Introducción al Aprendizaje Automático

Intuición, Algoritmos, y Mejores Prácticas

David Omar Flores Chávez[1] and
Héctor Aguilar Hernández[1]

[1]Escuela Superior de Cómputo

17 de octubre de 2017

Resumen

Extracellular enzymes are a key intermediate in the global carbon cycle, responsible for the initial step in the microbial transformation of organic matter to CO_2. Protein is a particularly important component of organic matter, becuase it contains both carbon and nitrogen, and because its component parts (amino acids) are used by all organisms. More than 300 peptidases (protein-degrading enzymes) are known, but despite the fact that many investigators have extensively studied the ecology and geochemistry of peptidases (protein-degrading enzymes) in aquatic ecosystems, all previous research in freshwater systems are limited to the analysis of a single peptidase, leucyl aminopeptidase. The purpose of this field trip was to continue a field program begun in 2013 to assess the activities of five distinct peptidases in diverse freshwater environments near the Poconos Environmental Education Center in the Delaware Water Gap National Recreation Area in eastern Pennsylvania. We show that at least four distinct peptidases are active in these waters: arginyl aminopeptidase, glycyl aminopeptidase, leucyl aminopeptidase, and trypsin. Enzymes hydrolyzing proteins containing terminal pyroglutamic acid were not observed. Patterns of peptidase activity varied among water bodies, and the ecological controls on that variability were not apparent. These data form a subset of a 3-year, multisite peptidase activity dataset which we are in the process of preparing for publication.

1. Introduction

Land plants 'fix' CO_2 into organic carbon. Most of this is rapidly reoxidized to CO_2, but approximately 0.2 % is mobilized from soils and transported by rivers to the ocean, where it may be sequestered over geological timescales Galy2015. The transport of terrestrial organic carbon to the ocean therefore represents a key term in the global carbon cycle, so it is critical to understand the factors that control how much terrestrial organic carbon reaches the ocean. A major sink is biological activity in rivers Cole2007, which emit $7{,}5 \times 10^{15}$ grams of CO_2 to the atmosphere per year.

Much of this CO_2 production is mediated by microorganisms, which produce extracellular enzymes in order to break down complex organic molecules outside of the cell Arnosti2013. The set of extracellular enzymes present in an environment therefore constrains the total amount of organic carbon the community can convert to CO_2. A full, mechanistic understanding of the global carbon cycle therefore requires a good understanding of the controls on the presence and activity of extracellular enzymes produced by microbial ecosystems.

For that reason, the environmental controls on microbial extracellular enzyme production have been intensively investigated Arnosti2011,Sinsabaugh2009,Arnosti2013. Nearly all of these studies, however, rely on a small class of compounds ('fluorogenic substrate proxies') which are capable of reporting the activities of only a small handful of distinct enzymes German2011. Notably, we are unaware of any studies of freshwater in which activities of peptidases (protein-degrading enzymes) were assayed using any

Figura 1: Map of sampling sites.

Figura 2: All raw data from the 2015 PEEC trip.

compound other than leucine 7-amido-4-methylcoumarin (or related leucine-based compounds), which reports the activity of an enzyme called leucyl aminopeptidase (E.C. 3.4.11.1). We know a great deal about the dynamics and controls of this enzyme - but there are more than 300 formally-recognized peptidases, any of which could be active and geochemically important in freshwaters.

Work by Suzuki and colleagues Obayashi2005, Obayashi2008a, Obayashi2009 has shown that in coastal Japanese seawater, endopeptidases (peptidases that cut proteins from the middle) are consistently more active than aminopeptidases (peptidases that cut proteins from one end). Our group has also observed this pattern in deep marine sediments from the Baltic Sea Lloyd2013.This pattern, however, is far from universal: in an Arctic fjord Steen2013 and the Amazon River (Steen et al. in prep), aminopeptidases are more active than endopeptidases. It seems, therefore that different sets of peptidases are present in different environments. The goals of this field trip were to address the following questions about freshwater peptidases:

- What is the range of peptidases present in ecologically diverse freshwaters of the Pocono mountains?
- Do patterns of peptidase activities vary from place to place?
- Are patterns of peptidases predicted by ecological factors such as the type of water body?

2. Methods

2.1. Sites and sample collection

Samples were collected from ecologically diverse sites in the vicinity of Pocono Environmental Education Center (Fig 1). Sample locations and brief descriptions are given in Table 1. Samples were collected from near shore in clean, 1-L polyethylene bottles. Sample temperature was measured in bulk water at the time of collection. Sample pH was measured onsite using pH paper, and again at the University of Tennessee using an electronic pH meter.

2.2. Enzyme assays

2.3. Data analysis

Enzyme data were analyzed using the R statistical package RCoreTeam2013 and the `enzalyze` package
(http://github.com/adsteen/enzalyze).

3. Results

3.1. Data quality control

Prior to analysis, three obvious outlier data points were removed, out of 1140 total data points collected. Plots of raw data showed that, in general, fluorescence increased linearly as a function of time in the live samples, and was constant in the killed controls (Fig. 2). Note that, since the gain of the three fluorimeters varied significantly, rates of fluorescence activity should not be compared among sites.

3.2. Levels of peptidase activity

Activities were highest in small, algae-covered ponds and lower in clearer ponds/lakes, streams, and the Delaware River (Fig 3). There was not a terribly obvious ecological pattern to the summed activities,

Figura 3: Summed activities of peptidases at each site.

Figura 4: Activities of individual peptidases at each site.

Figura 5: Hydrolysis rate of Leu-AMC vs Arg-AMC (left) and Leu-AMC vs GlyGlyArg-AMC

however: Front Pond and Loch Lomond, for instance, are both ponds of several hundred meters in diameter free from surface macroalgae, although they were on opposite ends of the spectrum of summed peptidase activities.

3.3. Patterns of peptidase activity

Some patterns in the activities of individual enzymes were evident (Fig 4). Hydrolysis rates of Gly-AMC and Pyr-AMC were always very close to zero, whereas hydrolysis rates of Arg-AMC, Leu-AMC, and GlyGlyArg-AMC were generally much faster. Leu-AMC was often the fastest-hydrolyzed substrate, but in some cases GlyGlyArg-AMC was much more active.

There was a relatively tight relationship between hydrolysis rate of arginyl aminopeptidase and leucyl aminopeptidase ($p = 0.00081$) and between glycyl aminopeptidase and leucyl aminopeptidase ($p = 0.0017$), whereas there was no significant relationship between trypsin and leucyl aminopeptidase ($p = 0.84$, Fig 5). The close correspondence between leucyl aminopeptidase and arginyl aminopeptidase, but not trypsin, is also evident from principal components analysis of the full dataset (with pyroglutamyl aminopeptidase removed, because those rates were generally not distinguishable from zero).

4. Discussion

4.1. Levels of peptidase activity

There was no clear factor driving summed peptidase activities. However, the lack of metadata makes it difficult to evaluate this further: for instance, knowing cell concentrations would be useful, as cell-specific enzyme activities might be more constant. It will be possible to collect such data in 2016.

4.2. Patterns of peptidase activity

Patterns of enzyme activities likewise varied. Because leucyl aminopeptidase, glycyl aminopeptidase, and arginyl aminopeptidase were all fairly tightly inter-correlated, the main axis of variation of enzyme activities was the ratio of endopeptidase activity (trypsin, which cleaves proteins in the middle) to aminopeptidase activities (the aminopeptidases, which remove aminopeptidases from the N-termini of proteins). This difference seems to point to a difference in microbial communities' strategies for obtaining protein: the communities at sites TW, PP and DR (i.e., those in the bottom half of the Fig 6) primarily access protein by degrading it from the middle, whereas communities at sites AC, LL, and BS primarily access protein from the ends. We note that we did not assay for carboxypeptidases (enzymes that cleave single amino acids from the C-terminus of proteins) because reliable fluorogenic substrates for those peptidases are not available, and that other peptidases (e.g. other endopeptidases, dipeptidyl aminopeptidases, etc.) may be present but poorly assayed by the substrates we have used. Thus, other dimensions of variability in protein acquisition may exist. The results presented here show - for the first time in any aquatic system - that substantial variability exists in the pathways by which microbial communities access protein.

These results do not point to an obvious factor driving protein acquisition strategies among sites: for instance, sites AC and TW are both small streams, yet they are on opposite ends of the endopeptidase/aminopeptidase spectrum. Likewise, sites PP and LL are both midsize (ca. 100-400 m major axis) ponds that are largely free from visible macroalgae, but have opposite protein degradation strategies.

We may hypothesize that protein acquisition strategies are driven by some combination of environmental and biological factors, possibly including nutrient status, dissolved orgnaic matter concentration

Figura 6: Principal components biplot of enzyme activities (Pyroglutamyl acid removed due because activities were generally below the detection limit).

or composition, microbial community composition (including microscopic eukaryotes), or other factors. Untangling those factors will require careful characterization of the chemical and biological environment of each site.

A substantial body of literature clearly demonstrates differences in ecosystem function at all spatial scales, at the level of 'gross' ecosystem function: for instance, rates of carbon preservation or CO_2 production, N transformation, etc. This work shows that functional variation may exist at very fine scale of individual enzyme production as well.

5. Project overview

5.1. Previous work

The 2015 trip represents the third year of the University of Tennessee / Malcolm X Shabazz High School collaboration to investigate enzyme activities around PEEC. In the first year, we collected saturation curves (measurements of enzyme activities at varying substrate concentrations) which told us what the correct substrate concentration to use in future experiments was, and which demonstrated clearly that the increase in fluorescence we measure in our samples is due to enzyme activities rather than to some artifact. The 2014 dataset looks a lot like the 2015 dataset, although we have not analyzed it as carefully at this point.

An undergraduate student, Lauren Mullen, has collected a parallel dataset of about 15 sites, using the same substrates and protocols as described in this report, in waters around the University of Tennessee and in the Smokies Mountains. Those data will be analyzed by the end of June 2015, and collated with the data presented here.

5.2. Future work

5.2.1. Coastal samples

During the end of the 2015 school year, Patrick Murray is expecting to add to our dataset by meausring enzyme activities in samples of coastal and estuarine water bodies around Newark, NJ. These samples will considerably expand the range of ecosystems we have measured, and will provide us with a true headwaters-to-ocean dataset.

5.2.2. Temporal stability

Over three years, we have sampled a wide range of environments, but none more than three times. **It will be important to determine how much temporal variability exists in protein degradation pathways**. We will use two strategies to address this. First, we can compare several sites that have been sampled in all three years of the PEEC experiments. Second, we can select several sites in Tennessee to repeatedly sample (say 1-2 times per week) for several weeks. We have reason to believe that the patterns will be broadly stable: the Suzuki lab, which has been using a range of peptidase substrates since 2005 but which has restricted their investigation to a single location in southern coastal Japan, has found that endopeptidases are consistently more active than exopeptidases Obayashi2005, Obayashi2008a, Obayashi2009.

5.2.3. Data integration and publication

Between the three years of PEEC experiments, the soon-to-be-collected data from coastal New Jersey, and the more recent, parallel work in east Tennessee, we have collected a substantial dataset about relative activities of the five enzymes listed in Table 1. It will take some time to confirm that the data have all been analyzed in the same way and to integrate the various datasets and synthesized the results. We expect to perform that data integration and synthesis, and to submit a manuscript (probably to the open-access journal Frontiers in Aquatic Microbiology) by the fall of 2015.

site	name	coordinates	temp., ^oC	date-time	description
FP	Front Pond	41.170668 N, 74.91492 W	25	5/27/15 12:10	clearwater pond
AC	Alicia's Creek	41.16923333 N, 74.91406667	20	5/27/15 13:58	first-order stream
PP	Pickerel Pond	41.167504 N, 74.917862 W	27	5/27/15 14:06	clearwater pond
BS	Briscoe Swamp	41.16796667 N, 74.9211 W	21	5/28/15 08:55	macroalgae-rich shallow swamp
TW	Tumbling Waters Creek	41.15451667 N, 74.91775 W	18	5/28/15 09:16	fast-flowing second-order stream
LL	Loch Lomond	41.20705 N, 74.89605 W	23	5/28/15 09:40	clearwater pond
DR	Delaware River	41.13724 N, 74.926029 W	25	5/28/15 12:35	major river
BB1	Basketball Court Pond 1[1]	pending	23	5/28/15 13:53	small, macroalgae-rich pond
BB2	Basketball Court Pond 2	pending	23	5/28/15 13:58	semi-connected shallow bay adjacent to BB1
VP	Vernal Pool	41.16935 N, 74.91406667 W	23	5/28/15 13:38	Vernal pool on Fossil Trail

Cuadro 1: Sites sampled in 2015.

Substrate	Abbrev.	Enzyme
Arginine-AMC	Arg-AMC	arginyl aminopeptidase
Glycine-AMC	Gly-AMC	glycyl aminopeptidase
Leucine AMC	Leu-AMC	leucyl aminopeptidase
Pyroglutamic acid-AMC	Pyr-AMC	pyroglutamyl aminopeptidase[2]
Glycine-Glycine-Argnine-AMC	GlyGlyArg-AMC	trypsin

Cuadro 2: List of substrates used in this experiment and the enzymes they assay.

Referencias