Published in final edited form as:

Environ Microbiol. 2009 January ; 11(1): 230–238. doi:10.1111/j.1462-2920.2008.01758.x.

Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11

H. James Tripp^{1,*}, Michael S. Schwalbach^{1,*}, Michelle M. Meyer², Joshua B. Kitner¹, Ronald R. Breaker^{2,3}, and Stephen J. Giovannoni^{1,§}

¹ Department of Microbiology, Oregon State University, Corvallis, OR 97333

² Department Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520

³ Howard Hughes Medical Institute, Yale University, New Haven CT 06520

Abstract

The genome sequence of the marine bacterium ‘*Candidatus Pelagibacter ubique*’ and subsequent analyses have shown that, while it has a genome as small as many obligate parasites, it nonetheless possesses a metabolic repertoire that allows it to grow as one of the most successful free-living cells in the ocean. An early report based on metabolic reconstruction indicated that SAR11 cells are prototrophs for all amino acids. However, here we report experimental evidence that ‘*Cand. P. ubique*’ is effectively auxotrophic for glycine and serine. With glucose and acetate added to seawater to supply organic carbon, the addition of 125 nM to 1.5 μ M glycine to growth medium containing all other nutrients in excess resulted in a linear increase in maximum cell density from 1.14×10^6 cells ml^{-1} to 8.16×10^6 cells ml^{-1} ($R^2 = 0.992$). Serine was capable of substituting for glycine at 1.5 μ M. ‘*Cand. P. ubique*’ contains a glycine-activated riboswitch preceding malate synthase, an unusual genomic context that is conserved in the SAR11 group. Malate synthase plays a critical role in central metabolism by enabling TCA intermediates to be regenerated through the glyoxylate cycle. *In vitro* analysis of this riboswitch indicated that it responds solely to glycine but not close structural analogs such as glycine betaine, malate, glyoxylate, glycolate, alanine, serine, or threonine. We conclude that ‘*Cand. P. ubique*’ is therefore a glycine-serine auxotroph that appears to use intracellular glycine level to regulate its use of carbon for biosynthesis and energy. Comparative genomics and metagenomics indicate that these conclusions may hold throughout much of the SAR11 clade.

Introduction

Since SAR11 bacteria are the most abundant and cosmopolitan heterotrophic bacteria in the world’s oceans (Morris et al., 2002), their carbon metabolism is of global ecological importance. Initial studies suggested that SAR11 possessed all the genes for amino acid biosynthesis, including glycine and serine, based on the presence of a putative threonine aldolase to compensate for missing *serBC* genes (Giovannoni et al., 2005). However, we noted in our studies of sulfur limitation that the addition of nanomolar amounts of glycine betaine to media containing a mixture of other carbon sources in micromolar concentrations increased maximum cell densities (Tripp et al., 2008). This observation combined with recent reports indicating that members of the SAR11 clade possess multiple putative glycine-

[§]Correspondence and requests for materials should be addressed to SJG (steve.giovannoni@oregonstate.edu).

*These authors contributed equally

binding riboswitches (Kazanov et al., 2007) prompted us to further investigate glycine-serine metabolism in SAR11.

The carbon compounds glycine and serine can account for up to 15% of carbon assimilated by *E. coli* growing on glucose (Pizer and Potochny, 1964). Apart from their direct use in protein synthesis, glycine and serine serve as precursors for cysteine, purines, heme groups, phospholipids, enterochelin, and tryptophan (Ravnikar and Somerville, 1987), and as sources of methyl groups for methionine and thymine biosynthesis. Thus, they are precursors for one of the main biosynthetic trunks branching from the glycolysis/gluconeogenesis backbone of central metabolism. In *E. coli*, glycine and serine biosynthesis proceeds from the glycolytic intermediate 3-phosphoglycerate to serine, catalyzed by the products of the *serABC* genes, then from serine to glycine, catalyzed by the product of the *glyA/shmt* gene. This path is utilized by *E. coli* when either glucose or acetate are used as sole carbon sources (Zhao and Shimizu, 2003). Two alternate pathways are known for glycine and serine biosynthesis in heterotrophs. Yeast and *E. coli* mutants that are missing the *glyA* or *SHMT* genes possess a threonine aldolase (*ltaE* or GLY1) that can degrade threonine to glycine, (Monschau et al., 1997).

Comparison of the putative pathways of glyoxylate metabolism in ‘*Cand. P. ubique*’ to the known pathways in *E. coli* suggested that glycine plays a novel regulatory role in the glyoxylate bypass of ‘*Cand. P. ubique*.’ *E. coli* has two operons for glyoxylate metabolism that are cross-induced by acetate and glycolate: *glcCDEFGB* and *aceBAK* (Pellicer et al., 1999). Typically, glyoxylate cycle activity is up-regulated during growth on acetate, to increase the flow of two carbon units from acetyl-CoA into biomass via the intermediates of the TCA cycle. The *glcCDEFGB* operon, which encodes the oxidation of glycolate to glyoxylate, and the formation of malate from glyoxylate and acetyl CoA, is controlled by *glcC*, which is missing from ‘*Cand. P. ubique*.’ The *aceBAK* operon, which encodes the cleavage of isocitrate by isocitrate lyase to form glyoxylate, then malate, is controlled in part by the product of *aceK*, which is also missing from ‘*Cand. P. ubique*.’ In fact, ‘*Cand. P. ubique*’ retains only the *aceA* (isocitrate lyase) gene from the first operon, and the *glcDEF* (glycolate oxidase) cluster and *glcB* (malate synthase) gene from the second operon. This affords ‘*Cand. P. ubique*’ the capacity for making malate from either glycolate or isocitrate using the glyoxylate bypass, but leaves it without the key regulatory enzymes that control activation of the bypass in *E. coli*. In their place is a novel glycine-activated riboswitch upstream of *glcB* (malate synthase), suggesting that glycine plays a novel role in controlling the glyoxylate bypass and thus the central carbon metabolism of ‘*Cand. P. ubique*.’

To further investigate glycine-serine metabolism in SAR11, we measured maximum cell density in response to additions of glycine, serine, and their potential precursors with all other nutrients provided in excess. As in an earlier study (Tripp et al., 2008), we inferred that a nutrient is required if maximum cell density increased in response to that nutrient when all other nutrients were provided in excess. We also performed *in vitro* structural probing assays on the unusual glycine riboswitch located upstream of malate synthase, as well as the common glycine riboswitch upstream of the glycine cleavage complex, to determine ligand specificity and sensitivity. In an attempt to determine how widely distributed the malate synthase riboswitch is, we searched for similar sequences in the NCBI non-redundant protein database and on environmental DNA fragments originating from members of the SAR11 group.

Results and Discussion

Genomic Analysis of Glycine and Serine Biosynthesis

Analysis of the SAR11 genome and metagenomic data indicated that genes for the common pathway of bacterial glycine and serine biosynthesis are missing with the possible exceptions of threonine aldolase and phosphoglycerate dehydrogenase (Table 1). The gene at locus SAR11_1366 is a member of COG0111 (phosphoglycerate dehydrogenase and related dehydrogenases). While it could be a homolog to D-3-phosphoglycerate dehydrogenase (*serA*), the absence of genes *serBC* indicates that it is likely a dehydrogenase of unknown function. The gene at locus SAR11_0689 is a homolog of a low-specificity threonine aldolase in *E. coli* (*ItaE*), which supports only poor growth in *E. coli* when glycine biosynthesis is suppressed by knocking out serine hydroxymethyltransferase (Liu et al., 1998). However, a similar threonine aldolase in yeast (GLY1) is able to support all glycine and serine biosynthesis (Monschau et al., 1997). Glycine hydroxymethyltransferase (*glyA*), which can interconvert glycine and serine via methylation or demethylation, is present in the ‘*Cand. P. ubique*’ genome (Table 1). Hence, the question of whether or not SAR11 is a glycine prototroph hinges on the ability of the threonine dehydrogenase to supply sufficient glycine for all cellular demand, an issue that required experimentation to resolve.

Response to Additions of Glycine

Batch culture experiments involving addition of glycine demonstrated that ‘*Cand. P. ubique*’ is in fact a functional glycine/serine auxotroph under the conditions tested. Figure 1 shows a linear, eight-fold increase in maximum cell density from 1.14×10^6 cells ml^{-1} to 8.16×10^6 cells ml^{-1} ($R^2=0.992$) in response to the addition of nanomolar glycine concentrations, with all other nutrients present in excess. Direct measurement of dissolved free amino acids (DFAA) indicated that nanomolar concentrations of serine (18.5 ± 2.4 nM), glycine (42.4 ± 1.2 nM) and other amino acids were present in seawater media following filtration and autoclaving procedures. These concentrations are comparable to *in situ* concentrations reported by others (Mopper and Lindroth, 1982; Fuhrman and Ferguson, 1986; Pomeroy et al., 1990). We postulate that these DFAA supported the growth observed in control cultures to which no glycine was added. The phrase “conditional auxotroph” has been proposed to describe yeast mutants that are auxotrophs under certain defined conditions (Monschau et al., 1997). This concept might usefully be applied to ‘*Cand. P. ubique*’.

Glycine-activated Riboswitches

Glycine riboswitches are typically located upstream of the glycine cleavage system (*gcvTHP* operon), and activate gene expression in the presence of glycine, ensuring that glycine is not degraded unless its intracellular concentration rises above a certain threshold (Barrick et al., 2004). However, we identified a second putative glycine riboswitch in the intergenic region preceding the glyoxylate cycle enzyme malate synthase (*glcB*), which was subsequently reported by (Kazanov et al., 2007), and decided to confirm its ligand. We aligned upstream regions of *glcB* genes from a collection of Sargasso Sea environmental sequences with best hits to ‘*Cand. P. ubique*’ and found remarkable conservation of nucleotides (Fig. 2). This conservation allowed us to confirm that the folding pattern of the mRNA for this intergenic region conforms to the consensus sequence of a tandem, cooperative, glycine-binding riboswitch (Mandal and Breaker, 2004), albeit one which is some 20% shorter in length than other reported glycine-binding riboswitches, possibly due to genome streamlining. A search of the NCBI nr database revealed only one other sequenced organism (*Granulibacter bethesdensis*CGDNIH1) with a putative glycine riboswitch upstream of malate synthase.

To determine the ligand identity of the putative glycine-binding riboswitch associated with malate synthase, we conducted *in vitro* structural probing assays in the absence of ligand and

in the presence of a variety of potential ligands including glycine. The assay takes advantage of differing spontaneous RNA cleavage rates to detect structural changes that result from ligand binding to the RNA (Soukup and Breaker, 1999). Structured regions of the RNA are typically more protected from cleavage than unstructured regions. Upon ligand binding, the RNA structure changes resulting in different spontaneous cleavage products. From the *in-vitro* assays, it is clear that the malate synthase associated riboswitch in '*Cand. P. ubique*' selectively binds glycine (Fig. 3A). No other ligands tested induce structural changes in the RNA at concentrations up to 1 mM, indicating that they likely do not interact with the RNA to alter gene expression.

In addition to the glycine riboswitch associated with malate synthase, a similar riboswitch was identified upstream of *gcvT* in '*Cand. P. ubique*'. Despite differences in the sequence length between the riboswitches and the predicted translation start site, both riboswitches appear to utilize a ribosomal binding site occlusion mechanism for gene regulation as no intrinsic terminator stems are apparent, and there are several opportunities for purine rich sequences preceding the predicted translation start site to base pair with CU-rich stretches within the riboswitches. While riboswitches are often kinetically rather than thermodynamically driven, thus obscuring the physiological relevance of K_D measurements (Wickiser et al., 2005), the malate synthase-associated riboswitch has a K_D approximately 10-fold lower than the *gcvT* riboswitch (25 μ M vs. 350 μ M Fig. 3B, C) and the K_D of both riboswitches is significantly higher than that observed for the *B. subtilis* glycine switch (1 μ M) (Mandal and Breaker, 2004). Also, the malate synthase riboswitch shows a more linear dissociation than the glycine cleavage riboswitch, as seen in Figure 3b, and as quantified by the Hill coefficients (1.1 for malate synthase and 1.4 for glycine cleavage). This means that the malate synthase riboswitch is less abrupt in its activation of the downstream gene, leading to a more gradual upregulation in response to increasing concentration of ligand.

The novel riboswitch on malate synthase appears to regulate the flow of carbon through the TCA cycle to biosynthesis, using glycine concentration as a cue. We postulate that it is a fail-safe switch that prevents anabolic pathways from competing with respiration for nutrients when intracellular glycine concentrations fall to maintenance levels. While this hypothesis is consistent with our observations, it raises the question, what selective pressures could have led to subfunctionalization of the glycine riboswitch to regulate biosynthetic output from TCA cycle, when glycine appears to be indirectly involved? As noted above, glycine and serine account for ca. 15% of microbial biomass (Pizer and Potochny, 1964), and glycine is a degradation product of choline, threonine, and the important marine osmolyte glycine betaine (Kiene and Williams, 1998). Glycine and its precursors glycine betaine and choline are likely to be at least as abundant in the ecosystem as any individual amino acid (Roulier et al., 1990; Kiene, 1998; Keller et al., 2004). Glycine, in particular, is unique because it is not produced endogenously by SAR11 cells, so unlike many other metabolic intermediates, glycine levels are probably controlled by the exogenous supply and the demands for protein synthesis. Moreover, NMR analysis of marine plankton revealed that proteins typically make up 50% to 70% of their weight while carbohydrates make up only 15% to 25% (Hedges et al., 2002). Given that glycine is the second-most frequently occurring amino acid in proteins (Mathews et al., 2000), its presence in the water column could well act as a signal that the most abundant fraction of plankton biomass - proteins - are present in the water. Viewed in this light, it is not surprising that glycine could be used as a proxy for elevated levels of important nutrients in general.

Glycine Substitution Experiments

Serine was the only compound tested that could substitute for glycine as a growth factor (Fig. 4). The ability of serine to substitute for glycine is consistent with the genome annotation of '*Cand. P. ubique*,' which identifies a glycine hydroxymethyltransferase

(SAR11_1048) that interconverts glycine and serine. The failure of glyoxylate to substitute for glycine is consistent with the finding that the genome contains no *agxt* or *spt* genes to transaminate glyoxylate using serine or alanine as amino donors. It also ruled out the possibility that an aspartate transaminase (*aspC*) gene upstream of the *glcDEF* operon, might allow formation of glycine from glyoxylate originating from glycolate, a photorespiration product found in seawater (Leboulanger et al., 1994; Leboulanger et al., 1997; Leboulanger et al., 1998). Therefore we rejected the hypothesis that the two glycine switches worked in tandem to maintain intracellular glycine levels by synthesizing glycine via glyoxylate when we found that the addition of glyoxylate or glycolate had no effect (Fig. 4). The failure of threonine to substitute for glycine is consistent with the hypothesis that the threonine aldolase identified in the genome annotation does not have sufficient activity to satisfy cellular glycine and serine demands. However, an alternate explanation is that threonine aldolase could not maintain intracellular glycine concentrations at levels sufficient to activate the glyoxylate cycle and allow organic acid assimilation into biomass.

While glycine betaine could not substitute for glycine at 1.5 μM in these experiments (Fig. 4), 100 nM glycine betaine was capable of supporting maximum yields in the range of 10^7 cells ml^{-1} in other experiments using a broader mix of dilute carbon compounds as primary carbon sources (Tripp et al., 2008). The effectiveness of low concentrations of glycine betaine under some carbon source conditions and its inhibitory effect under others is still under investigation.

Conclusion

The metabolic and regulatory features we describe are unusual, simple, and indicate a key role for glycine in SAR11 carbon metabolism. The evolution of these cells has clearly adapted them to compete successfully for dissolved organic carbon in highly oligotrophic marine ecosystems. These new observations support the broad conclusion that metabolic versatility has been sacrificed for simplicity and genome reduction, rendering SAR11 cells able to use ambient nutrient resources efficiently but reducing their versatility. We show that genes for glycine and serine biosynthesis are missing, and, with the compounds glucose and acetate providing carbon, glycine or serine are required by ‘*Cand. P. ubique*’ for growth. We also show that an unusual riboswitch that is activated by glycine mediates synthesis of malate synthase and the flux of carbon through the TCA cycle into biosynthesis, replacing common control mechanisms. We postulate that these features operate collectively to maximize cell yield when multiple carbon compounds are simultaneously available, and regulate metabolism at a fundamental level to conserve resources when the flux of exogenous compounds into the cell is too low to support growth.

One of the limitations of our study is the dearth of information about the composition of marine dissolved organic carbon, which is possibly the largest pool of reactive carbon on the planet (Benner, 2002). Our study points to an important role for glycine and related compounds, and may stimulate further study of the environmental distributions of these molecules.

Experimental Procedures

Metabolic Reconstruction and Missing Genes

All biosynthetic pathways in the KEGG diagrams (Kanehisa et al., 2004) for the sequenced and annotated genome of ‘*Cand. P. ubique*’ (Giovannoni et al., 2005) were compared to the same KEGG diagrams for *E. coli*. When it emerged that the *serBC* genes from *E. coli* were not annotated for ‘*Cand. P. ubique*,’ the nucleotide sequence of ‘*Cand. P. ubique*’ (NCBI accession number NC_007205) was searched for the *E. coli*, *Caulobacter crescentus*, and

Silicibacter sp. TM1040 nucleotide sequences for *serBC* (NCBI accession numbers gi|49175990:4622918-4623886, gi|49175990:956876-957964, gi|99079841:3186312-3187187, gi|56694928:3552346-3553527, gi|16124256:2307427-2308317, gi|16124256:3478372-3479562) with tBLASTx using a cutoff of $1e^{-3}$. Alternate pathways for serine and glycine biosynthesis were identified from literature searches and MetaCyc (Caspi et al., 2008). This resulted in a list of homologs from *Rattus norvegicus*, *Silicibacter pomeroyi* DSS-3, and *Roseobacter denitrificans* Ch 144 that are involved in alternate glycine and serine biosynthetic pathways (NCBI accession numbers gi|56694928:3233121-3234329, gi|110677421:1677707-1678939, gi|89043286:11133-12314, gi|207032|gb|M35270.1|, gi|56694928:1657594-1658841, gi|56972364|gb|BC088133.1, gi|110677421:1869449-1870576, 17227497, gi|49175990:3789378-3790574, gi|110677421:196214-197401, gi|49175990:3788343-3789368, gi|56694928:3558594-3559778, gi|110677421:195174-196205). These were compared to the nucleotide sequence of '*Cand. P. ubique*' with tBLASTx using a cutoff of $1e^{-3}$.

Riboswitch Fold

Identification of riboswitch motifs—The nucleotide sequence of '*Cand. P. ubique*' was split into overlapping, 50 kb segments and put into the RibEx program (Abreu-Goodger and Merino, 2005). The program output identified partial motifs of putative glycine-activated riboswitches upstream of the *glcB* and *gcvT* genes ($1.60e^{-12}$ for *glcB* and $4.00e^{-19}$ for *gcvT*).

Extraction and alignment of metagenomic data—The 300 base pair upstream regions for *glcB* and *gcvT* genes found in a collection of putative SAR11 environmental fragments from the Sargasso Sea binned phylogenetically according to the method of (Wilhelm et al., 2007) were aligned using ClustalW in MEGA with default parameters. The number of environmental fragments searched was over 800,000 and the number binned as belonging to SAR11 was over 100,000. The 50 best alignments for the upstream regions of *glcB* and *gcvT* on fragments binned as SAR11 were retained for modeling the fold of the glycine-activated riboswitches.

Identification of folding pattern—The upstream regions of *glcB* and *gcvT* from '*Cand. P. ubique*' that showed 100% nucleotide level conservation with the aligned metagenomic data were put into the Mfold program (Zuker, 2003) with a constraint of maximum distance for base pairing equal to 100. The resulting folds were examined at the nucleotide level for placement of immutable sites in known riboswitches from *B. subtilis* and *V. cholerae* (Mandal et al., 2004).

Structural Probing to Identify Riboswitch Ligand—The intergenic regions of '*Cand. P. ubique*' corresponding to the glycine riboswitches associated with *gcvT* and *glcB* were PCR amplified with the following respective primer pairs: 5'-acatatctggtagattattcttatatcggg, 5'-attaactactgtacccatctgtatg and 5'-ctattaaaaggtactcttaattgatatcg, 5'-tttaccatctgtattttgcctg. The resulting products were cloned into pCR2.1 using a TOPO TA cloning kit (Invitrogen) and the plasmid insert sequenced. DNA template for *in vitro* transcription was generated by PCR amplification from the sequenced plasmids with forward primers containing a T7 promoter sequence (5'-TAATACGACTCACTATAGGttcttatatcgggagaga, for *gcvT* and 5'-TAATACGACTCACTATAGGtaattgatatcgggagag for *glcB*) and reverse primers indicated above. RNA was generated by *in vitro* transcription, PAGE purified, and 5'-radiolabeled with [γ - 32 P] ATP as described elsewhere (Meyer et al., 2008).

Structural probing assays were conducted by adding RNA (~2 nM) to buffer (50 mM Tris, pH. 8.3, at 23°C, 20 mM MgCl₂, and 100 mM KCl) containing 1 mM of differing candidate

ligands. The reactions were incubated at 25°C for ~40 hours. The samples were separated by denaturing 10% PAGE and imaging and quantification were performed using a Molecular Dynamics PhosphorImager and ImageQuaNT software. Binding and Hill constant values were determined by conducting assays at a series of glycine concentrations. The normalized fraction of RNA cleaved at sites indicated in Figure 3A was calculated for each ligand concentration and the constants were calculated from this data as described previously (Mandal et al., 2004).

Growth Experiments

Natural seawater media collection and preparation procedures are described elsewhere (Connon and Giovannoni, 2002). Briefly, natural media was prepared by 0.2 µm filtering and autoclaving seawater collected from a depth of 10m at the NH5 station (44° 39.1, -124° 10.6) approximately 10 kilometers off the coast of Oregon. Batch cultures were grown in the dark for 30 days at 16° C in 250 ml polycarbonate Erlenmeyer flasks containing 50 ml of growth medium. Cell counts were obtained every five days from an Easy-Cyte flow cytometer (Guava Technologies) after staining with SYBR-Green I (Invitrogen) for one hour.

High performance liquid chromatography (HPLC) measurements of dissolved free amino acid were conducted as previously described (Kaiser and Benner, 2005). HPLC analysis of seawater media collected September, 2007 used for this study contained glycine and serine as well as other dissolved free amino acids (threonine, glutamate, alanine) at nanomolar concentrations following media preparation. We speculate that additional glycine precursors, such as glycine betaine, and choline were also likely present but were not directly assayed in this study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a Marine Microbiology Initiative award from the Gordon and Betty Moore Foundation. M.M.M. is supported by an NIH post-doctoral fellowship (F32GM079974) and the Breaker Lab receives support from the Howard Hughes Medical Institute. We thank Craig Carlson and Stuart Goldberg for the HPLC analysis.

References

- Abreu-Goodger C, Merino E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* 2005; 33:W690–692. [PubMed: 15980564]
- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A.* 2004; 101:6421–6426. [PubMed: 15096624]
- Benner, R. Chemical composition and reactivity. In: Hansell, D.; Carlson, C., editors. *Biogeochemistry of marine dissolved organic matter*. New York: Academic Press; 2002. p. 59-90.
- Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research.* 2008; 36:D623–D631. [PubMed: 17965431]
- Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol.* 2002; 68:3878–3885. [PubMed: 12147485]
- Davis HC, Guillard RR. Relative value of ten genera of microorganisms as food for oyster and clam larvae. *US First Wildlife Serv, Fish Bull.* 1958; 58:293–304.

- Fuhrman JA, Ferguson RL. Nanomolar concentrations and rapid turnover of dissolved free amino-acids in seawater - agreement between chemical and microbiological measurements. *Mar Ecol Prog Ser.* 1986; 33:237–242.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science.* 2005; 309:1242–1245. [PubMed: 16109880]
- Hedges JJ, Baldock JA, Gelinas Y, Lee C, Peterson ML, Wakeham SG. The biochemical and elemental compositions of marine plankton: A NMR perspective. *Mar Chem.* 2002; 78:47–63.
- Kaiser K, Benner R. Hydrolysis-induced racemization of amino acids. *Limnol Oceanogr Methods.* 2005; 3:318–325.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004; 32:D277–280. [PubMed: 14681412]
- Kazanov MD, Vitreschak AG, Gelfand MS. Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics.* 2007; 8:347. [PubMed: 17908319]
- Keller MD, Matrai PA, Kiene RP, Bellows WK. Responses of coastal phytoplankton populations to nitrogen additions: dynamics of cell-associated dimethylsulfoniopropionate (DMSP), glycine betaine (GBT), and homarine. *Can J Fish Aquat Sci.* 2004; 61:685–699.
- Kiene RP. Uptake of choline and its conversion to glycine betaine by bacteria in estuarine waters. *Appl Environ Microbiol.* 1998; 64:1045–1051. [PubMed: 16349511]
- Kiene RP, Williams LPH. Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnol Oceanogr.* 1998; 43:1592–1603.
- Leboulanger C, Descolasgros C, Jupin H. HPLC determination of glycolic acid in seawater - an estimation of phytoplankton photorespiration in the Gulf of Lions, Western Mediterranean Sea. *J Plankton Res.* 1994; 16:897–903.
- Leboulanger C, Oriol L, Jupin H, Descolas-Gros C. Diel variability of glycolate in the eastern tropical Atlantic Ocean. *Deep Sea Res (I Oceanogr Res Pap).* 1997; 44:2131–2139.
- Leboulanger C, Martin-Jezequel V, Descolas-Gros C, Sciandra A, Jupin HJ. Photorespiration in continuous culture of *Dunaliella tertiolecta* (Chlorophyta): Relationships between serine, glycine, and extracellular glycolate. *J Phycol.* 1998; 34:651–654.
- Liu JQ, Dai T, Itoh N, Kataoka M, Shimizu S, Yamada H. Gene cloning, biochemical characterization and physiological role of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. *Eur J Biochem.* 1998; 255:220–226. [PubMed: 9692922]
- Mandal M, Breaker RR. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol.* 2004; 5:451–463. [PubMed: 15173824]
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science.* 2004; 306:275–279. [PubMed: 15472076]
- Mathews, CK.; van Holde, KE.; Ahern, KG. *Biochemistry.* San Francisco: Addison Wesley Longman, Inc; 2000.
- Meyer MM, Roth A, Chervin SM, Garcia GA, Breaker RR. Confirmation of a second natural preQ(1) aptamer class in Streptococcaceae bacteria. *RNA.* 2008; 14:685–695. [PubMed: 18305186]
- Monschau N, Stahmann KP, Sahm H, McNeil JB, Bognar AL. Identification of *Saccharomyces cerevisiae* GLY1 as a threonine aldolase: a key enzyme in glycine biosynthesis. *FEMS Microbiol Lett.* 1997; 150:55–60. [PubMed: 9163906]
- Mopper K, Lindroth P. Diel and depth variations in dissolved free amino acids and ammonium in the Baltic Sea determined by shipboard HPLC analysis. *Limnol Oceanogr.* 1982; 27:336–347.
- Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature.* 2002; 420:806–810. [PubMed: 12490947]
- Pellicer MT, Fernandez C, Badia J, Aguilar J, Lin ECC, Baldoma L. Cross-induction of glc and ace operons of *Escherichia coli* attributable to pathway intersection: characterization of the glc promoter. *J Biol Chem.* 1999; 274:1745–1752. [PubMed: 9880556]
- Pizer LI, Potochny ML. Nutritional and regulatory aspects of serine metabolism in *Escherichia coli*. *J Bacteriol.* 1964; 88:611–619. [PubMed: 14208496]

- Pomeroy LR, Macko SA, Ostrom PH, Dunphy J. The microbial food web in Arctic seawater - concentration of dissolved free amino-acids and bacterial abundance and activity in the Arctic Ocean and in Resolute Passage. *Mar Ecol Prog Ser.* 1990; 61:31–40.
- Ravnikar PD, Somerville RL. Genetic characterization of a highly efficient alternate pathway of serine biosynthesis in *Escherichia coli*. *J Bacteriol.* 1987; 169:2611–2617. [PubMed: 3108237]
- Roulier MA, Palenik B, Morel FMM. A method for the measurement of choline and hydrogen-peroxide in seawater. *Mar Chem.* 1990; 30:409–421.
- Soukup GA, Breaker RR. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA.* 1999; 5:1308–1325. [PubMed: 10573122]
- Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature.* 2008
- Wickiser JK, Winkler WC, Breaker RR, Crothers DM. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell.* 2005; 18:49–60. [PubMed: 15808508]
- Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct.* 2007; 2:27. [PubMed: 17988398]
- Zhao J, Shimizu K. Metabolic flux analysis of *Escherichia coli* K12 grown on ¹³C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J Biotechnol.* 2003; 101:101–117. [PubMed: 12568740]
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003; 31:3406–3415. [PubMed: 12824337]

Figure 1.

Molar growth yield curve for glycine. The maximum cell density achieved for batch cultures grown in the dark at 16° C is plotted against addition of glycine in the presence of excess nutrients (avg., N=2 for each data point, $R^2=0.992$, error bars show complete range). Carbon was supplied as 10 μ M glucose and 50 μ M acetate. Other nutrients were 10 μ M NH_4Cl , 1 μ M KH_2PO_4 , 53.6 μ M FeCl_3 and a mix of vitamins (Davis and Guillard, 1958).

Figure 2.

Secondary structure of SAR11 riboswitches and alignment to environmental data. Aptamers are labeled with Roman Numerals I and II, and base pairing stems are labeled P1, P2, and P3 with subsections labeled “a” and “b.” Base pairs with an asterisk were shown to be accessible to spontaneous cleavage, as would be expected if they are not part of a base pair. Base pairs shaded in grey show 100% conservation in upstream regions of metagenomic data. *gcvT*, glycine cleavage protein T; *glcB*, malate synthase.

Figure 3. Structural probing and K_D measurements for *gcvT* and *glcB* riboswitches. A. Spontaneous cleavage products of the *glcB* associated riboswitch in the presence of 1 mM of ligand candidates and the absence of ligand. Selected RNase T1 cleavage products (cleaves 3' of G residues) are identified on the left. Arrows on the right indicate bands used to quantify extent of RNA cleavage. B. Plot of percentage RNA cleaved (normalized between no ligand and saturating ligand conditions) versus the logarithm of the molar concentration of glycine used to determine K_D and h constants for the *glcB* associated riboswitch. C. Identical to B for the *gcvT* associated riboswitch.

Fig 4. Maximum cell yields for glycine substitutes (avg., N=2 for each data point, error bars show complete range). Response to 1.5 μ M additions of compounds. Gly, glycine; Bet, betaine; Ser, serine; Glc, glycolate; Glx, glyoxylate; Thr, threonine; No Gly, no glycine; No Glu, no glucose; No ace, no acetate. Excess nutrients were supplied as in Fig. 1.

Table 1

Glycine/serine biosynthesis genes. The gene names and corresponding enzyme names appear in the first two columns; yeast genes are shown in all capital letters. The gene loci for genes present in '*Cand. P. ubique*' appear in the right column. EC = Enzyme Commission.

Glucose/3-Phosphoglycerate Precursor		EC Number	Gene Locus
<i>serA</i>	D-3-phosphoglycerate dehydrogenase	1.1.1.95	SAR11_1366
<i>serB</i>	phosphoserine phosphatase	3.1.3.3	missing
<i>serC</i>	3-phosphoserine aminotransferase	2.6.1.52	missing
Glyoxylate Precursor			
<i>agxt/spt</i>	alanine/serine:glyoxylate aminotransferase	2.6.1.44	missing
Threonine Precursor			
<i>tdh</i>	threonine dehydrogenase	1.1.1.103	missing
<i>kbl</i>	2-amino-3-ketobutyrate CoA ligase	2.3.1.29	missing
<i>ItaE/GLY1</i>	threonine aldolase	4.1.2.5	SAR11_0689
Serine/Glycine Interconversion			
<i>glyA/SHMT</i>	glycine hydroxymethyltransferase	2.1.2.1	SAR11_1048