formulation, calibration, validation) is carried out in co-ordination with empirical studies, in order to exploit the multiple feedbacks between data and simulations to iteratively refine the model and achieve reliable MCP predictions (Fig. 3). Laboratory experiments (i.e. microcosm and mesocosm batch cultures and chemostats) using the techniques described earlier are able to provide a detailed description of RDOC production and microbial consumption in a controlled environment. The understanding generated by these experiments allows the formalization of a conceptual model (i.e. a diagram describing the process being investigated). The conceptual model can be seen as the central hypothesis driving subsequent modelling steps. The conceptual model then needs to be ‘translated’ into a mathematical form (process model) and calibrated to reproduce the dynamics observed in the experiments. The development of the process model might highlight knowledge gaps and inform new, more informative, experimental designs. When the process model successfully reproduces the dynamics observed in the experiments, it is ready to be implemented in a full ecosystem model. Field studies (including incubation experiments) provide measurements (standing stock concentrations and rates) to scale the conceptual model up to the ecosystem level and to validate the ecosystem model. Eventually, after the ecosystem model is satisfactorily validated, it can be used for climate predictions. During the validation of the ecosystem model, it is possible to highlight knowledge gaps and inconsistencies which imply that the conceptual model should be revised and new experiments planned. Mesocosms and macrocosms represent a valuable compromise between the laboratory and the real environment, providing additional information to refine the conceptual model, inform the development of the process model and validate the ecosystem model.

Global and decadal scale perspective
In order to predict future changes in the capacity and efficiency of the MCP, global spatial-scale and decadal temporal-scale surveys of robust, high-throughput proxies of RDOC accumulation will be required alongside chemical fingerprints of RDOC transformation pathways, such as D-enantiomers of amino acids, and the more analytically challenging comprehensive chemical characterization of DOC. High-throughput proxies might include the fluorescence intensity of humic-like chromophoric dissolved organic material (CDOM), the production of which correlates with apparent oxygen utilization (i.e. microbial degradation of organic material) in the deep ocean.

Global and decadal monitoring of the MCP would require the development of sensors for components of the DOM pool (e.g. CDOM, fDOM, DOC, RDOC fingerprints) appropriate for deployment on a range of platforms including moorings, gliders, volunteer observing ships (such as ferries and container ships) and biogeochemical ARGO floats (http://www.argo.ucsd.edu/), as well as remote sensing algorithms for surface distributions. An international co-ordinated programme would be required, including standardized methods, inter-calibration exercises, and scientific training and capacity building. This should be embedded within established time-series stations such as the Hawaii Ocean Time-series (HOT; http://hahana.soest.hawaii.edu/hot/) and those within the OceanSITES (http://www.oceansites.org/index.html) worldwide network, as well as global shipboard surveys such as those co-ordinated within GO-SHIP (http://www.go-shipt/index.html) and GEOTRACES (http://www.geotraces.org/) to ensure that appropriate biogeochemical, biodiversity and omics data are collected concurrently. The Global Ocean Observing System (GOOS; http://www.gooscean.org/) Biogeochemistry Panel has recognized the need for sustained ocean observations of DOC, designating it as an essential ocean variable (EOV) required for addressing societal drivers such as the role of biogeochemistry in climate and human impacts on ocean biogeochemistry. The future aspiration should be to designate RDOC or a proxy of DOC_lability as an EOV.

Progress in MCP quantification will also rely on continued advances in quantification of the microbial processes related to DOC and RDOC production and transformation, including heterotrophic bacterial respiration, viral lysis and chemosynthesis. Improvements in bioinformatics and data manipulation and visualization tools will enhance the opportunities for connections to be made between the molecular complexity of RDOC and meta-genomics, -proteomics and -transcriptomics [20].

Given the complexity and diversity of measurements involved (i–iii) and the value of repeated measurement and experimentation for model validation (vii), a logical starting point for a systematic MCP field study would be the addition of MCP appropriate measurements at the locations of already established moored- and ship-based multidisciplinary ocean time series that examine...
a diverse range of ocean ecosystems and biogeochemical conditions. Overall, MCP quantification would be most efficiently achieved through a co-ordinated international research initiative linking each of these programme components (i–vii) and including appropriate policies for method standardization and inter-calibration, data-management, capacity-building, outreach and education, science-policy-society activities to address the societal driver of understanding the role of marine biogeochemistry in climate (Fig. 4).

CONCLUSION

Quantitative understanding of the mechanisms that produce and utilize RDOC is still in its infancy, despite their fundamental role in global climate. Significant progress towards quantification of the MCP requires an internationally co-ordinated multidisciplinary research initiative that interlinks in-situ time-series studies and micro- and macro-scale manipulation experiments within a robust conceptual to numerical modelling framework. The analytical tools and intellectual interaction between chemists and microbiologists are now at a stage where this societally driven question can, and must, become a high priority research topic.

ACKNOWLEDGEMENTS

We sincerely thank the participants of the first Yan-qi Lake conference and the joint International Council for the Exploration of the Sea/North Pacific Marine Science Organisation (ICES/PICES) working group on Climate Change and Biologically-driven Ocean Carbon Sequestration (WGCCBCOS; http://www.ices.dk/community/groups/Pages/WGCCBCOS.aspx, http://meetings.pices.int/members/working-groups/wg33).

FUNDING

This work was supported by the United Kingdom Natural Environment Research Council (NERC) (NE/R000956/1 awarded to C.R.), the Leverhulme Trust (RPG-2017-089 awarded to C.R.), the Korean Long-term Marine Ecological Research (K-LTMER) Program funded by the Korean Ministry of Oceans and Fisheries (awarded to J.-H.-H.) and the United States of America National Science Foundation (1233373 awarded to R.B.). L.P. was funded by the UK NERC National Capability for Marine Modelling, Y.Z. by the National Key Research and Development Programs of China (2016YFA0601400) and D.W. was supported by the Canada Excellence Research Chair in Ocean Science and Technology.

REFERENCES

12. Dittmar T. Reasons behind the long-term stabili-ty of dissolved organic matter. In: Hansell DA and

Figure 4. Schematic of a co-ordinated international research initiative linking each of these programme components (i–vii) with data-management, capacity-building and science–policy–society activities.


National Science Review 5: 474–480, 2018
doi: 10.1093/nsr/nwy070
Advance access publication 7 July 2018