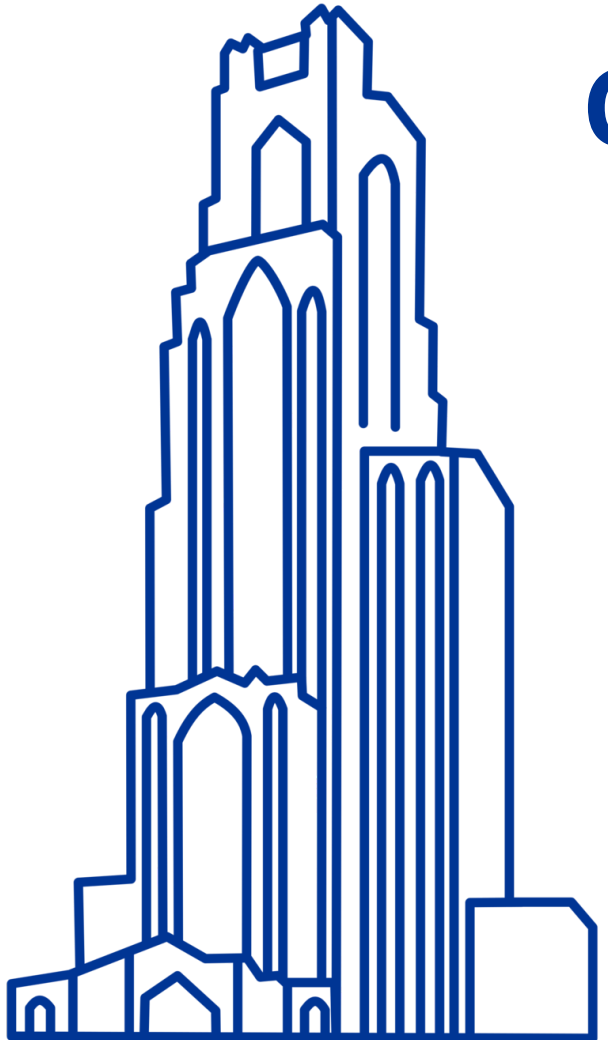


Computational Biology Seminar

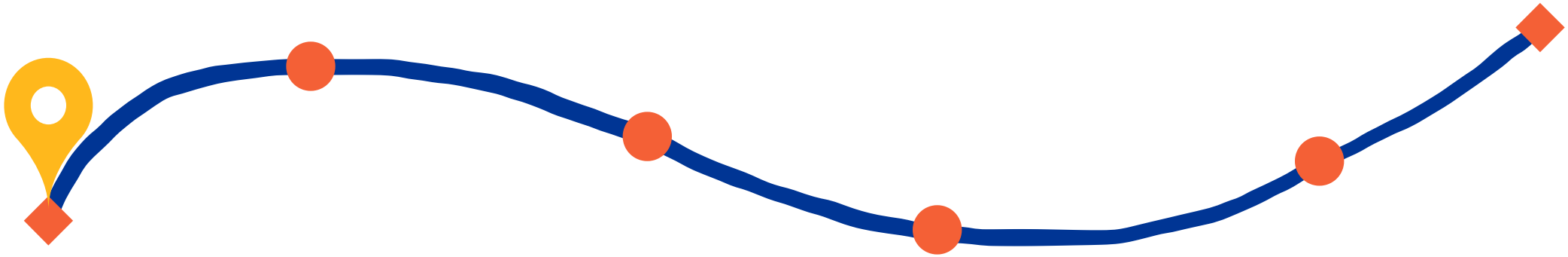
(BIOSC 1630)

Lecture 01

Aug 28, 2024

University of Pittsburgh

After today, you should be able to



1. **Understand course structure and expectations.**
2. Outline the requirements and goals of the perspective primers.
3. Compare and contrast different types of scientific articles.
4. Explain the key components of the research ecosystem.
5. Apply effective strategies to find relevant literature.

Meet your teaching team

Instructor
Alex Maldonado, PhD
he/him/his

Postdoctoral Associate in Computational Biology

B.S.E in Chemical Engineering, 2018
Western Michigan University

Ph.D. in Chemical Engineering, 2023
University of Pittsburgh

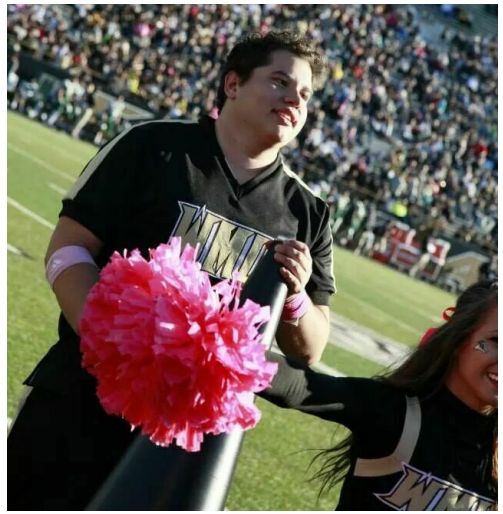
Office hours: By appointment

Email: alex.maldonado@pitt.edu

Acceptable ways to address me: **Alex** (preferred) Dr. Maldonado Dr. Alex Dr. M

Alex's fun facts

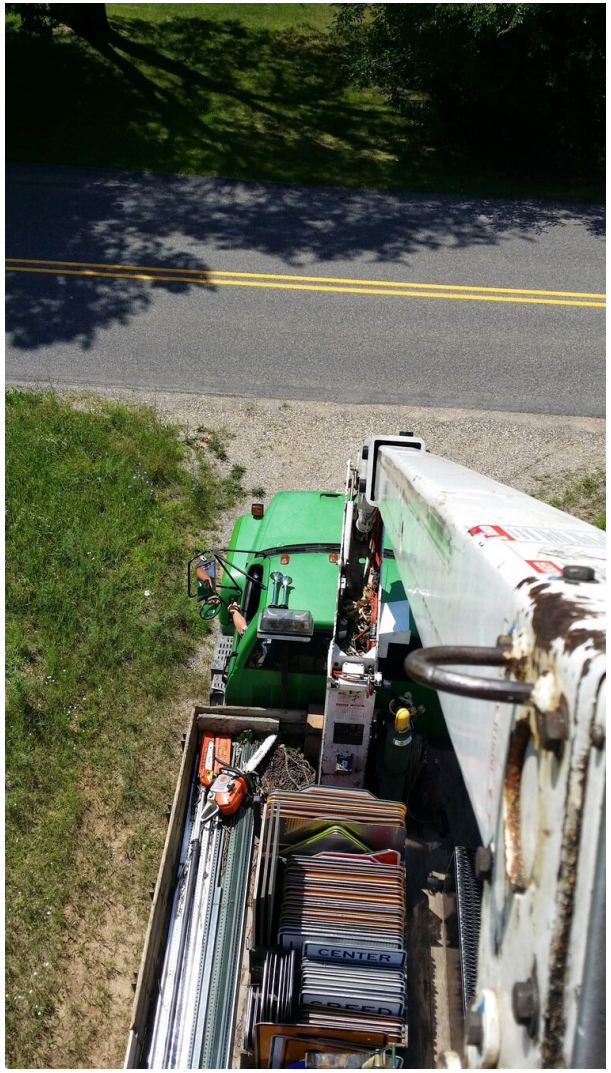
Every male in my (maternal) family played football —I rebelled



Alex's fun facts

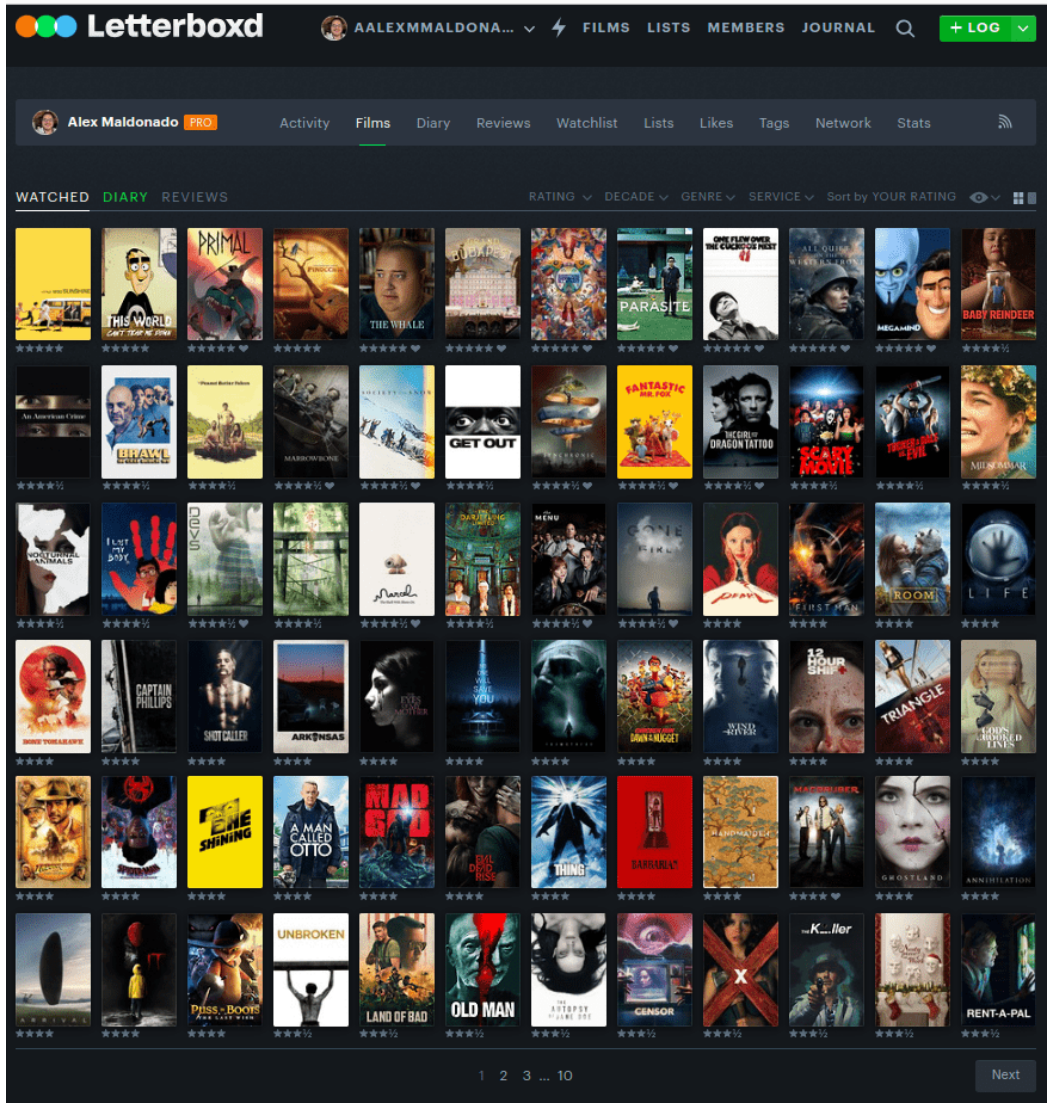
Part-time jobs

- Construction
- UPS package handler
- Kent County Traffic safety
- Jimmy John's delivery driver
- Wings West ice events

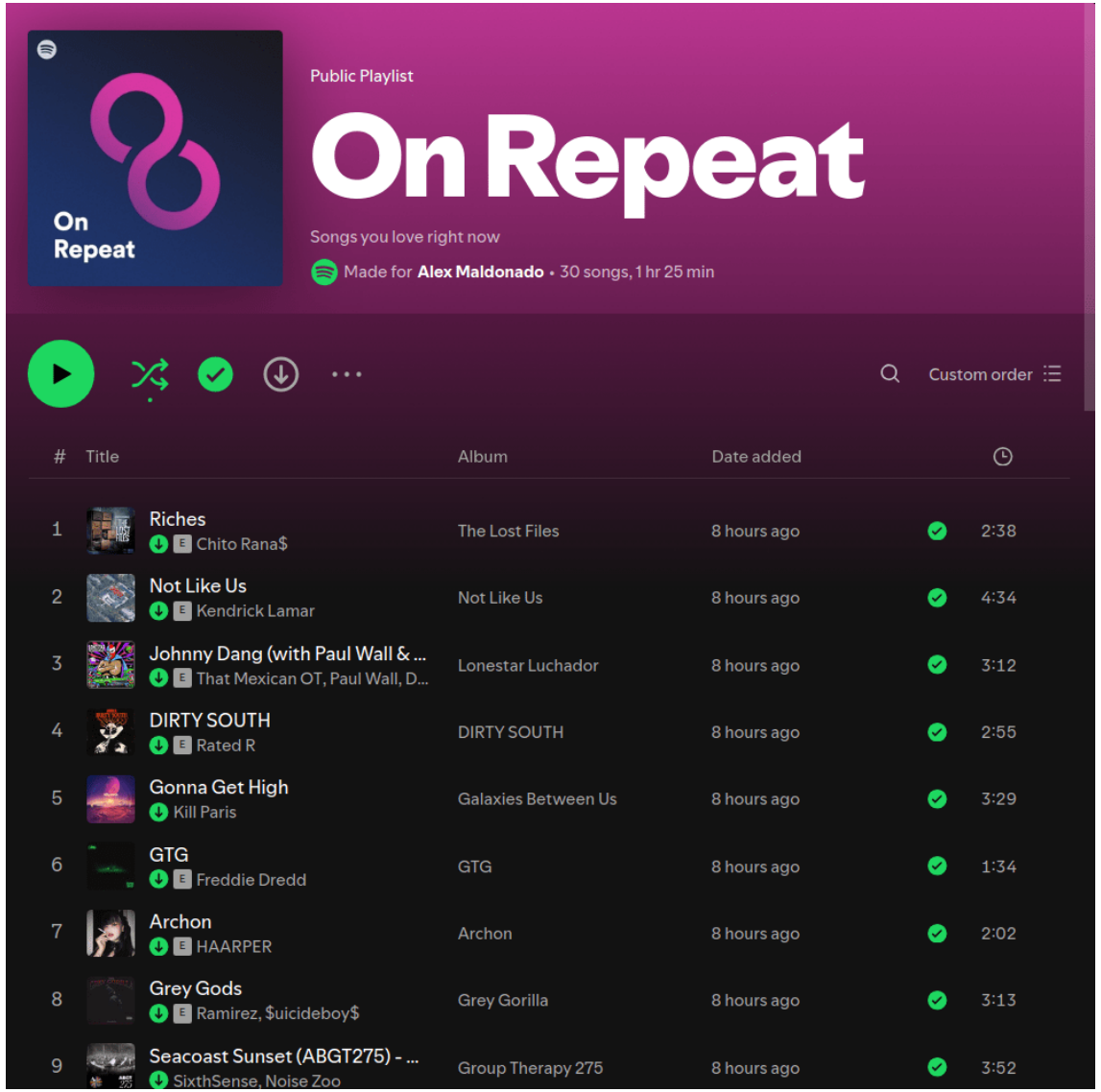


Get to know my ...

Movie taste



Music taste



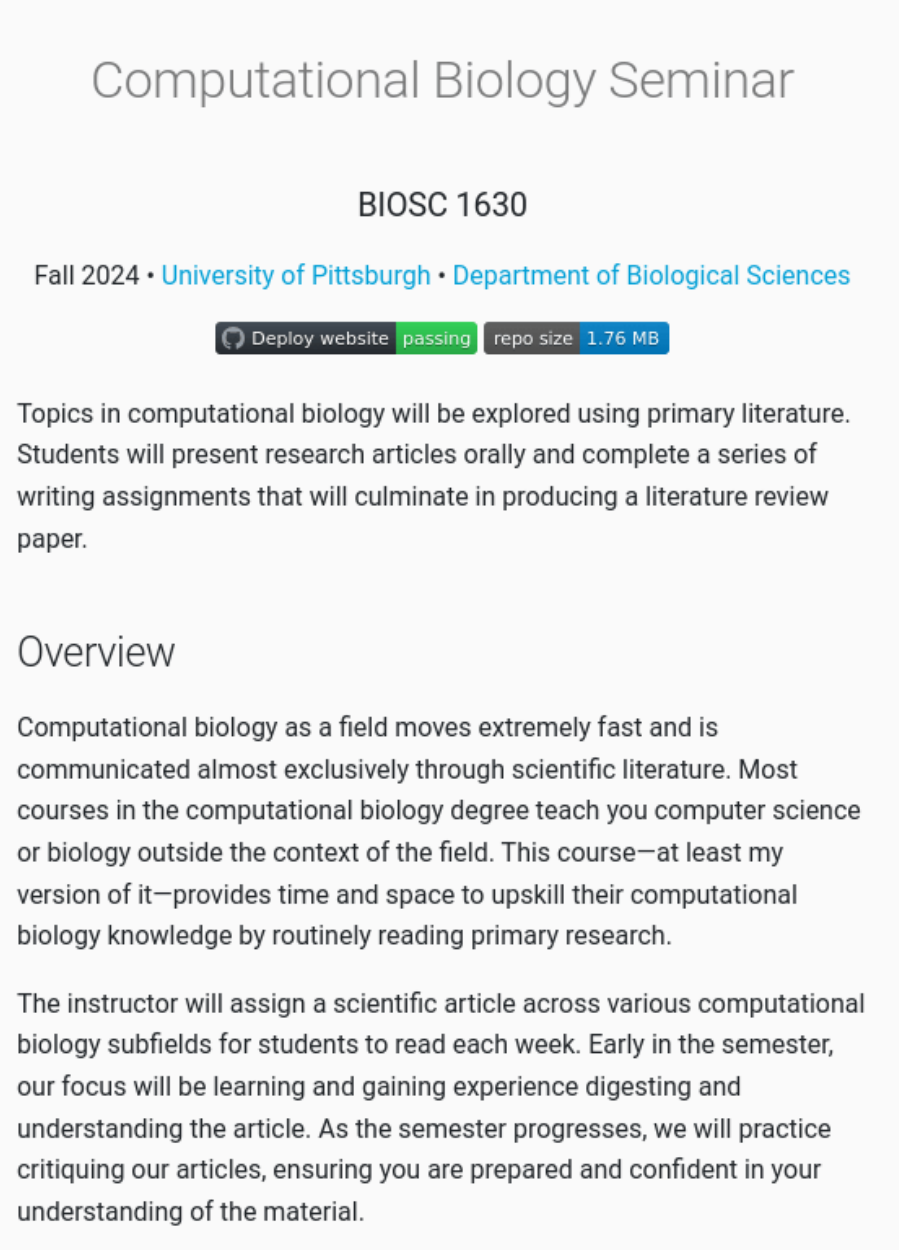
Tessa the cat

All course material is hosted online

Website: pitt-biosc1630-2024f.oasci.org/

Things that contain student information will be only on Canvas to be FERPA compliant

Assignments will be submitted on Gradescope



Computational Biology Seminar

BIOSC 1630

Fall 2024 • University of Pittsburgh • Department of Biological Sciences

Deploy website passing | repo size 1.76 MB

Topics in computational biology will be explored using primary literature. Students will present research articles orally and complete a series of writing assignments that will culminate in producing a literature review paper.

Overview

Computational biology as a field moves extremely fast and is communicated almost exclusively through scientific literature. Most courses in the computational biology degree teach you computer science or biology outside the context of the field. This course—at least my version of it—provides time and space to upskill their computational biology knowledge by routinely reading primary research.

The instructor will assign a scientific article across various computational biology subfields for students to read each week. Early in the semester, our focus will be learning and gaining experience digesting and understanding the article. As the semester progresses, we will practice critiquing our articles, ensuring you are prepared and confident in your understanding of the material.

Reading and critiquing research

We will be routinely reading primary research articles within computational biology

New methods come out all of the time, which ones should you use?

Will enhance your knowledge of the field

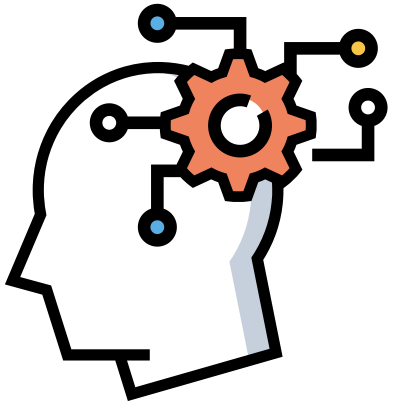
You will read one research article every two weeks

I have picked several papers across computational biology published in 2023

A pre-class assignment will ensure you are prepared for discussion-based activities

Enhancing your critical thinking by reading literature

"I'm not going to graduate school, so why am I taking a whole class reading literature?"



Reading literature is a great activity to practice **critical thinking and skepticism**

Peer reviewed does not mean infallible

BIOSC 1640 is where you problem solve

Problems, solutions, and ideas need to be shared

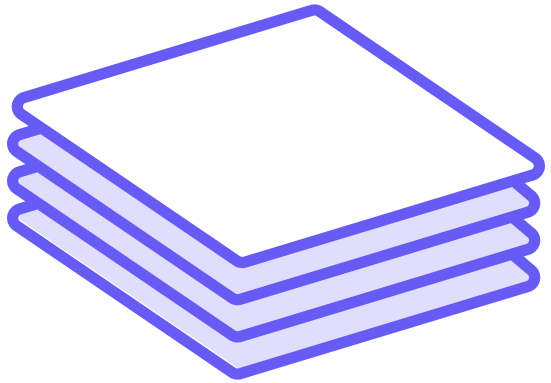
Every job description likely contains "Excellent communication skills"

We will help you hone your communication skills with writing a perspective paper

You will write a perspective paper during the semester

This is an atypical two-credit course

Writing-intensive courses require 11 to 13 single-spaced pages of writing



Paper is chunked into smaller assignments due periodically throughout the semester

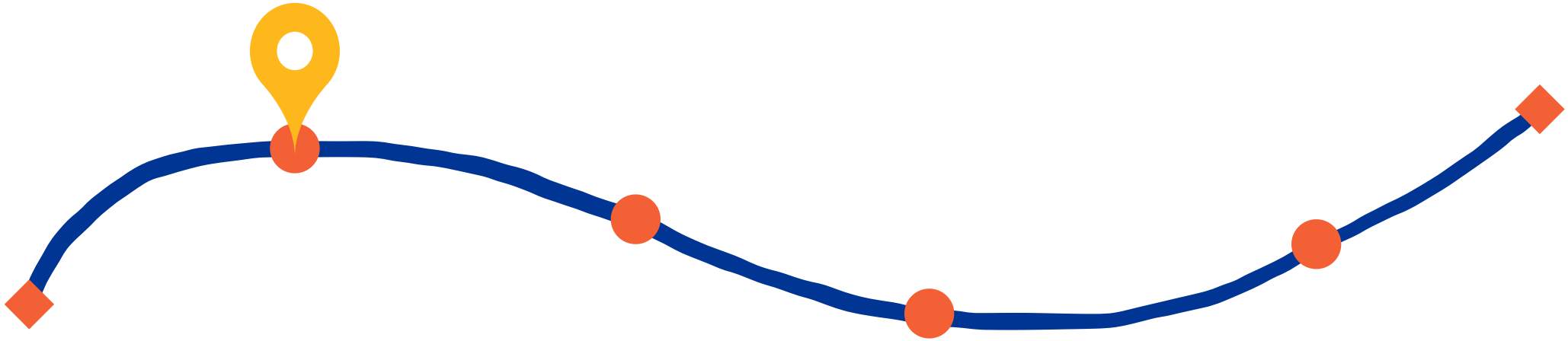
- **Theme analysis**: 6%
- **Literature review**: 8%
- **Outline**: 6%
- **Introduction**: 6%
- **Field overview**: 8%
- **Key challenges**: 8%
- **Future directions**: 4%
- **Conclusion**: 4%
- **Peer review**: 10%
- **Final draft**: 10%

I will accept the following graduate fellowship applications in lieu of writing a perspective

- **NSF GRFP**: Humanities and STEM
- **DOE CSGF**: High-performance computing
- **Hertz Fellowship**: Applied STEM

Research statements for these applications are more rigorous than this perspective

After today, you should be able to

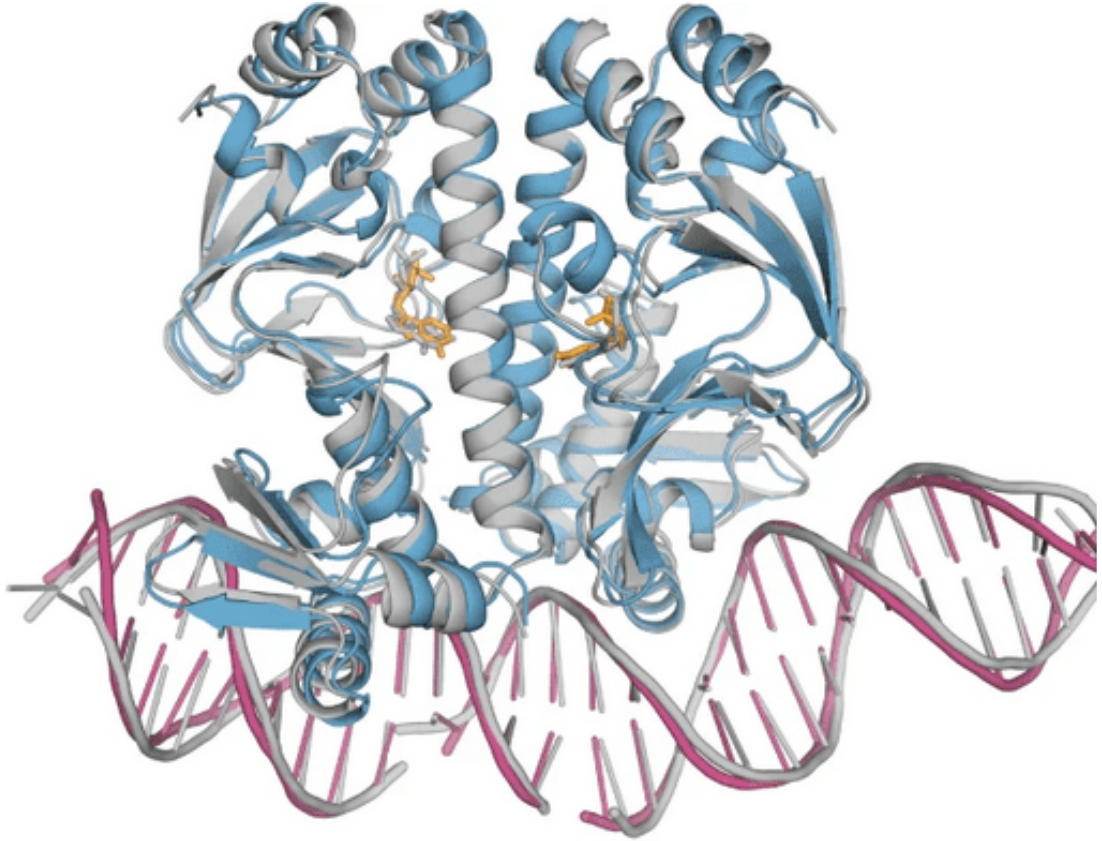


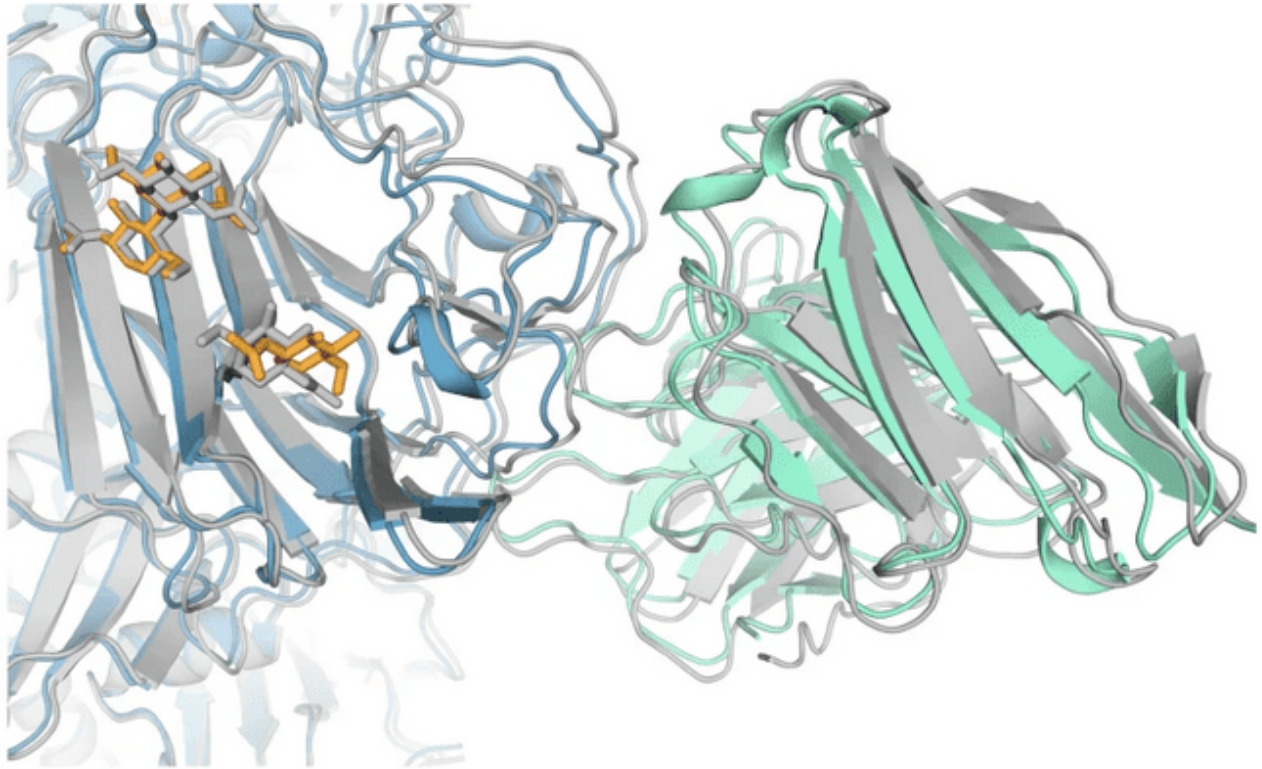
1. Understand course structure and expectations.
2. **Outline the requirements and goals of the perspective primers.**
3. Compare and contrast different types of scientific articles.
4. Explain the key components of the research ecosystem.
5. Apply effective strategies to find relevant literature.

Primer 1: Protein structure prediction

AlphaFold has undoubtedly revolutionized protein structure prediction

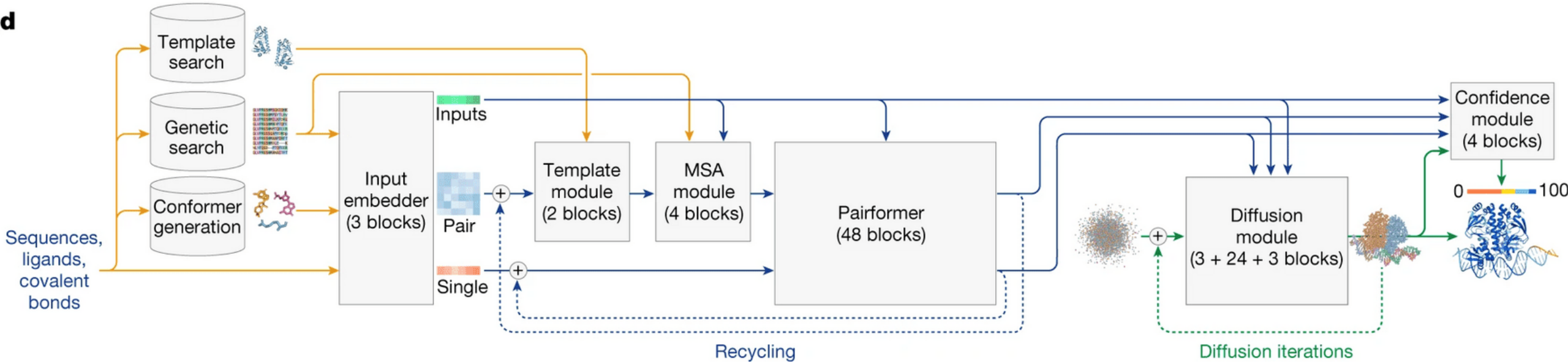
AF3 can predict any (ordered) biomolecule with post-translational modifications





Primer 1: Protein structure prediction

AlphaFold is a complicated machine learning model that is trained to **reproduce experimental structures**

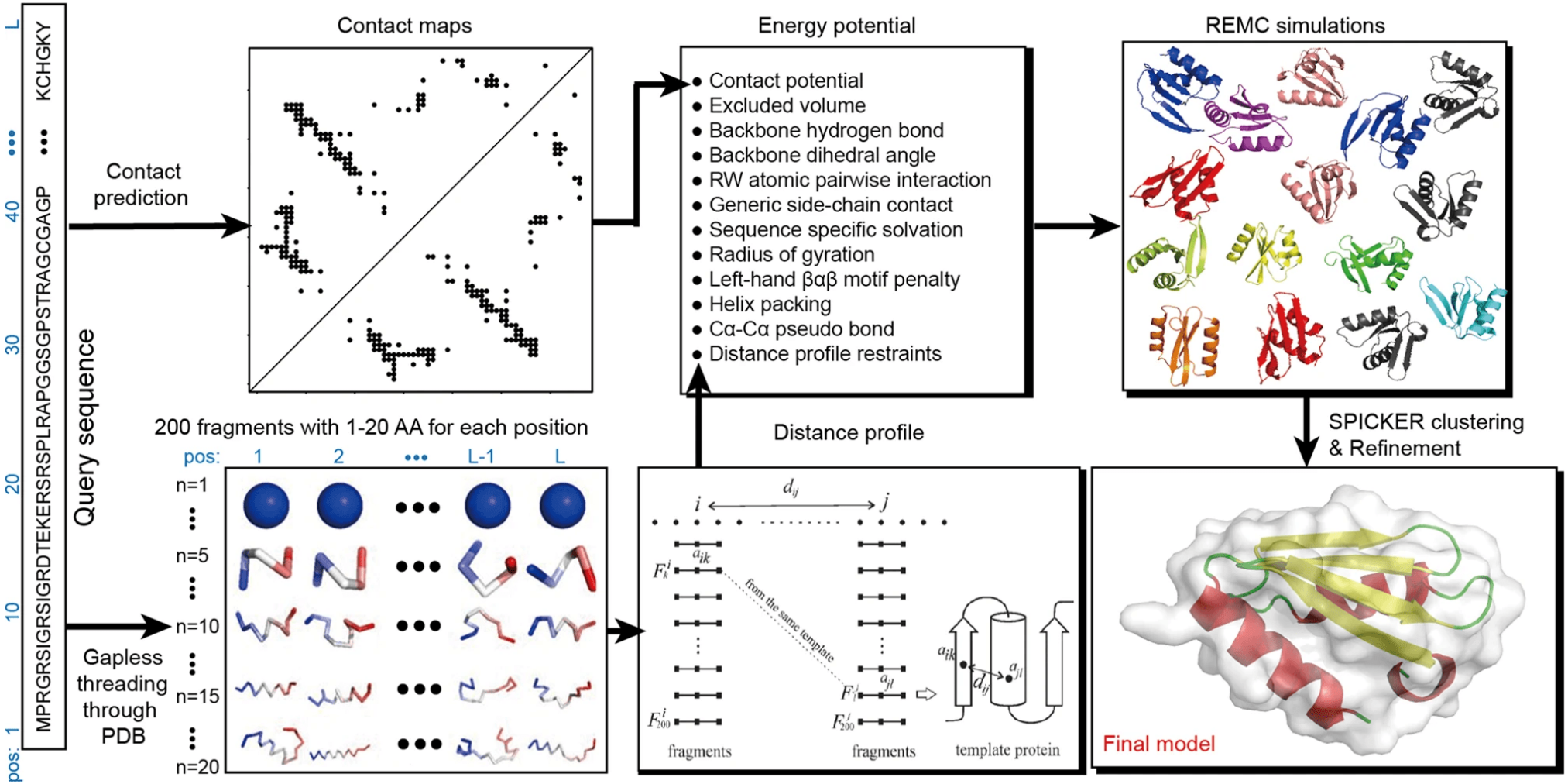


It achieves transferability by learning to find coevolutionary signals of residue-residue proximity

Primer 1: Protein structure prediction

Ab initio methods use physical representations and calculations to compute structures

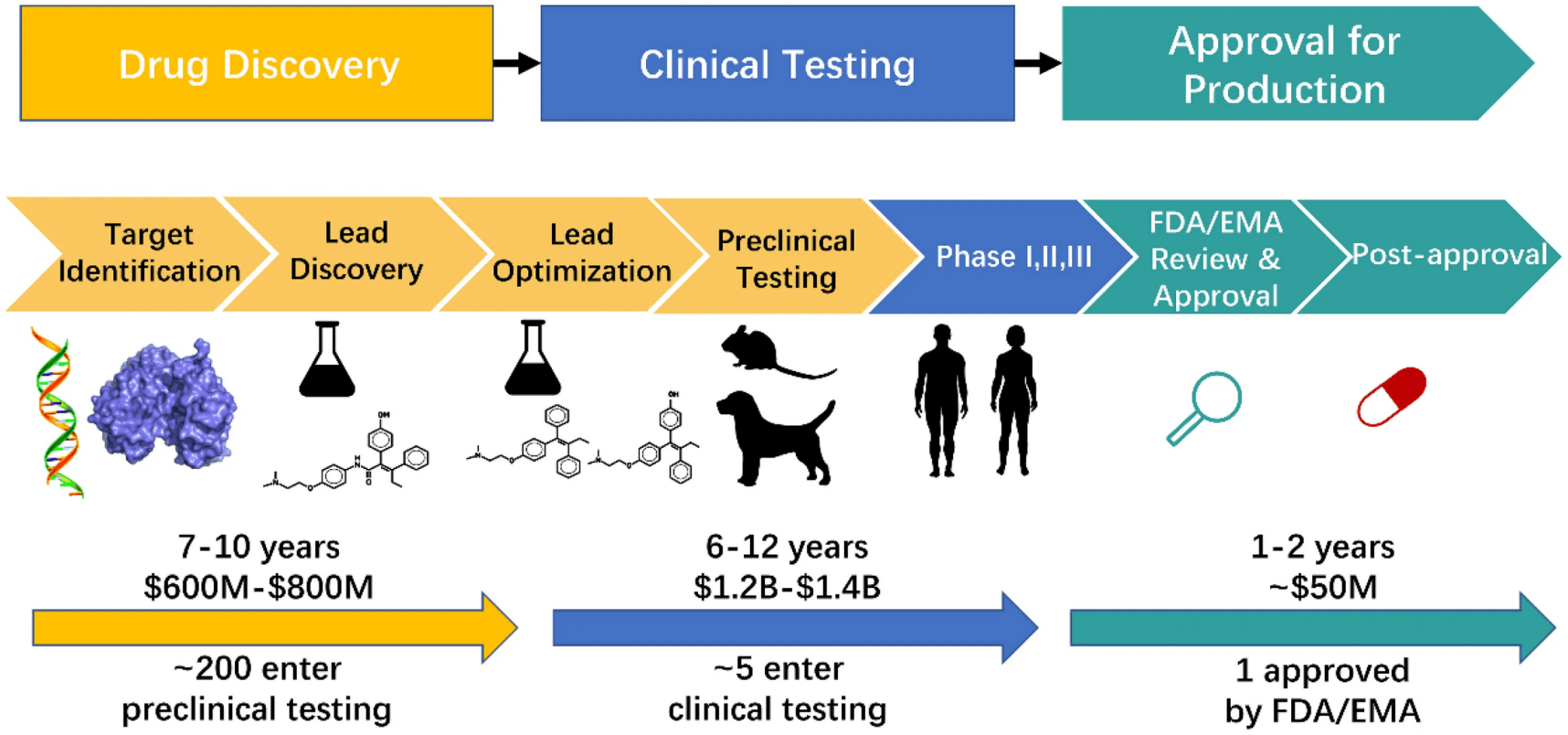
Expensive but interpretable



Should we abandon *ab initio* methods for protein structure prediction?

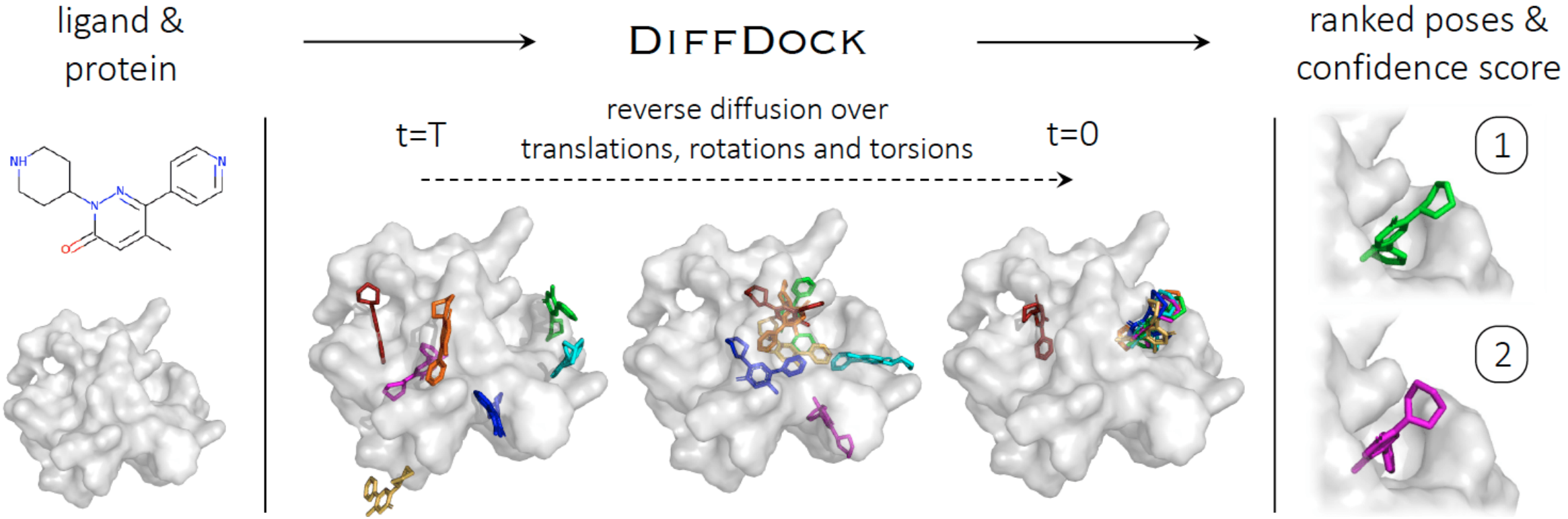
Primer 2: Molecular dynamics for drug design

Drug discovery aims to identify drug candidates from a large search space



Primer 2: Molecular dynamics for drug design

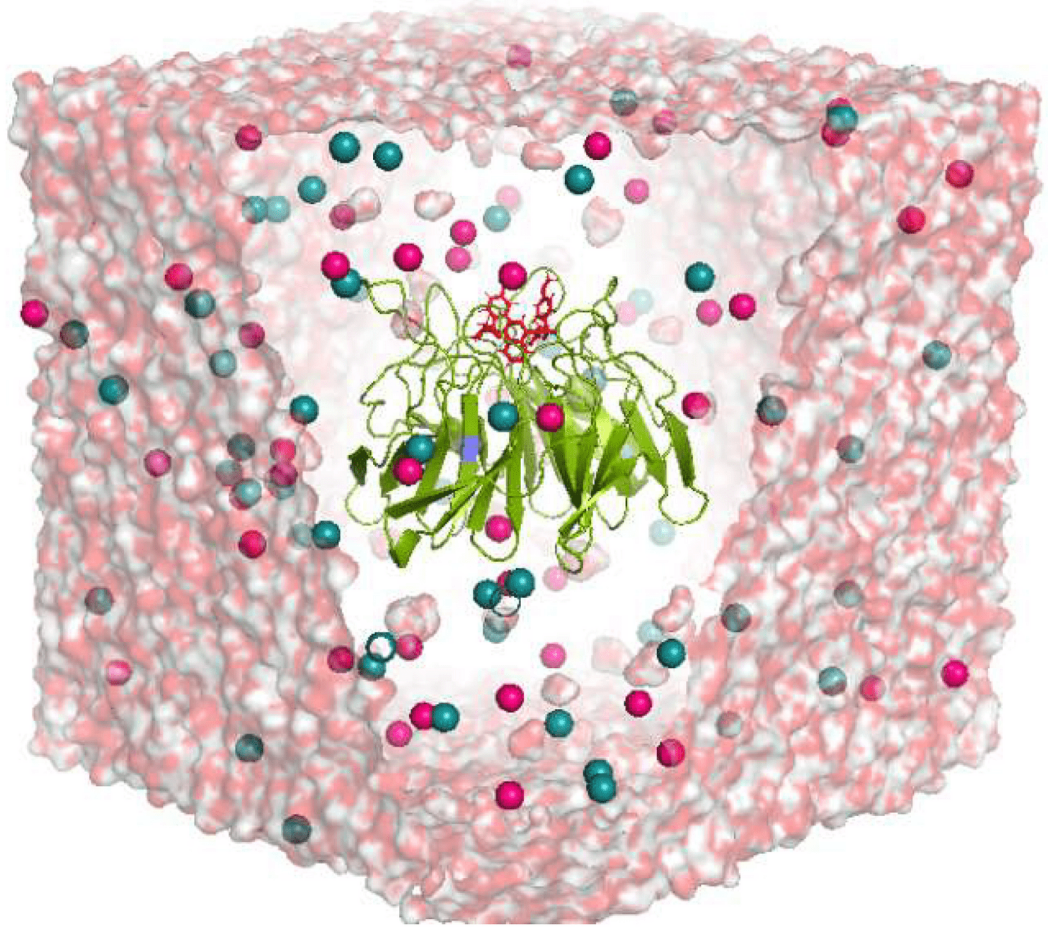
Protein-ligand docking can rapidly identify binding poses based on parameterized scoring functions and/or machine learning



Primer 2: Molecular dynamics for drug design

Molecular simulations account for solvent, protein, and ligand dynamics

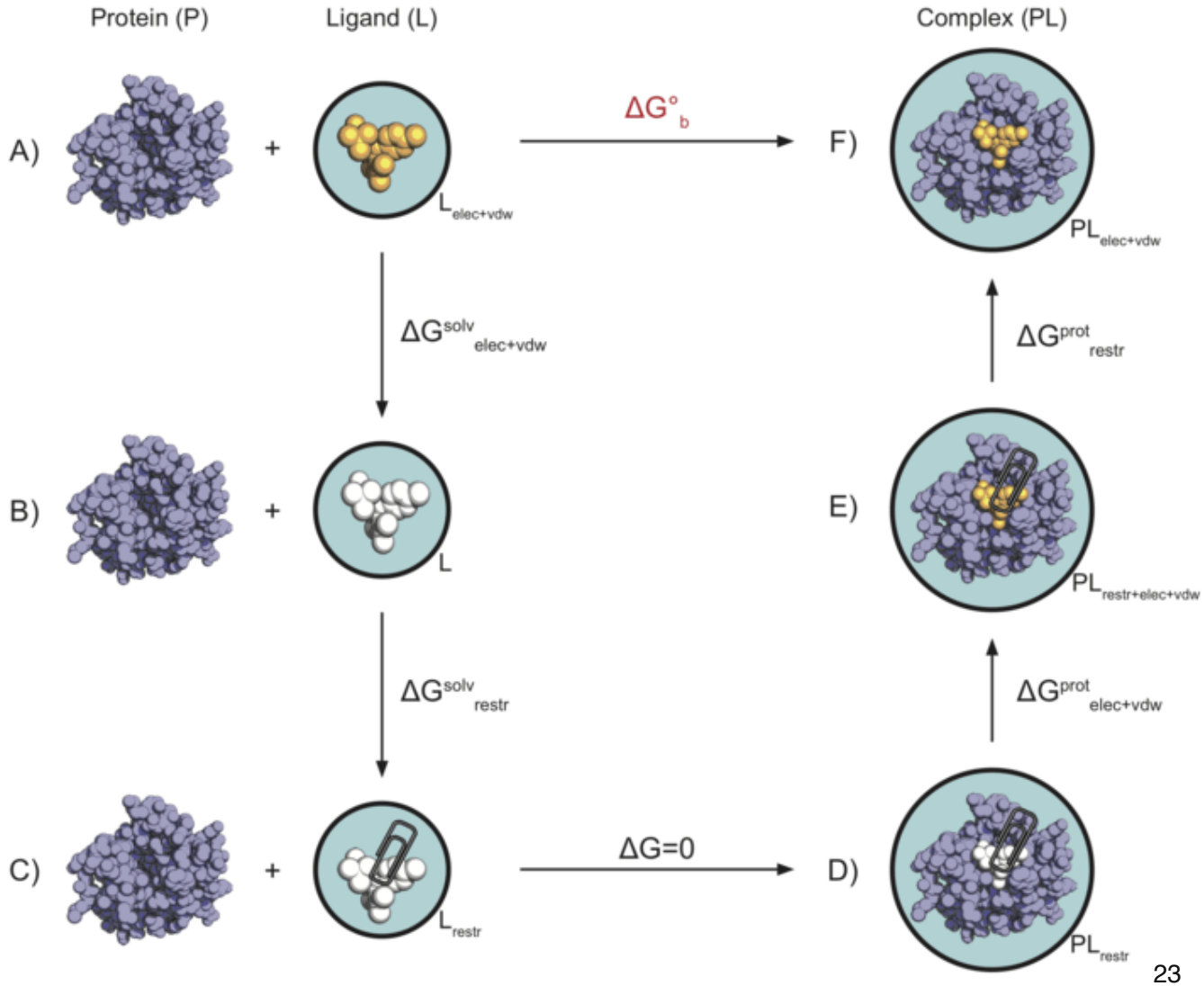
Significantly enhanced conformational sampling and free energies (e.g., entropy)



Primer 2: Molecular dynamics for drug design

Alchemical simulations systematically turn off the ligand's interaction with proteins

With experimental assays becoming more economical, why spend the time running simulations?

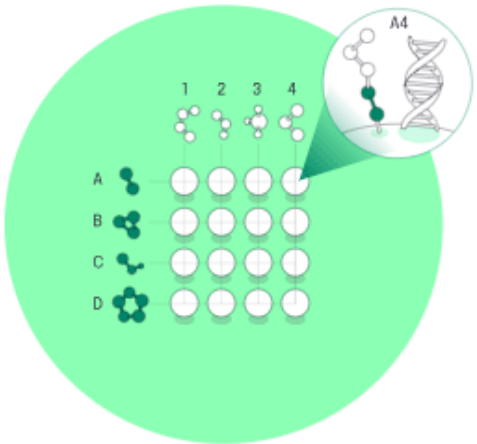


Reproduced from Aldeghi *et al.*, Chem. Sci., 2016, **7**, 207-218. © The Royal Society of Chemistry

Primer 2: Molecular dynamics for drug design

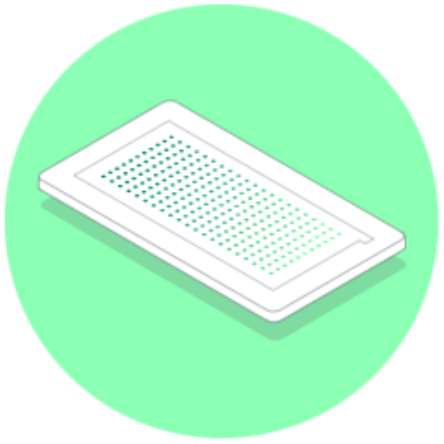
Experimental techniques probe drug action before investing in clinical trials

Example: Terray Therapeutics



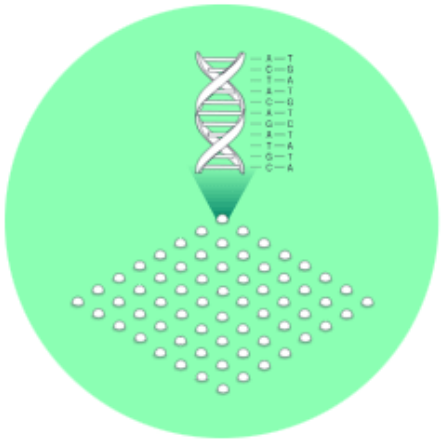
Make encoded libraries on nanobeads

Molecules are combinatorially synthesized on nanoparticle beads and DNA barcodes are attached to identify which molecule is on each bead.



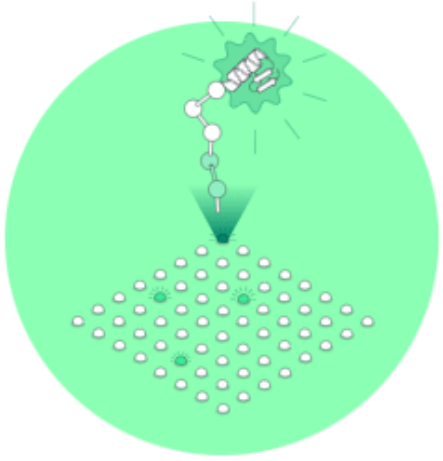
Immobilize beads on tArray chip

Beads are randomly placed into 32M wells on a chip the size of a nickel.



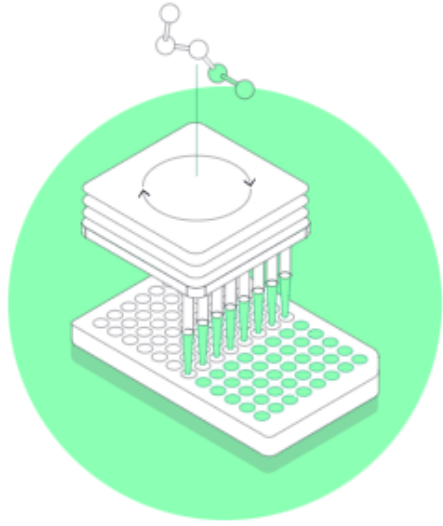
Map position of beads

DNA barcodes are sequenced to map the position of each molecule, then DNA is cleaved off the beads.



Measure target-ligand binding at scale

Precise binding affinity is measured based on fluorescence intensity in less than 4 minutes per chip.

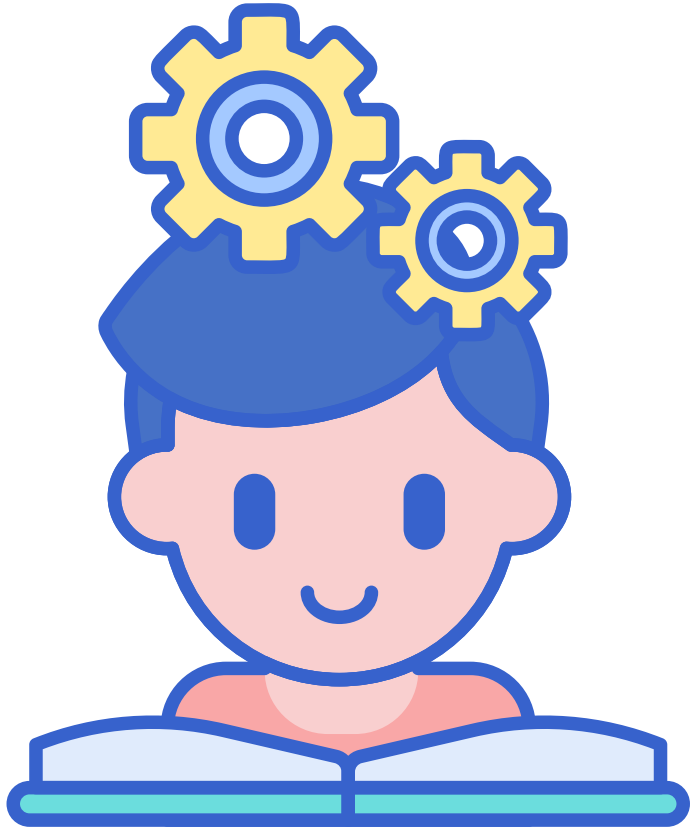


Resynthesize and test additional properties

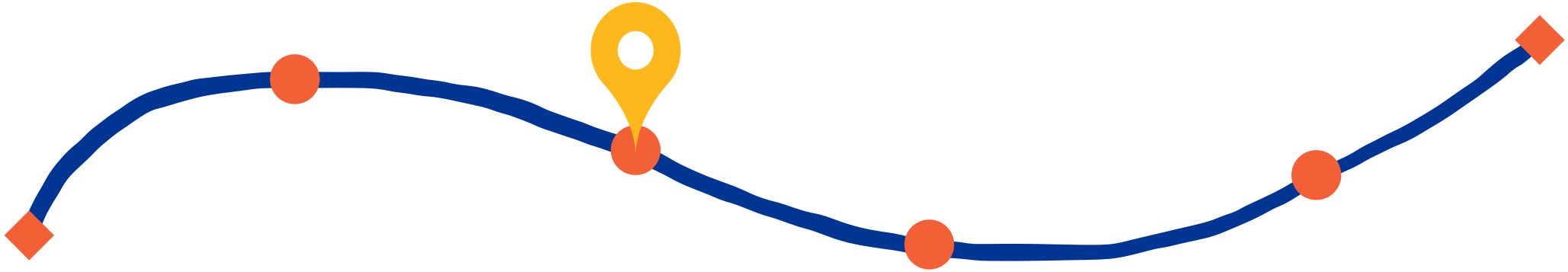
Hits are resynthesized at microscale and further evaluated using high-throughput biological testing.

Should we bypass molecular simulations and go straight to experimental assays?

Each primer has no right answer; I'm just looking for your opinion with supporting evidence



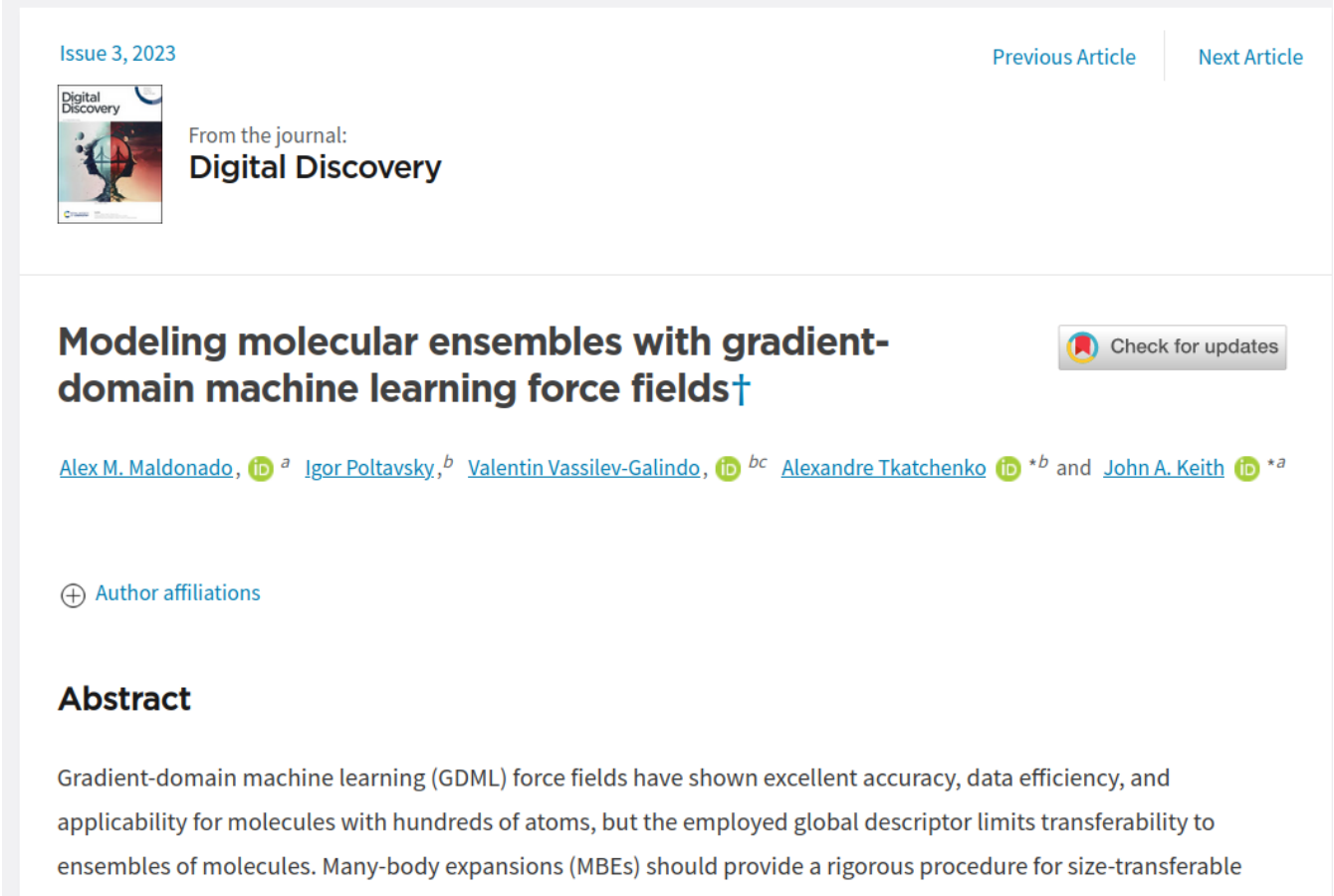
After today, you should be able to



1. Understand course structure and expectations.
2. Outline the requirements and goals of the perspective primers.
3. **Compare and contrast different types of scientific articles.**
4. Explain the key components of the research ecosystem.
5. Apply effective strategies to find relevant literature.

Primary literature: Original research by scientists in the field

- Contains novel computational methods, algorithms, or analyses
- Published in peer-reviewed journals
- Includes detailed methodology, datasets, and statistical analyses
- Often accompanied by open-source code or software tools



Issue 3, 2023

Previous Article | Next Article

From the journal:
Digital Discovery

Modeling molecular ensembles with gradient-domain machine learning force fields†

Check for updates

Alex M. Maldonado, [a] Igor Poltavsky,[b] Valentin Vassilev-Galindo, [bc] Alexandre Tkatchenko [*b] and John A. Keith [*a]

Author affiliations

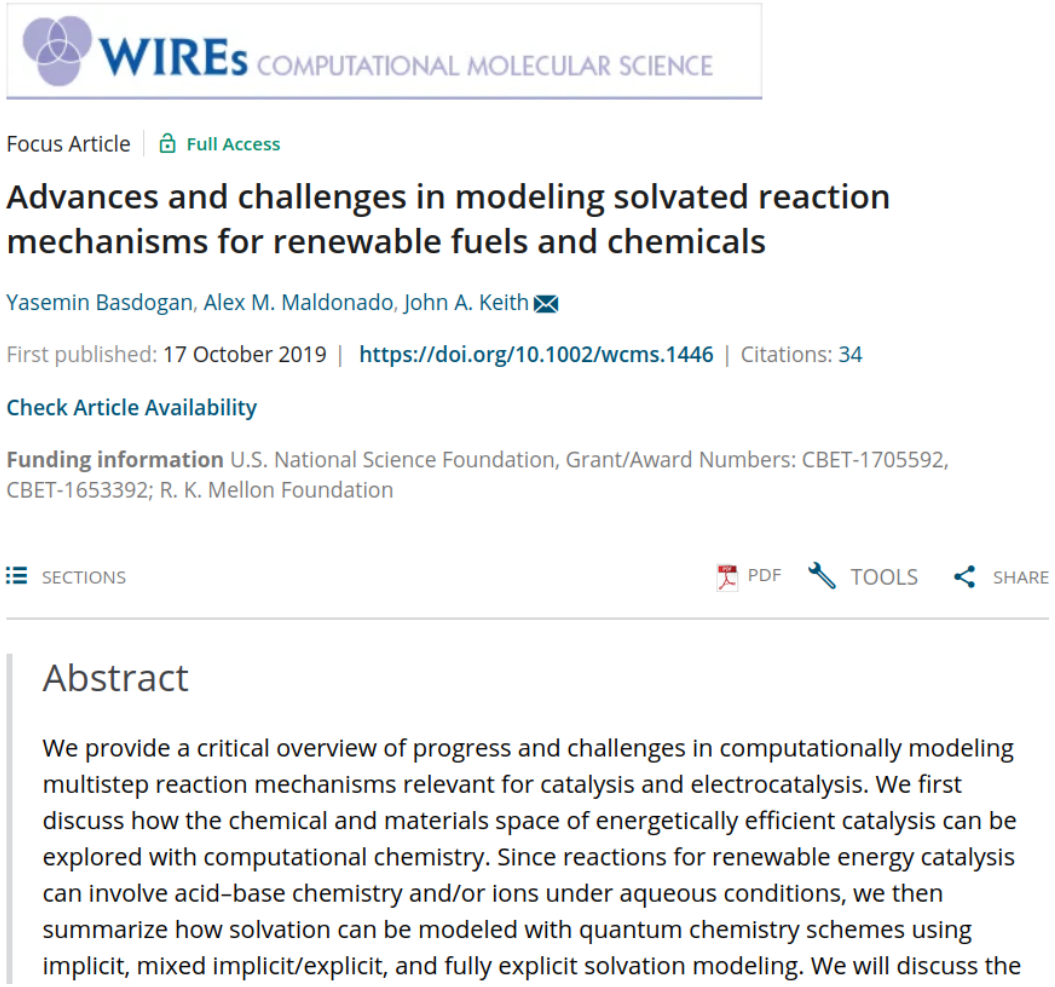
Abstract

Gradient-domain machine learning (GDML) force fields have shown excellent accuracy, data efficiency, and applicability for molecules with hundreds of atoms, but the employed global descriptor limits transferability to ensembles of molecules. Many-body expansions (MBEs) should provide a rigorous procedure for size-transferable

Example

Review: Summarizes the current state of knowledge in a specific area

- Synthesis of multiple primary research articles
- No new experimental data presented
- Analyzes trends, gaps, and future directions in research



WIREs COMPUTATIONAL MOLECULAR SCIENCE

Focus Article | Full Access

Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals

Yasemin Basdogan, Alex M. Maldonado, John A. Keith

First published: 17 October 2019 | https://doi.org/10.1002/wcms.1446 | Citations: 34

Check Article Availability

Funding information U.S. National Science Foundation, Grant/Award Numbers: CBET-1705592, CBET-1653392; R. K. Mellon Foundation

SECTIONS PDF TOOLS SHARE

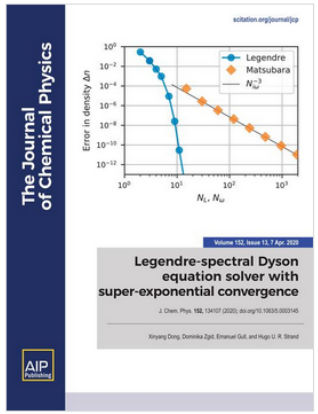
Abstract

We provide a critical overview of progress and challenges in computationally modeling multistep reaction mechanisms relevant for catalysis and electrocatalysis. We first discuss how the chemical and materials space of energetically efficient catalysis can be explored with computational chemistry. Since reactions for renewable energy catalysis can involve acid–base chemistry and/or ions under aqueous conditions, we then summarize how solvation can be modeled with quantum chemistry schemes using implicit, mixed implicit/explicit, and fully explicit solvation modeling. We will discuss the

Example

Perspective: Presents author's viewpoint on a specific topic or issue

- Discusses potential impacts of new technologies or approaches
- Often addresses controversial or emerging areas
- May propose new research directions or hypotheses
- Not peer-reviewed as rigorously as primary research



Volume 152, Issue 13
7 April 2020

‹ Previous Article Next Article ›

Article Contents

I. INTRODUCTION

II. CHALLENGES IN MODELING MIXED SOLVENTS

A. Phase Potential

RESEARCH ARTICLE | APRIL 01 2020

First-principles modeling of chemistry in mixed solvents: Where to go from here?

Alex M. Maldonado ; Yasemin Basdogan ; Joshua T. Berryman ; Susan B. Rempe ; John A. Keith

Check for updates

+ Author & Article Information

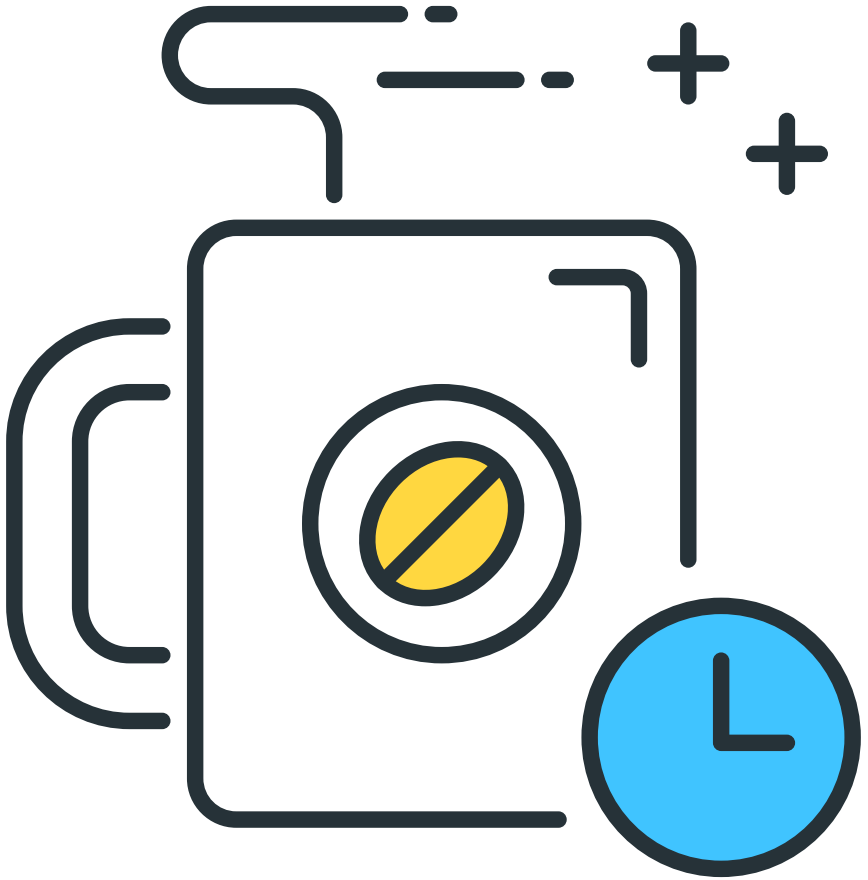
J. Chem. Phys. 152, 130902 (2020)
https://doi.org/10.1063/1.5143207 **Article history**

Split-Screen Views PDF Share Tools

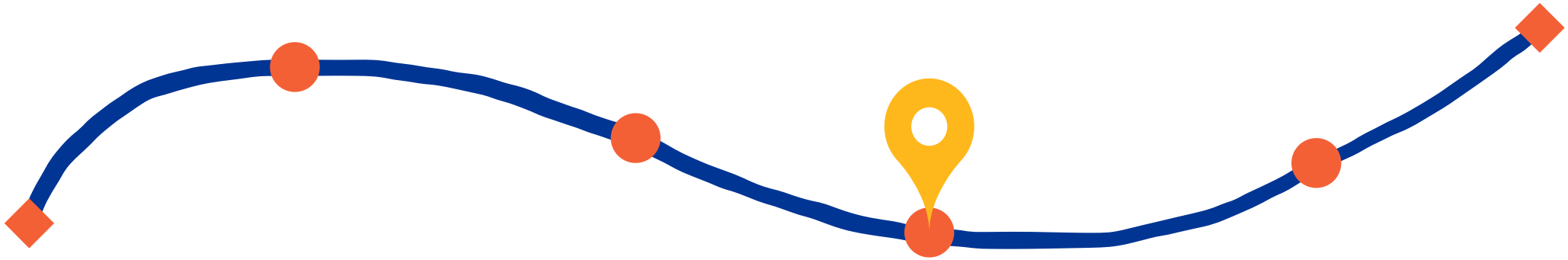
Mixed solvents (i.e., binary or higher order mixtures of ionic or nonionic liquids) play crucial roles in chemical syntheses, separations, and electrochemical devices because they can be tuned for specific reactions and applications. Apart from fully explicit solvation treatments that can be difficult to parameterize or computationally expensive, there is currently no well-established first-principles regimen for reliably modeling atomic-scale chemistry in mixed solvent environments. We offer our perspective on how this process could be achieved in the near future as mixed solvent systems become more explored using theoretical and computational chemistry. We first outline what makes mixed solvent systems far more complex compared to single-component solvents. An overview of current and promising techniques for modeling mixed solvent environments is provided. We focus on

Example

Ten-minute break



After today, you should be able to



1. Understand course structure and expectations.
2. Outline the requirements and goals of the perspective primers.
3. Compare and contrast different types of scientific articles.
4. **Explain the key components of the research ecosystem.**
5. Apply effective strategies to find relevant literature.

Understanding the research ecosystem sheds light on literature caveats

Carnegie classifications of Universitites

Doctorate-granting universities

R1: Very high research activity

- ~ 3% of universities
- Destination of major federal and private funding

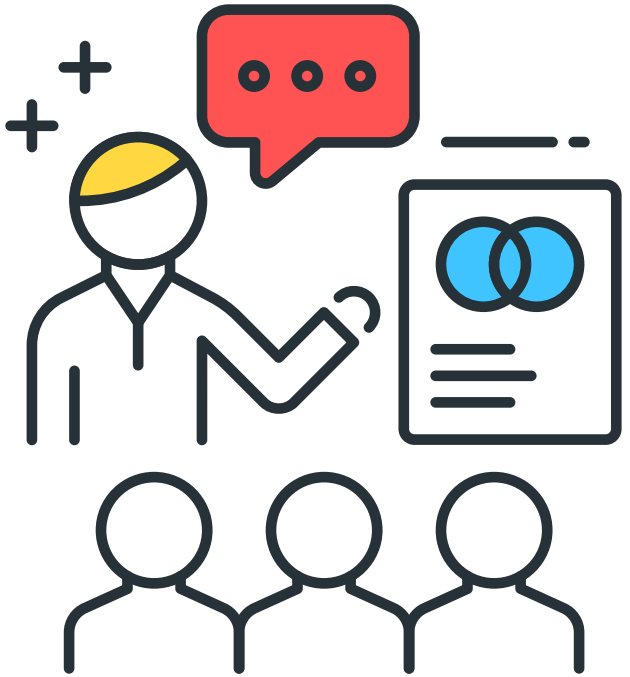
R2: High research activity

Basically how much money a university spends on research

Hiring tenure-stream professors

Application process

- Often 300+ applications for a single position
- Comprehensive application package
- Multiple phone/video interviews
- Campus visits with seminars, chalk talks, meetings, etc.



Selection criteria

- High performance in research
- Grant potential
- Fit with department research priorities and culture

Tenure-stream expectations

- Independently publish by year three
- Mentor research trainees
- Invited to speak at institutions and conferences
- Provide service to department and profession
- Develop one undergraduate course

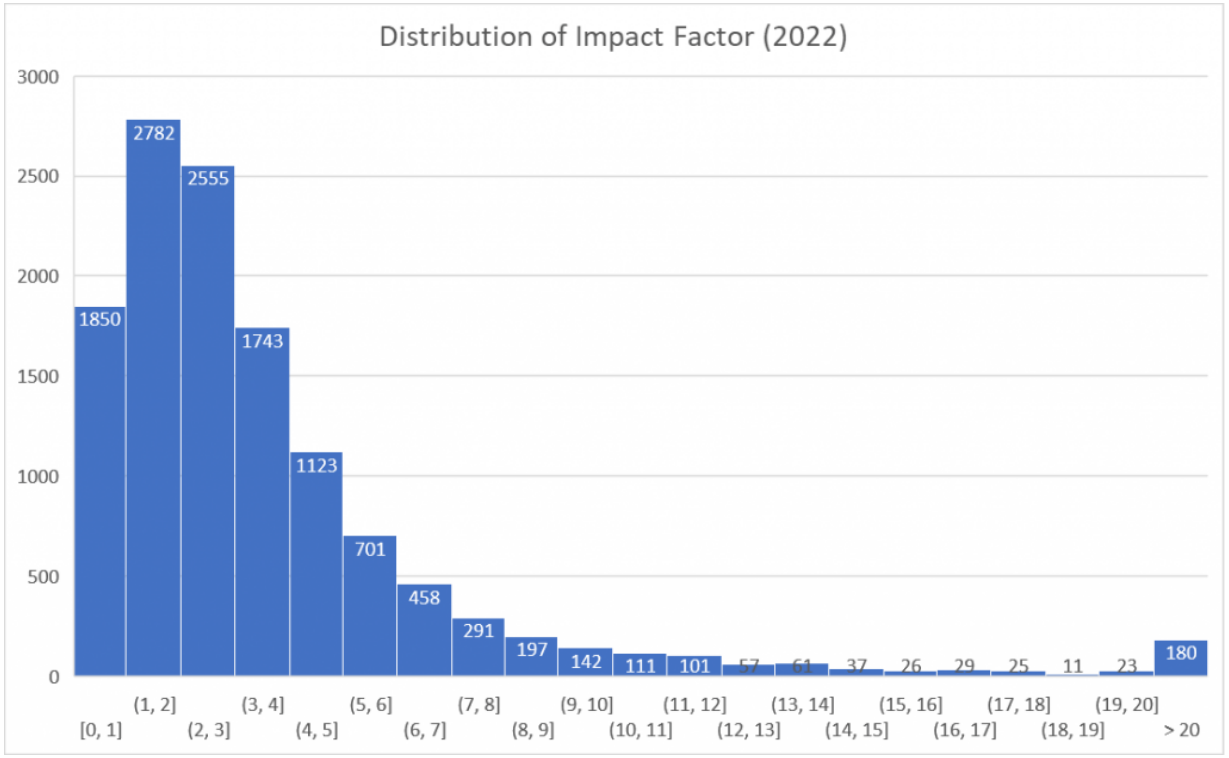
Getting tenure

- Quality, impact, and consistency of research
- Number, size, and prestige of funding sources
- Teaching excellence not required
- Faculty and dean vote; provost has final say

Tenure catalyzes a publish or perish culture

Journal prestige

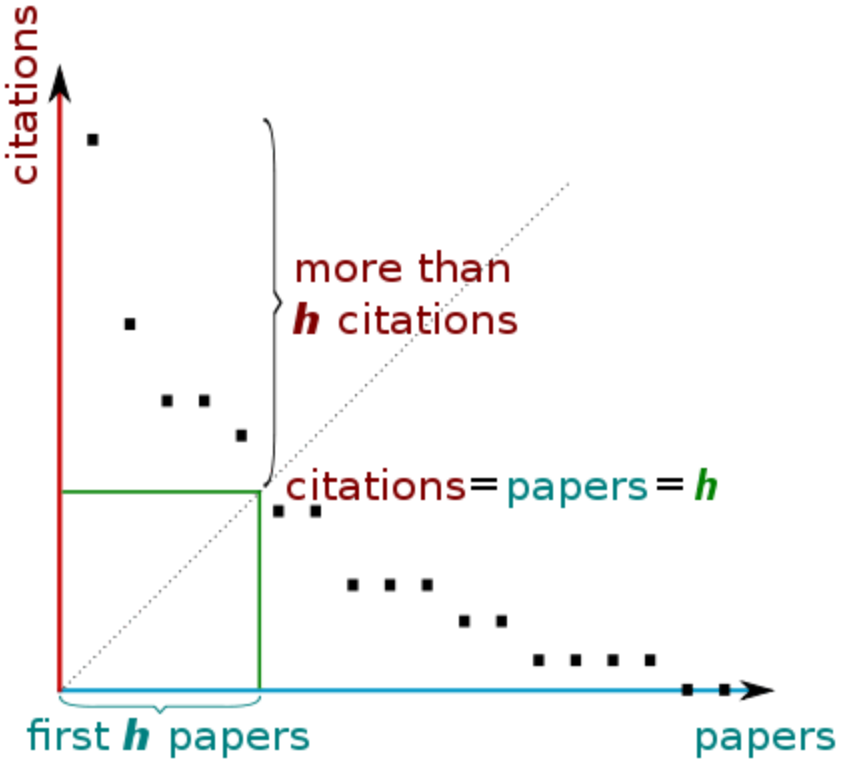
Impact factor tries to communicate how "groundbreaking" publications in that journal are

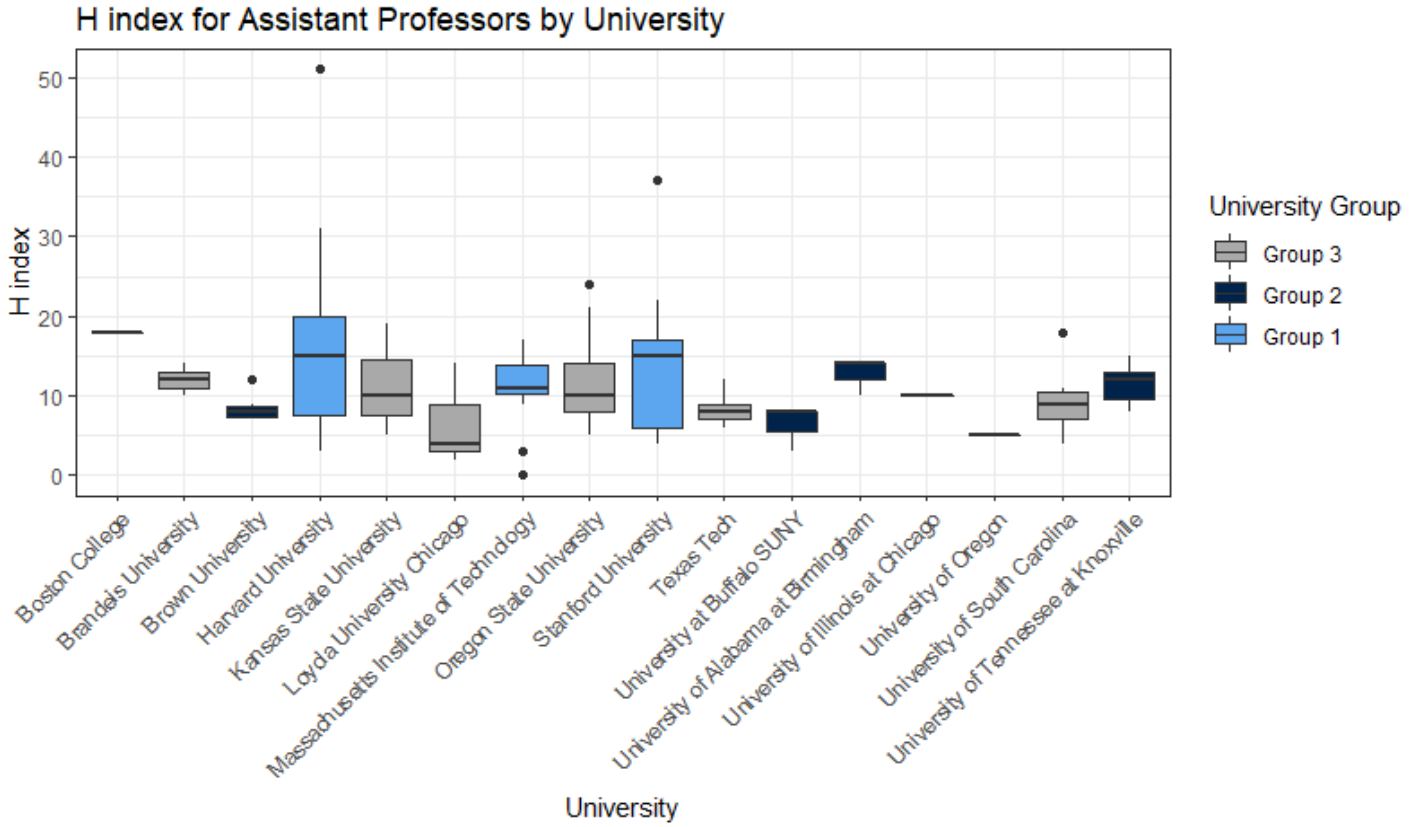


Example journals

Research productivity

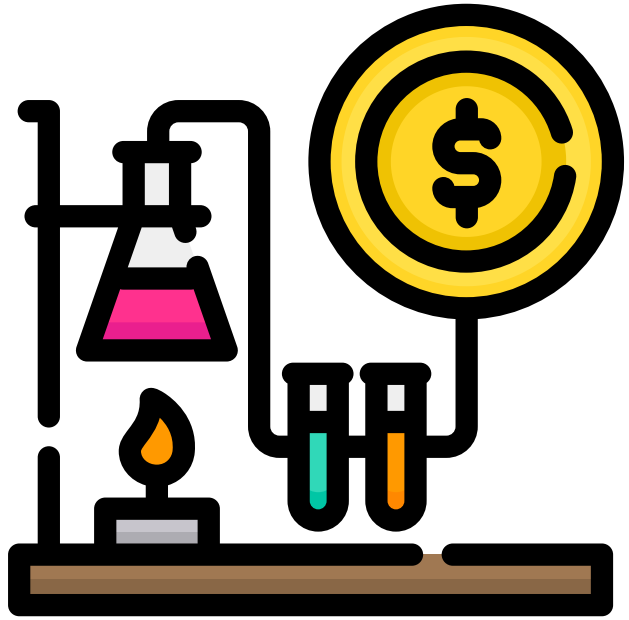
H-index is a quick, imperfect way to
quantify research quality of a professor





Grant acquisition

- Professors pay (and train) others to do their research
- Provides funding for research projects and lab personnel
- Major Funding Sources:
  - **Federal:** NIH, NSF, DOE, DOD, NASA
  - **Private foundations:** Gates Foundation, Howard Hughes Medical Institute
  - Industry partnerships

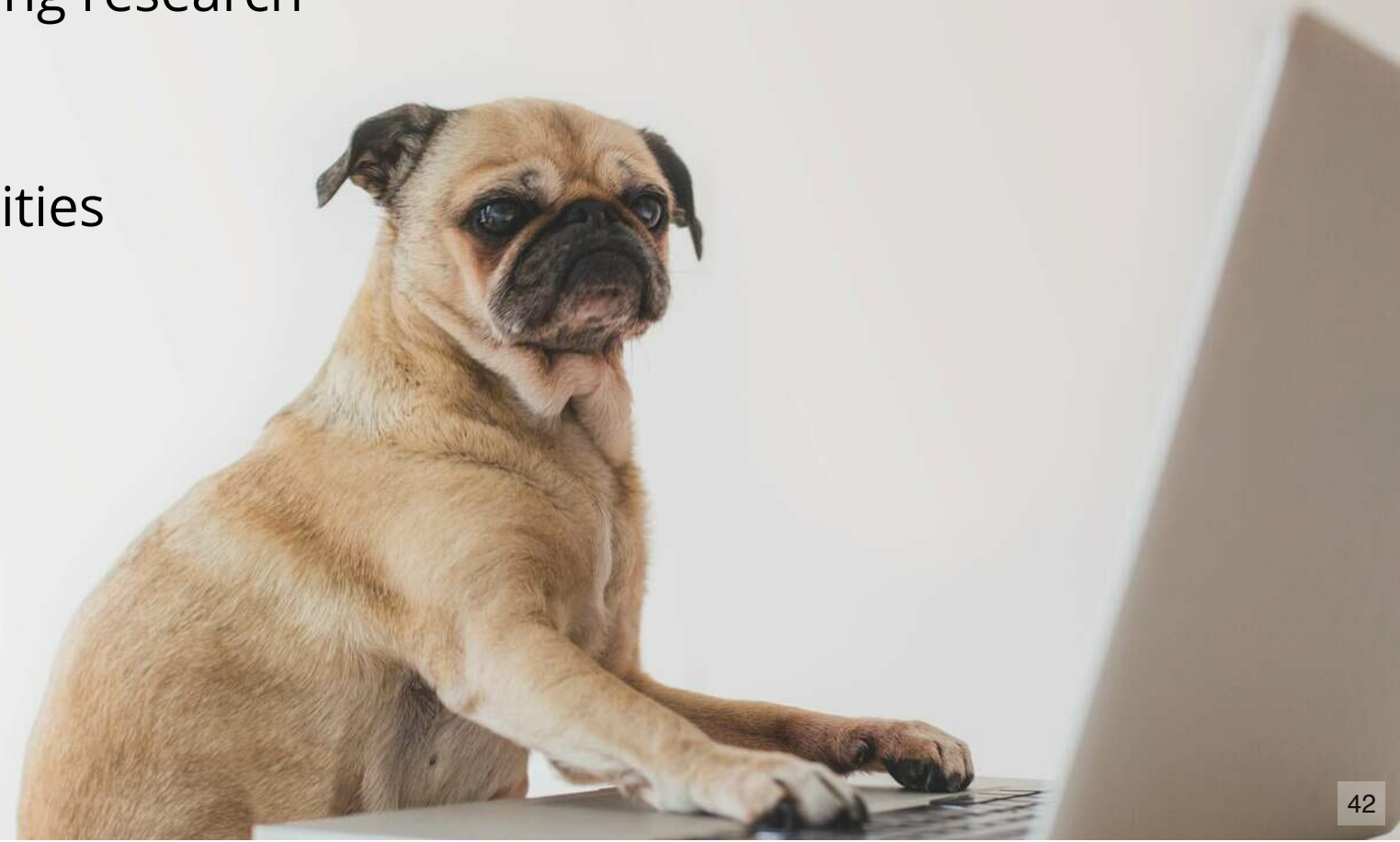


Grant databases

- NIH Reporter
- NSF Award Search
- Granteome

Obtaining grants is extremely difficult

- Time-consuming application process
- Balancing grant writing with ongoing research and teaching
- Success rates often below 20%
- Navigating changing funding priorities
- Multiple submission rounds



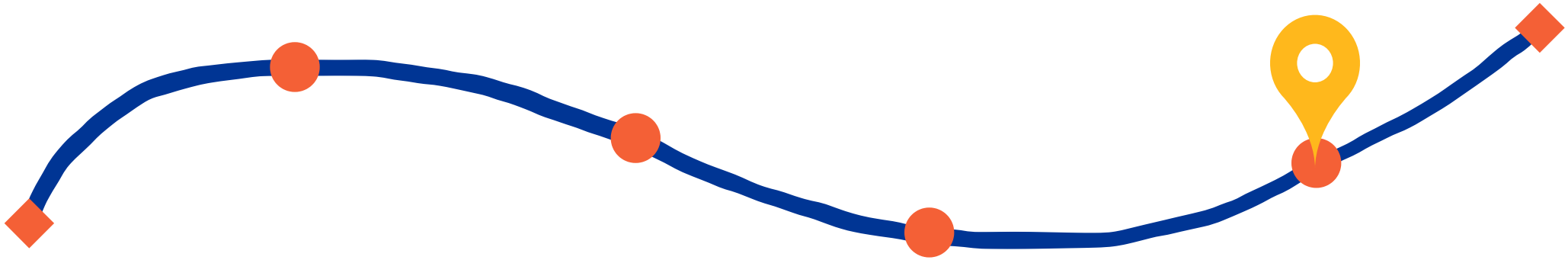
Critical evaluation

Just because it is peer-reviewed does not mean it is perfect

Things you should be considerate of

- Do the authors have the appropriate expertise?
- Are the methods state-of-the-art or justified?
- Is the journal reputable?
- Do the results support their conclusions?
- Are the results reproducible?

After today, you should be able to



1. Understand course structure and expectations.
2. Outline the requirements and goals of the perspective primers.
3. Compare and contrast different types of scientific articles.
4. Explain the key components of the research ecosystem.
5. **Apply effective strategies to find relevant literature.**

Literature search activity

Before the next class, you should

Lecture 01:
Course overview

Lecture 02:
Reading literature

Today

Next Wednesday

- Turn in the "Theme analysis" assignment (will be released shortly)