Research Group Meeting, Week 2.0

Reading and Searching Scientific Literature, Skill Development 1.0: Solutions, Dilutions, Buffers, pH

Inspiration

If we all worked on the assumption that what is accepted as true
were really true,
there would be little hope of advance.

Orville Wright

Anyone who has never made a mistake
has never tried anything new.

Albert Einstein

Every artist was first an amateur.

Ralph Waldo Emerson

In other words: It is OK to make mistakes! It is expected. I expect you all NOT to be professionals at this point! An "A" is not received in this class based on your lab being successful, but did you learn from it? I can tell by the effort in the Skill Developments and your engagement in the lab.

Administrative

- It is important that we get to know each other. (3 minutes)
- Availability between 11-noon on Mondays.
- I have three time slots for "sub-group meetings": Sign up, maximum 12 per time slot. Meet once/3 weeks. Counts for 1 lab hour.

Group Delta: Tuesdays 1-2pm, Jan 31, Feb21, March20, April 10 (Kevin, Kenera, Whitney, Mignote, Monica, Alexander, Reagan, Jesus, Sarah, Hollan, Justyn)
Group Echo: Tuesdays 2-3pm, Feb7, Feb28, March27,April17 (Divya, Varsha, Steven, Kimberly, Diana, Grace, Arjun, Johnson)
Group Foxtrot: Wednesdays 11-noon, Feb8, March7, April 4, April 18 (Morgan, David, Huy, Duy, Ivy, Rachel, Uyen, Anjali, Hwang, Kathryn, Jonathan)

Chalkboard Basics

- Lab Notebook: look at Jen La's
- Some definitions
- Micropipettors
- Volume measurements
- Solution make up
- pH and Buffers

In class, I reviewed the following on the chalkboard: Molarity: M vs. mM; Stock silutions and dilutions. Single component vs. Multiple componet solutions, Derived the Henderson-Hasselbach equation in my review of buffers (next slide) and drew a graph to illustrate buffering capacity. Highlighted that buffers are important to living organisms to maintain stable pH even though the solution may have a number of reactions occurring (e.g. metabolism), components in it etc. (blood, for example is bufferedat ph of 7.4)

Chalkboard basics

- **Brief pH/buffer summary, highlighting important equations.**

$$HA + H_2O \xleftrightarrow{K_a} H_3O^+ + A^-$$

Weak acid dissociates to a strong conjugate base by Equilibrium constant Ka

$$K_a = \frac{[H_3O^+][A^-]}{[HA]}$$

Solve for Ka, which equals moles of products over moles of reactants

$$-\log[H_3O^+] = -\log K_a - \log\frac{[A^-]}{[HA]}$$

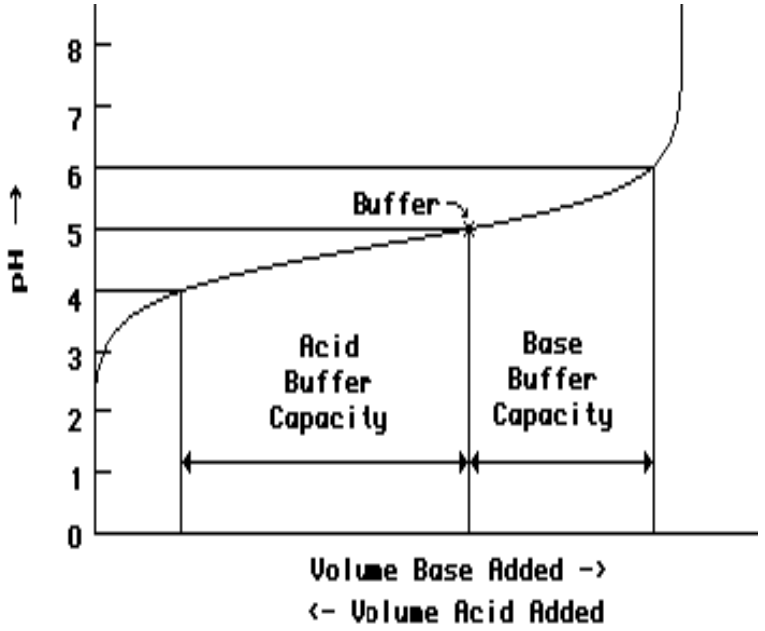
Rearrange and take the –log of both sides

$$pH = pK_a + \log\frac{[HA]}{[A^-]}$$

Recognize that the log of [H+] is pH and rearrange, you have the Henderson-Hasselbalch equation. pKa is characteristic of acids (e.g acetic acid pKa is 4.74).

Chalkboard basics

- Brief pH/buffer summary, highlighting important equations. Cont'd.



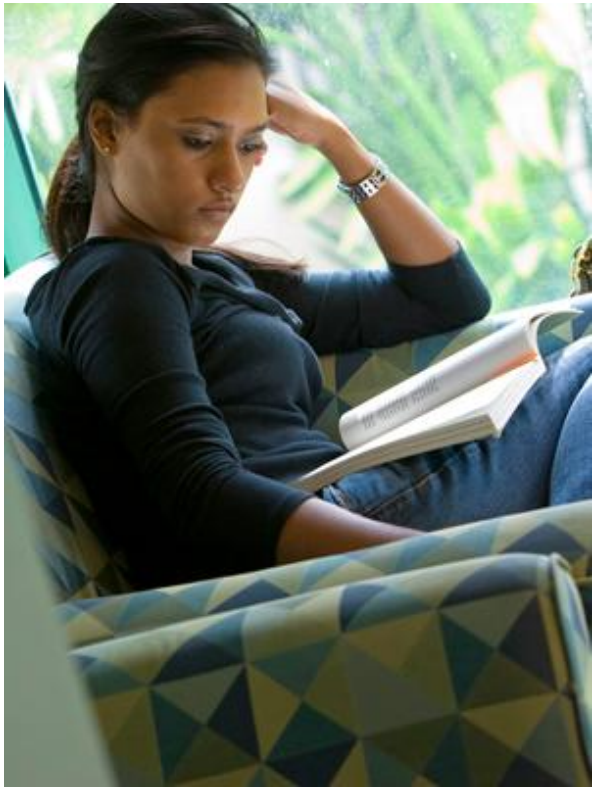
At [A-] = [HA] the pH = pK_a(because log [A-] / [HA] = log 1 = 0 in the previous equation). This solution will be equally effective as a buffer towards either H_3O^+ or OH^- because equal amounts of acid and conjugate base are present.

Journal Literature
The textbook of science

- Never too early to start thinking about your proposal
- Start with literature searches
- Looking for an anti-bacterial? Start with a LOCAL plant that has anti-bacterial properties

We did not get to go over this in class, I will review briefly next time!

The Problem



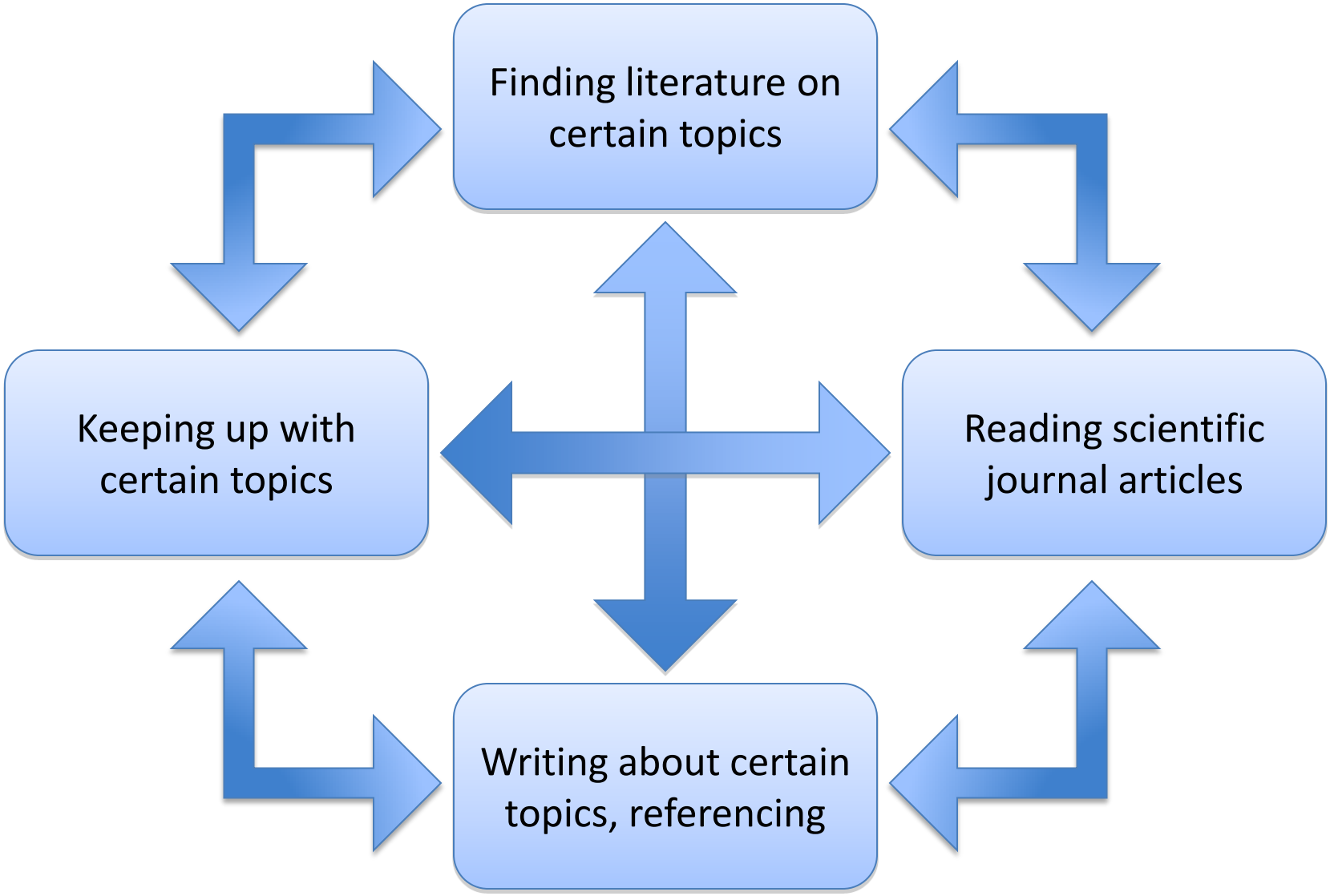
- Students have been taught to read cover to cover, or entire articles/chapters
- Scientific literature is dry and boring
- Scientific vocabulary is complicated
- Students do not know how to find relevant literature for their questions
- Students have no time to read and less to think or write about it

Must Overcome the Problem Because

Scientific Literature …

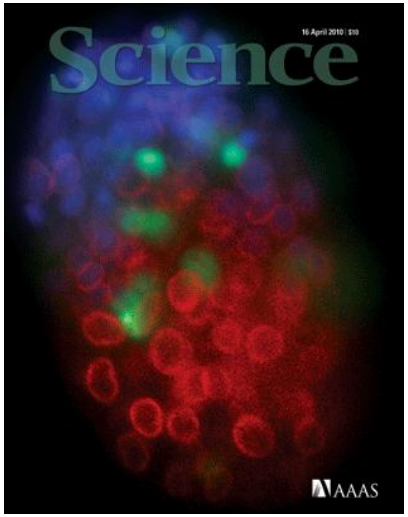
- … sparks creativity
- … is critical to development as a scientist
- … provides insight into own research
- … helps to fully appreciate a topic
- … informs what colleagues have been up to
- … allows scientists to get funded

Break it Down to Four Main Tasks



The Innovative Temporal Encyclopedia

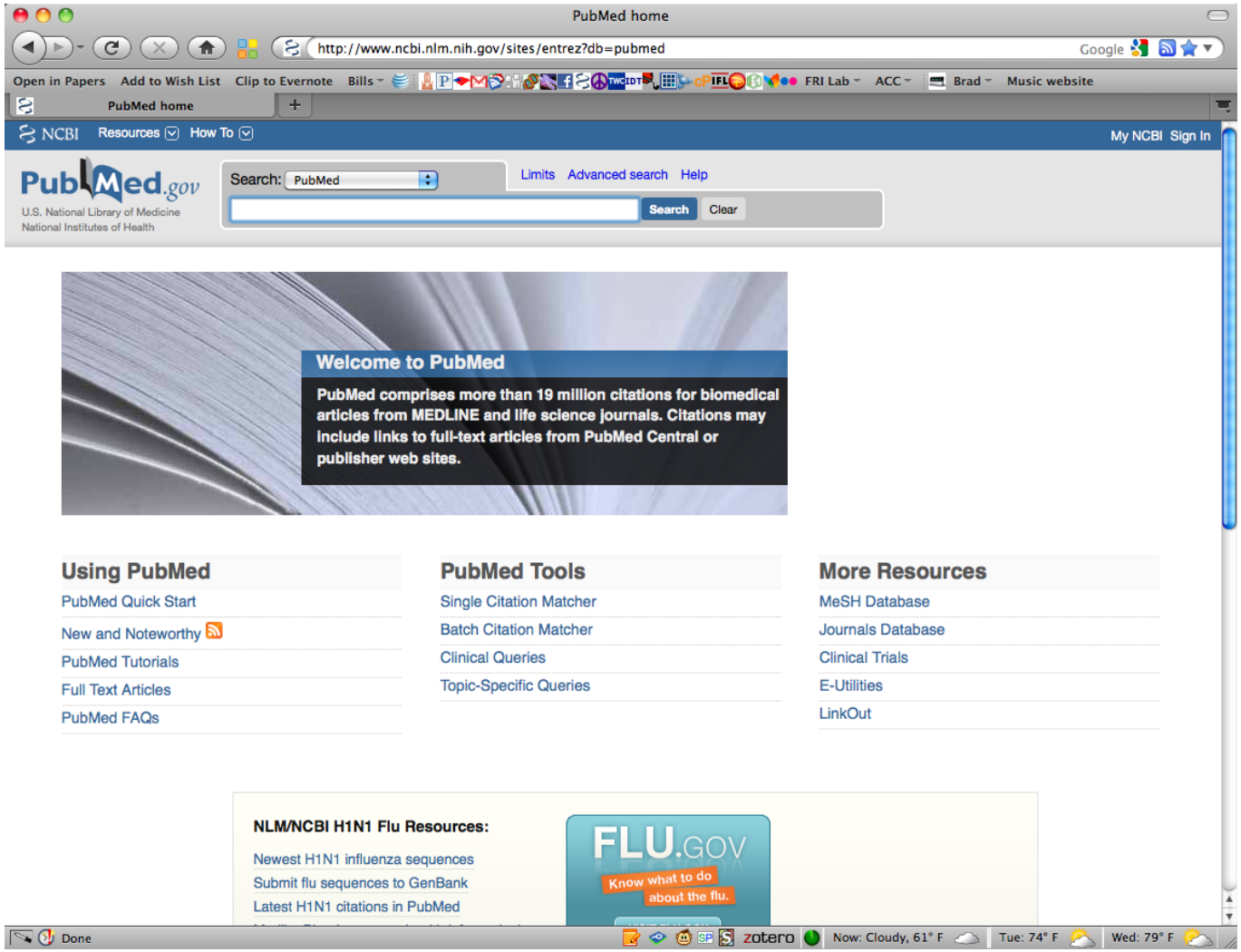
- **Scientific Literature is like a constantly updated history encyclopedia.**
  - Don't read through spatially
  - Always re-read for depth
- **Temporal relevance**
  - Research not done in vacuum
  - Many groups contribute
  - What happened, What's next?
- **State of the field**
  - Find main researchers
  - Know where the field is by putting results from everyone together.



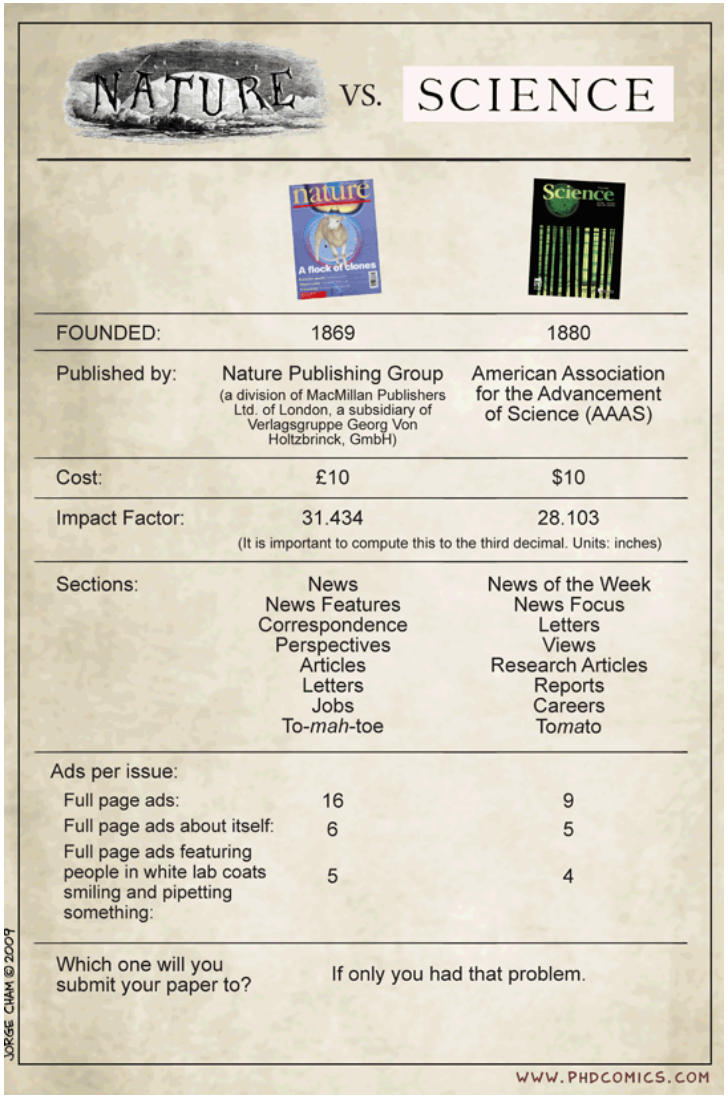
Where To Start?

1. **Skim Wikipedia (Grand overview)**
2. **Search Online (Current/Past)**
   - Foundations/Youtube/Pubmed/WOS/Google Scholar
   - Follow linking structures
3. **Read Review Article (specialized overview)**
4. **Read Primary Literature**
   - Find References from Intro
5. **Look at publishers websites for collections**
   - High impact journals
6. **Flip through hard copy journals – go to the library**
7. **Colleagues**
   - Journal Clubs, Professional Societies/Meetings, On campus lectures and seminars
8. **Books (expensive but comprehensive)**
   - Textbooks or focused topic books

Pubmed

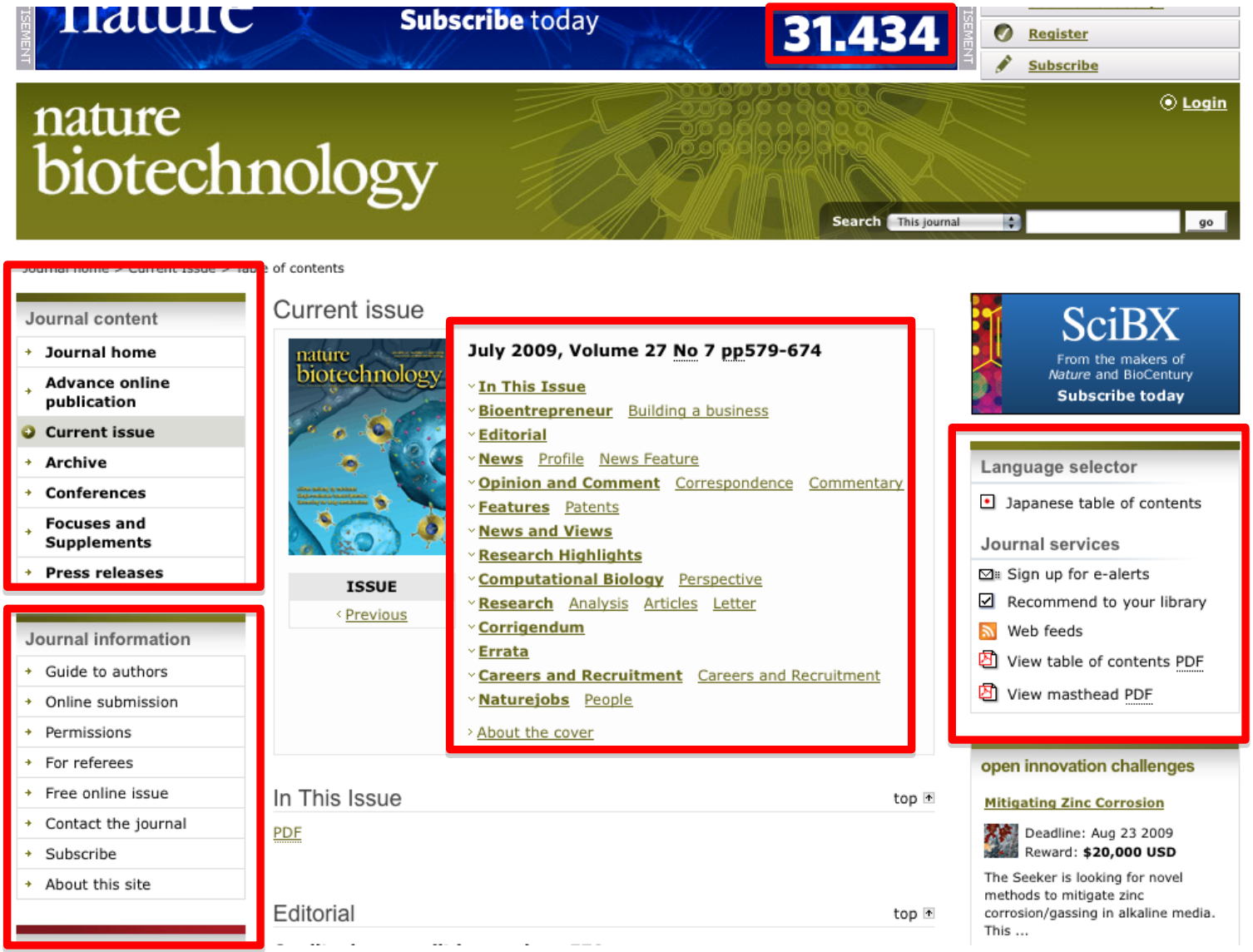

PubMed home
http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed
Open in Papers
Add to Wish List
Clip to Evernote
Bills
FRI Lab
ACC
Brad
Music website
NCBI
Resources
How To
My NCBI
Sign In
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed
Limits
Advanced search
Help
Search
Clear
Welcome to PubMed
PubMed comprises more than 19 million citations for biomedical articles from MEDLINE and life science journals. Citations may include links to full-text articles from PubMed Central or publisher web sites.
Using PubMed
PubMed Quick Start
New and Noteworthy
PubMed Tutorials
Full Text Articles
PubMed FAQs
PubMed Tools
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Topic-Specific Queries
More Resources
MeSH Database
Journals Database
Clinical Trials
E-Utilities
LinkOut
NLM/NCBI H1N1 Flu Resources:
Newest H1N1 influenza sequences
Submit flu sequences to GenBank
Latest H1N1 citations in PubMed
FLU.GOV
Know what to do about the flu.
Done
zotero
Now: Cloudy, 61° F
Tue: 74° F
Wed: 79° F

Structure of an Actual Journal

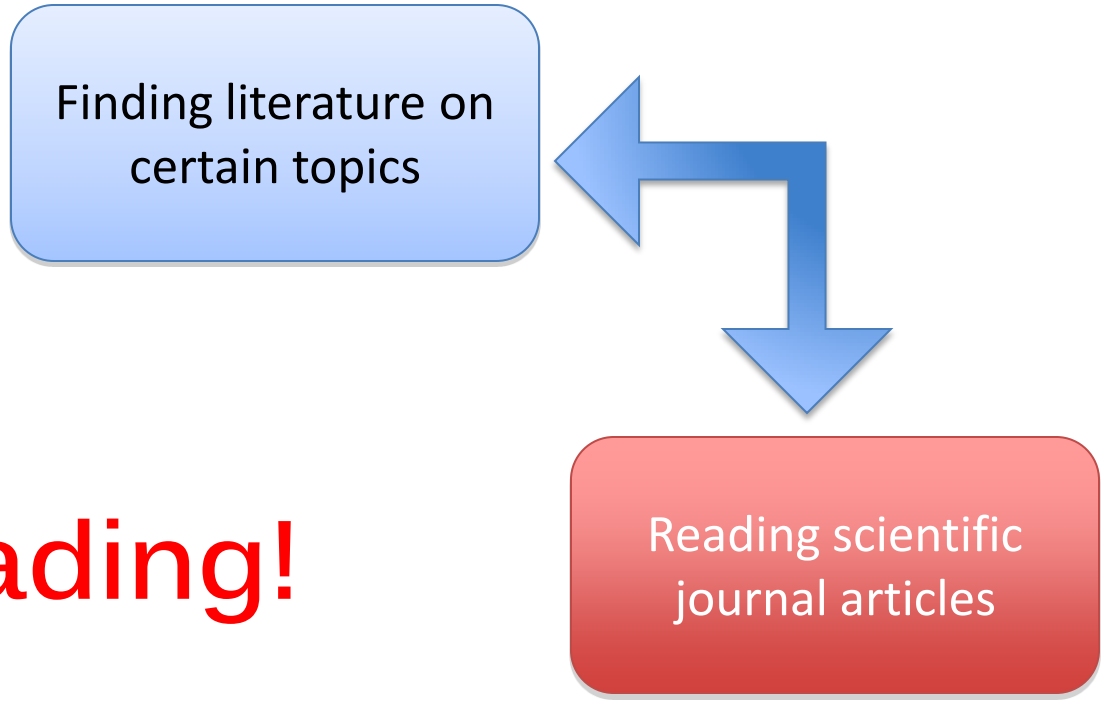


- **Recent Advances**
  - News
- **Opinions**
  - Commentary
  - Views
  - Perspectives
- **Short Research**
  - Letters
  - Communications
  - Reports
- **Long Research**
  - Research
  - Articles

Journal Website



Once You Have an Article…



Start reading!

Do Not Read it Like a Book

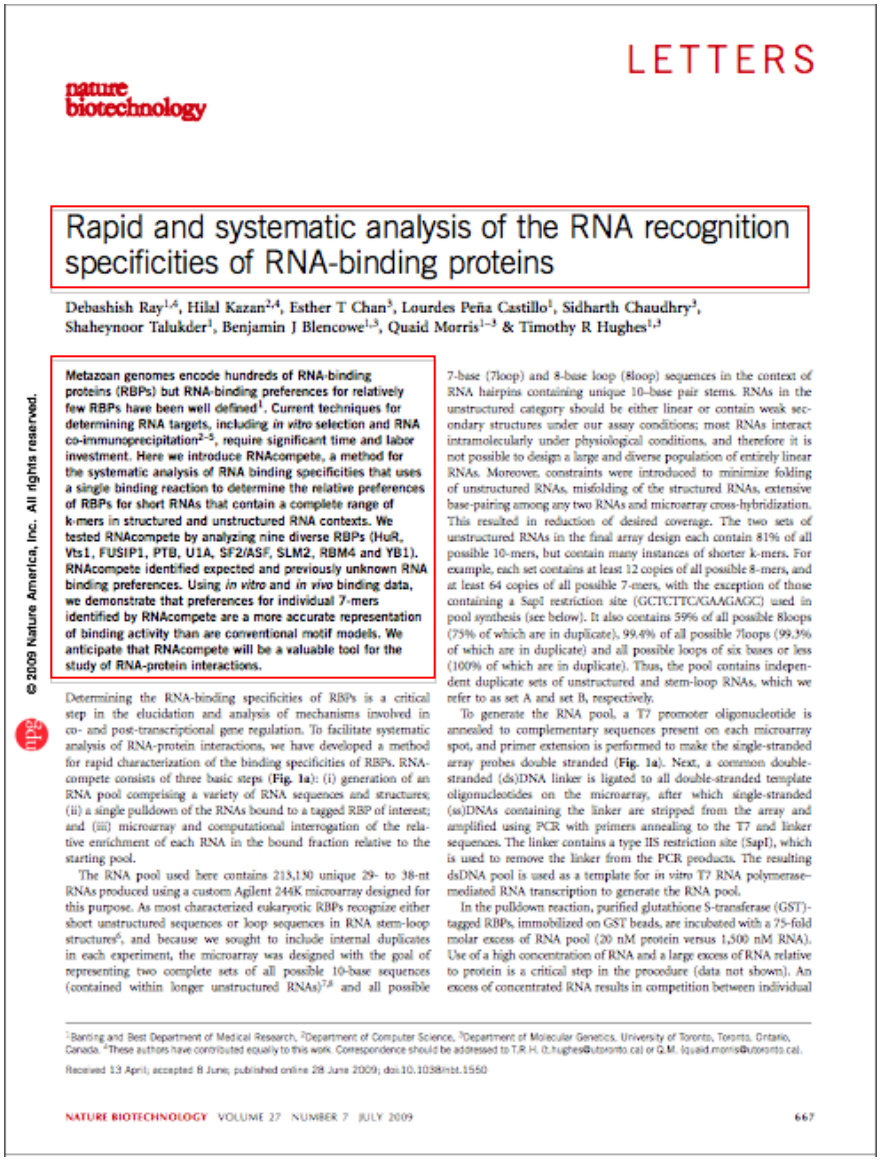
- **Spatially Parse the information**
  - Title
  - Abstract
  - Introduction
  - Methods
  - Figures
  - Results/Discussion
  - Conclusion
  - References
  - Supplementary information

Type	Read
Title	100
Abstract	50
Figures	20
Conclusion	10
Full Paper	5

- **Organize the important stuff**
  - Writing a figure/results outline with main result
  - Writing a subheading outline

Learn how to Read, Again

- Title
- Abstract
- Last paragraph of Introduction
  - take home message
- Figures
  - graphical story
- Conclusions
- Introduction
  - again if needed
- Results and Discussion
- Methods



LETTERS

nature biotechnology

Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins

Debashish Ray[1,4], Hilal Kazan[2,4], Esther T Chan[3], Lourdes Peña Castillo[1], Sidharth Chaudhry[3], Shaheynoor Talukder[1], Benjamin J Blencowe[1,3], Quaid Morris[1–3] & Timothy R Hughes[1,3]

Metazoan genomes encode hundreds of RNA-binding proteins (RBPs) but RNA-binding preferences for relatively few RBPs have been well defined[1]. Current techniques for determining RNA targets, including *in vitro* selection and RNA co-immunoprecipitation[2–5], require significant time and labor investment. Here we introduce RNAcompete, a method for the systematic analysis of RNA binding specificities that uses a single binding reaction to determine the relative preferences of RBPs for short RNAs that contain a complete range of k-mers in structured and unstructured RNA contexts. We tested RNAcompete by analyzing nine diverse RBPs (HuR, Vts1, FUSIP1, PTB, U1A, SF2/ASF, SLM2, RBM4 and YB1). RNAcompete identified expected and previously unknown RNA binding preferences. Using *in vitro* and *in vivo* binding data, we demonstrate that preferences for individual 7-mers identified by RNAcompete are a more accurate representation of binding activity than are conventional motif models. We anticipate that RNAcompete will be a valuable tool for the study of RNA-protein interactions.

Determining the RNA-binding specificities of RBPs is a critical step in the elucidation and analysis of mechanisms involved in co- and post-transcriptional gene regulation. To facilitate systematic analysis of RNA-protein interactions, we have developed a method for rapid characterization of the binding specificities of RBPs. RNAcompete consists of three basic steps (**Fig. 1a**): (i) generation of an RNA pool comprising a variety of RNA sequences and structures; (ii) a single pulldown of the RNAs bound to a tagged RBP of interest; and (iii) microarray and computational interrogation of the relative enrichment of each RNA in the bound fraction relative to the starting pool.

The RNA pool used here contains 213,130 unique 29- to 38-nt RNAs produced using a custom Agilent 244K microarray designed for this purpose. As most characterized eukaryotic RBPs recognize either short unstructured sequences or loop sequences in RNA stem-loop structures[6], and because we sought to include internal duplicates in each experiment, the microarray was designed with the goal of representing two complete sets of all possible 10-base sequences (contained within longer unstructured RNAs)[7,8] and all possible 7-base (7loop) and 8-base loop (8loop) sequences in the context of RNA hairpins containing unique 10–base pair stems. RNAs in the unstructured category should be either linear or contain weak secondary structures under our assay conditions; most RNAs interact intramolecularly under physiological conditions, and therefore it is not possible to design a large and diverse population of entirely linear RNAs. Moreover, constraints were introduced to minimize folding of unstructured RNAs, misfolding of the structured RNAs, extensive base-pairing among any two RNAs and microarray cross-hybridization. This resulted in reduction of desired coverage. The two sets of unstructured RNAs in the final array design each contain 81% of all possible 10-mers, but contain many instances of shorter k-mers. For example, each set contains at least 12 copies of all possible 8-mers, and at least 64 copies of all possible 7-mers, with the exception of those containing a SapI restriction site (GCTCTTC/GAAGAGC) used in pool synthesis (see below). It also contains 59% of all possible 8loops (75% of which are in duplicate), 99.4% of all possible 7loops (99.3% of which are in duplicate) and all possible loops of six bases or less (100% of which are in duplicate). Thus, the pool contains independent duplicate sets of unstructured and stem-loop RNAs, which we refer to as set A and set B, respectively.

To generate the RNA pool, a T7 promoter oligonucleotide is annealed to complementary sequences present on each microarray spot, and primer extension is performed to make the single-stranded array probes double stranded (**Fig. 1a**). Next, a common double-stranded (ds)DNA linker is ligated to all double-stranded template oligonucleotides on the microarray, after which single-stranded (ss)DNAs containing the linker are stripped from the array and amplified using PCR with primers annealing to the T7 and linker sequences. The linker contains a type IIS restriction site (SapI), which is used to remove the linker from the PCR products. The resulting dsDNA pool is used as a template for *in vitro* T7 RNA polymerase–mediated RNA transcription to generate the RNA pool.

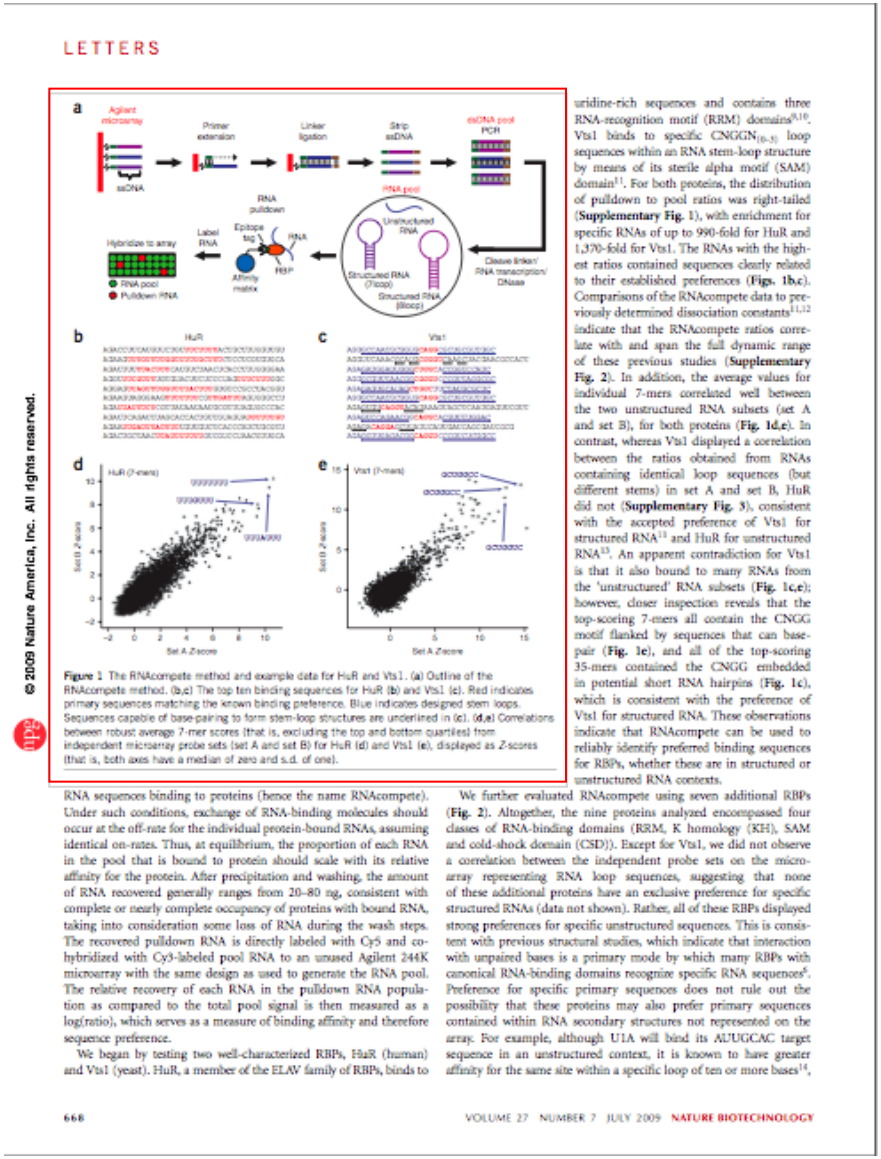
In the pulldown reaction, purified glutathione S-transferase (GST)-tagged RBPs, immobilized on GST beads, are incubated with a 75-fold molar excess of RNA pool (20 nM protein versus 1,500 nM RNA). Use of a high concentration of RNA and a large excess of RNA relative to protein is a critical step in the procedure (data not shown). An excess of concentrated RNA results in competition between individual

[1]Banting and Best Department of Medical Research, [2]Department of Computer Science, [3]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [4]These authors have contributed equally to this work. Correspondence should be addressed to T.R.H. (t.hughes@utoronto.ca) or Q.M. (quaid.morris@utoronto.ca).

Received 13 April; accepted 8 June; published online 28 June 2009; doi:10.1038/nbt.1550

Learn how to Read, Again

- Title
- Abstract
- Last paragraph of Introduction
  - take home message
- Figures
  - graphical story
- Conclusions
- Introduction
  - again if needed
- Results and Discussion
- Methods



LETTERS

Figure 1 The RNAcompete method and example data for HuR and Vts1. (a) Outline of the RNAcompete method. (b,c) The top ten binding sequences for HuR (b) and Vts1 (c). Red indicates primary sequences matching the known binding preference. Blue indicates designed stem loops. Sequences capable of base-pairing to form stem-loop structures are underlined in (c). (d,e) Correlations between robust average 7-mer scores (that is, excluding the top and bottom quartiles) from independent microarray probe sets (set A and set B) for HuR (d) and Vts1 (e), displayed as *Z*-scores (that is, both axes have a median of zero and s.d. of one).

RNA sequences binding to proteins (hence the name RNAcompete). Under such conditions, exchange of RNA-binding molecules should occur at the off-rate for the individual protein-bound RNAs, assuming identical on-rates. Thus, at equilibrium, the proportion of each RNA in the pool that is bound to protein should scale with its relative affinity for the protein. After precipitation and washing, the amount of RNA recovered generally ranges from 20–80 ng, consistent with complete or nearly complete occupancy of proteins with bound RNA, taking into consideration some loss of RNA during the wash steps. The recovered pulldown RNA is directly labeled with Cy5 and co-hybridized with Cy3-labeled pool RNA to an unused Agilent 244K microarray with the same design as used to generate the RNA pool. The relative recovery of each RNA in the pulldown RNA population as compared to the total pool signal is then measured as a log(ratio), which serves as a measure of binding affinity and therefore sequence preference.

We began by testing two well-characterized RBPs, HuR (human) and Vts1 (yeast). HuR, a member of the ELAV family of RBPs, binds to uridine-rich sequences and contains three RNA-recognition motif (RRM) domains[9,10]. Vts1 binds to specific CNGGN$_{(0-5)}$ loop sequences within an RNA stem-loop structure by means of its sterile alpha motif (SAM) domain[11]. For both proteins, the distribution of pulldown to pool ratios was right-tailed (**Supplementary Fig. 1**), with enrichment for specific RNAs of up to 990-fold for HuR and 1,370-fold for Vts1. The RNAs with the highest ratios contained sequences clearly related to their established preferences (**Figs. 1b,c**). Comparisons of the RNAcompete data to previously determined dissociation constants[11,12] indicate that the RNAcompete ratios correlate with and span the full dynamic range of these previous studies (**Supplementary Fig. 2**). In addition, the average values for individual 7-mers correlated well between the two unstructured RNA subsets (set A and set B), for both proteins (**Fig. 1d,e**). In contrast, whereas Vts1 displayed a correlation between the ratios obtained from RNAs containing identical loop sequences (but different stems) in set A and set B, HuR did not (**Supplementary Fig. 3**), consistent with the accepted preference of Vts1 for structured RNA[13] and HuR for unstructured RNA[15]. An apparent contradiction for Vts1 is that it also bound to many RNAs from the 'unstructured' RNA subsets (**Fig. 1c,e**); however, closer inspection reveals that the top-scoring 7-mers all contain the CNGG motif flanked by sequences that can base-pair (**Fig. 1e**), and all of the top-scoring 35-mers contained the CNGG embedded in potential short RNA hairpins (**Fig. 1c**), which is consistent with the preference of Vts1 for structured RNA. These observations indicate that RNAcompete can be used to reliably identify preferred binding sequences for RBPs, whether these are in structured or unstructured RNA contexts.

We further evaluated RNAcompete using seven additional RBPs (**Fig. 2**). Altogether, the nine proteins analyzed encompassed four classes of RNA-binding domains (RRM, K homology (KH), SAM and cold-shock domain (CSD)). Except for Vts1, we did not observe a correlation between the independent probe sets on the microarray representing RNA loop sequences, suggesting that none of these additional proteins have an exclusive preference for specific structured RNAs (data not shown). Rather, all of these RBPs displayed strong preferences for specific unstructured sequences. This is consistent with previous structural studies, which indicate that interaction with unpaired bases is a primary mode by which many RBPs with canonical RNA-binding domains recognize specific RNA sequences[6]. Preference for specific primary sequences does not rule out the possibility that these proteins may also prefer primary sequences contained within RNA secondary structures not represented on the array. For example, although U1A will bind its AUUGCAC target sequence in an unstructured context, it is known to have greater affinity for the same site within a specific loop of ten or more bases[14],

Make Notes (always!)

selection are positively charged groups (e.g. primary amino groups), the presence of hydrogen bond donors and acceptors and planarity (aromatic compounds) (Wilson and Szostak, 1999; Rimmele, 2003). The aptamer selection is more difficult for targets with largely hydrophobic character and for negatively charged molecules (e.g. containing phosphate groups). These target requirements are caused by the basic principles of the intermolecular interactions in an aptamer–target complex. The aptamers bind to their targets by a combination of complementarity in shape, stacking interactions between aromatic compounds and the nucleobases of the aptamers, electrostatic interactions between charged groups or hydrogen bondings (Patel, 1997; Patel et al., 1997; Hermann and Patel, 2000). In presence of the target, and on formation of the binding complex, the aptamers undergo adaptive conformational changes. The folding into defined three-dimensional structures permits the aptamers to completely encapsulate small target molecules by generating a specific binding pocket. In higher molecular weight targets like proteins, different substructures on the molecule surfaces are involved in aptamer binding. For example, side chains of basic amino acids (lysine, arginine) are often responsible for intermolecular hydrogen bondings.

The binding complexes of aptamers targeting nucleic acids

special sens
the T7 pron
to convert t
dsDNA is t
resulting in
RNA SELE
each round
amplified b
The new RI
again by in
Importa
are the size
and chemic
tide librarie
used for ir
selected ap
sequence a
(1992) sele
library of 9
sequence.
aptamer va
was retaine
regions are
libraries ar

Types of interactions

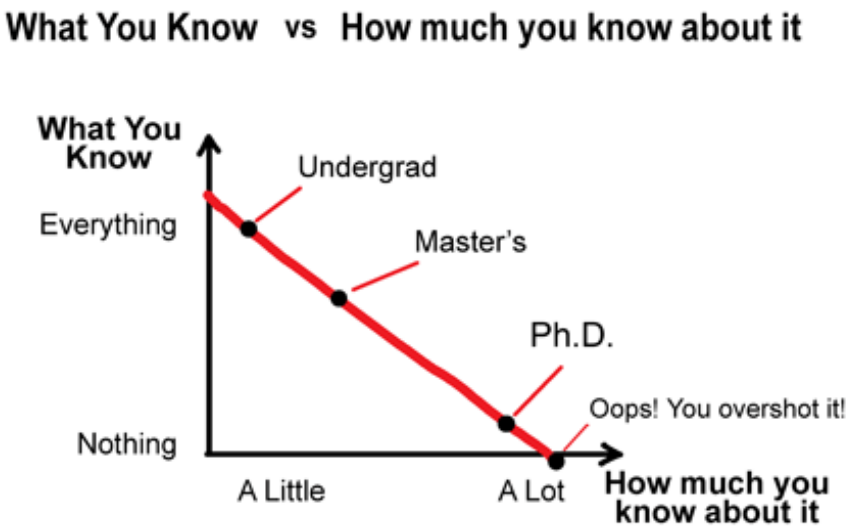
- Take Notes
  - In the margin, common notebook or gdocs
  - Use a pen or highlighter, box important info
  - Summarize the figures and main findings with page numbers or figure numbers
  - Circle the important references

Understanding Journal Types

- JBC – Top Rated
- Science
- Nature
- JACS – Society Journals
- Biotechniques – Methods!
- Natural Product Research
- Biochemistry
- Cell
- Journal of Natural Products - Reputation continues to grow!

Now That You Know Everything…

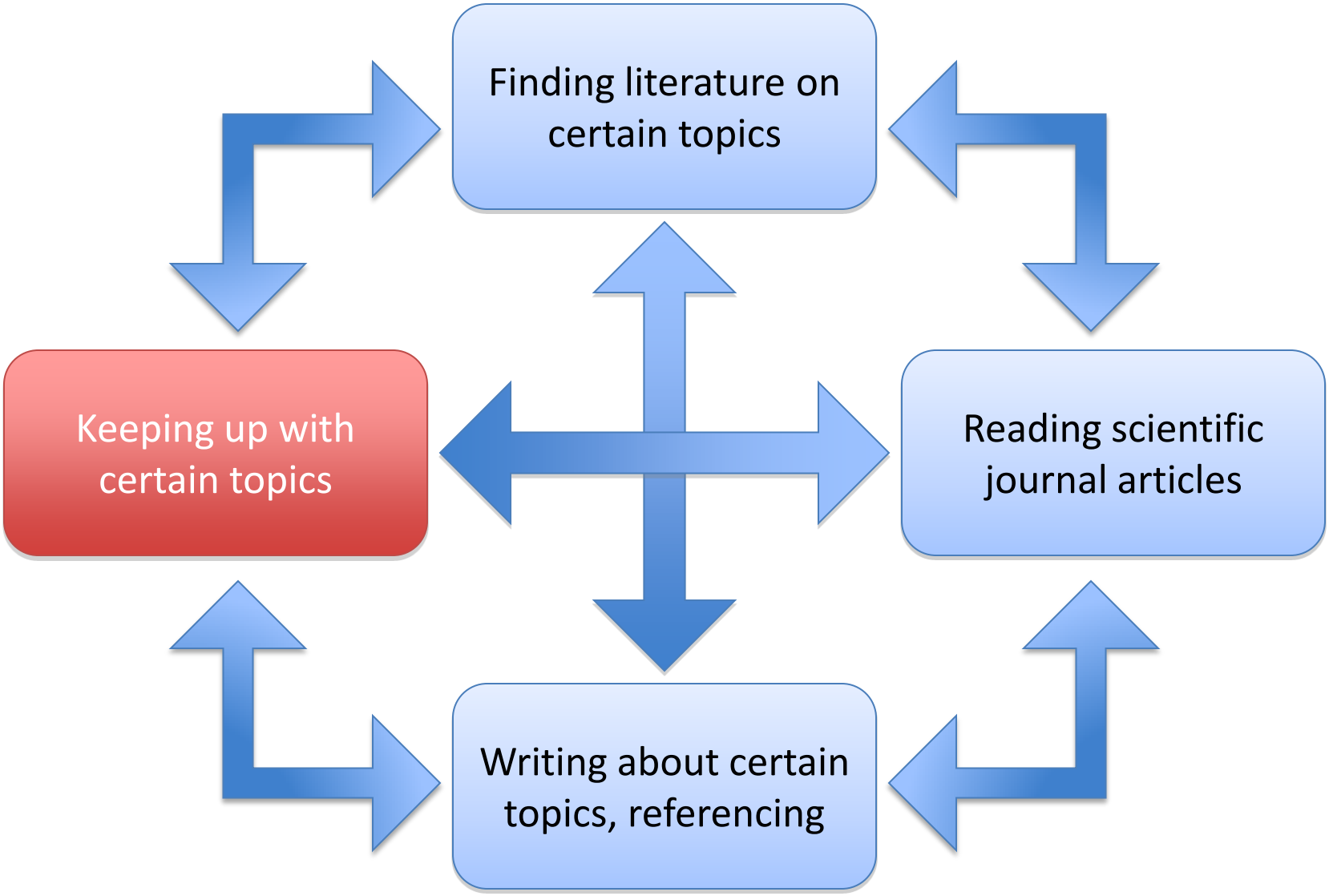
Finding literature on certain topics



Reading scientific journal articles

Writing about certain topics, referencing

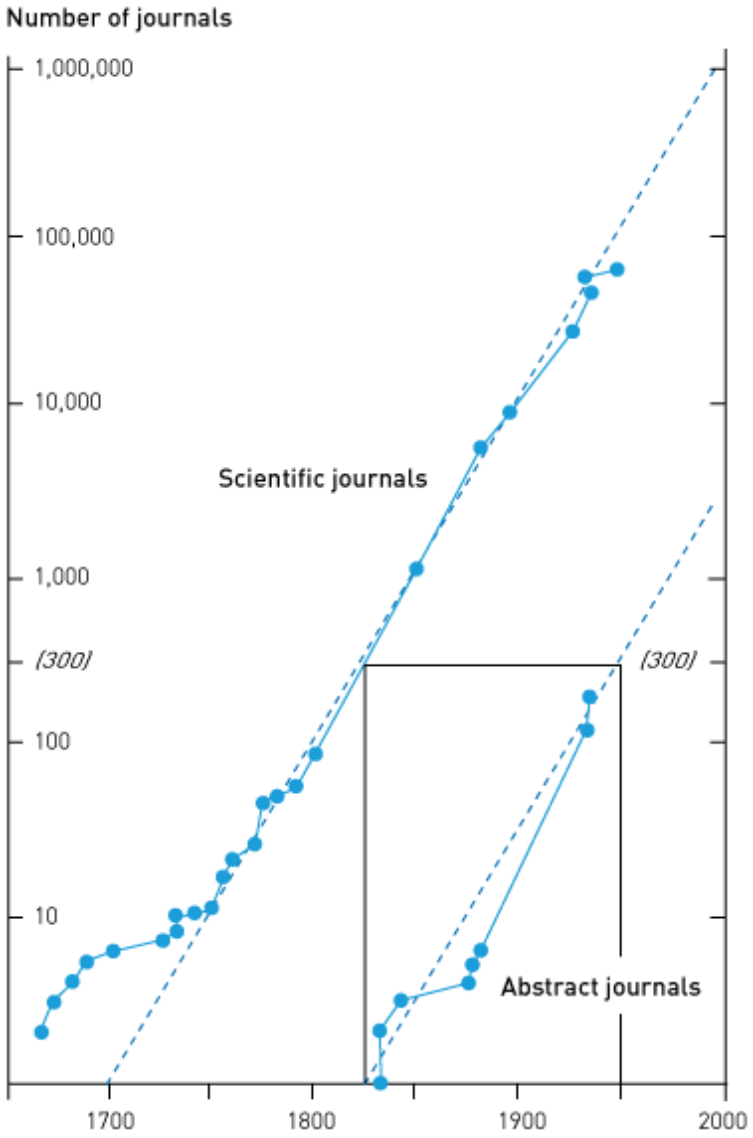
Break it into the Four Main Tasks



The Problem

How To...

- ... find relevant literature?
- ... prepare time to read, reread and read again?
- ... not fall asleep when reading?
- ... store (un)read articles?
- ... organize knowledge gleaned in an easily accessible manner?
- ... reduce time needed for all this?

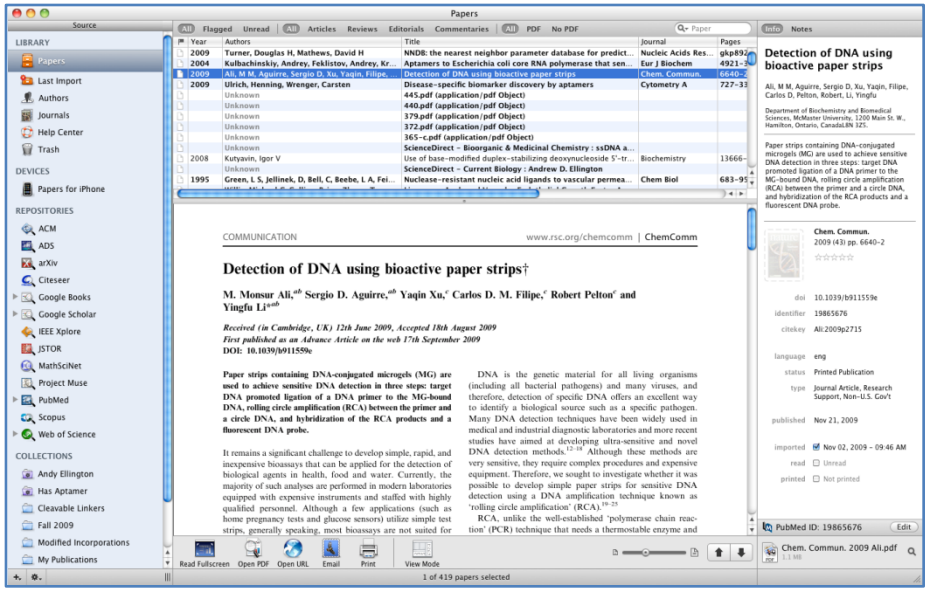


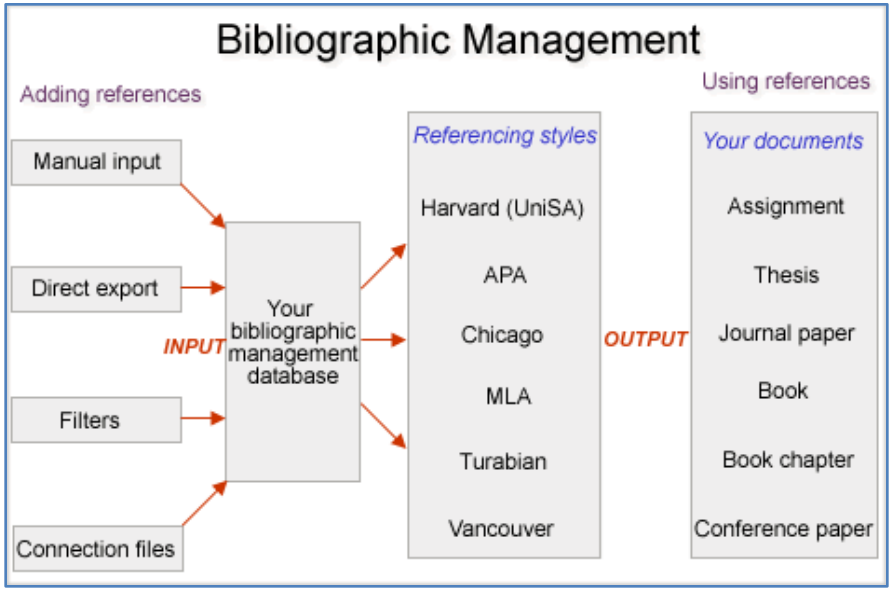
Keeping Up With the Jones

- Pubmed/WoS
  - Search weekly
- Podcasts
- Feeders
  - Pubcrawler
  - RSS feeds
- Flip through hard copy journals or look at publishers websites regularly
  - High impact journals
- Others
  - Journal Clubs, Professional Societies/Meetings, On campus lectures and seminars

What are you looking for?

- PDF Organization?
- Cite While You Write Refs?

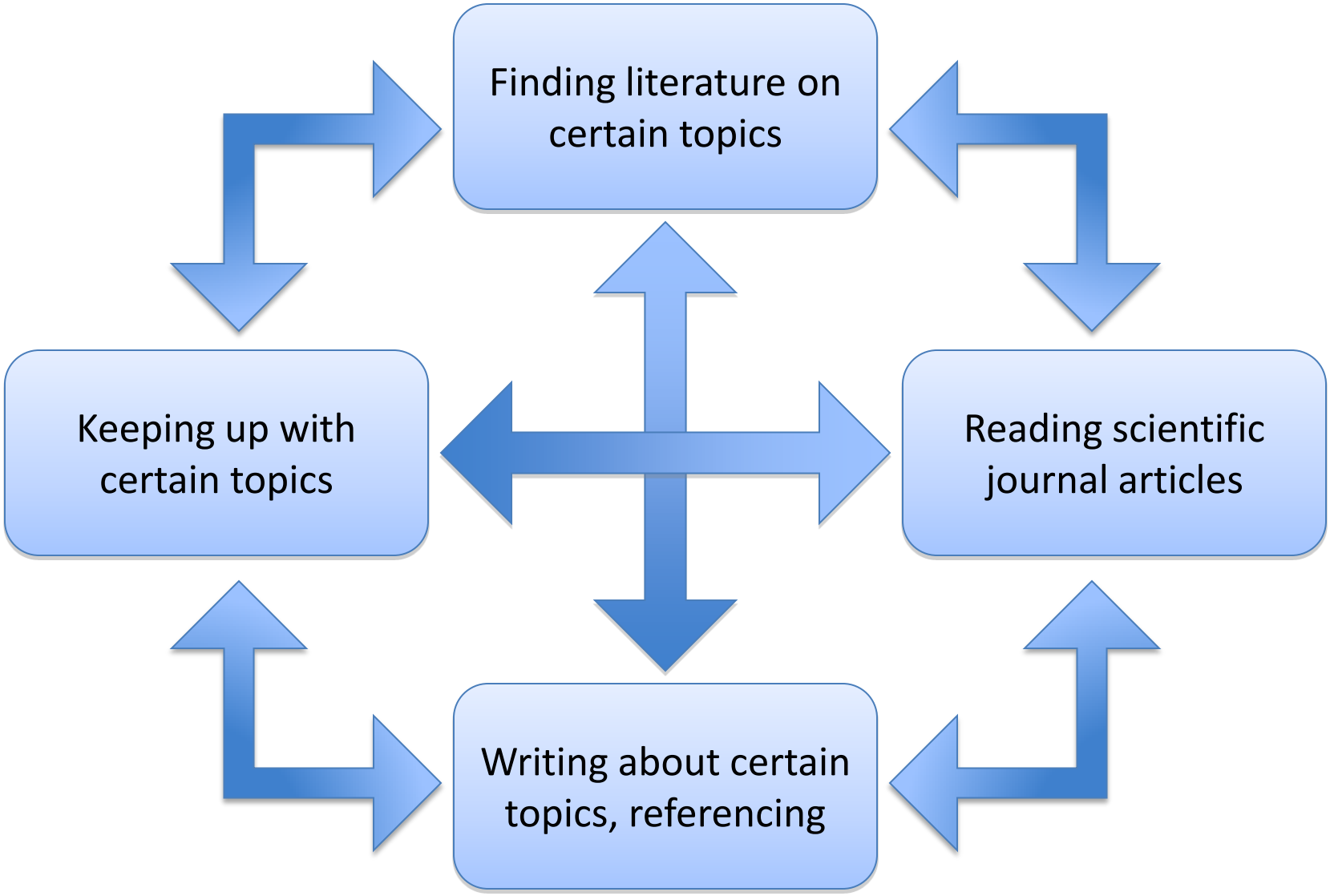




Sente 6

zotero

In Summary


Finding literature on certain topics
Reading scientific journal articles
Writing about certain topics, referencing
Keeping up with certain topics