



Image: IBM 7094, originally built for missile defence systems and repurposed to develop the first protein sequencing program. Courtesy of NASA. Article on page 11.

~/BINFsights/issue_02/contents

SOCIETY NEWS

LOOKBACK

--

Peer Mentoring
Subcom Induction

BIOINFORMATICS DEBUGGED: CAREER PANEL

--

Ajinkya Joshi
Jack Zhao
Sara Ballouz
Sarah Kummerfeld

> FEATURED

--

International Women's Day:
The Mother of Bioinformatics

Amino acid of the week

-

ATTGCCTTTAGATUNSW
ACTBIOINFORMATICS
GCCGATCGASOCIETY_



Peer Mentoring.

Mentors and mentees from BINFSOC's Peer Mentoring Program gathered for a casual 'Paint and Sip' on Saturday. As the program's first social session, the evening was an excellent way for everyone to introduce each other by relaxing and getting creative. Throughout the event each participant attempted to recreate an artwork using paints or digital software, whilst enjoying their favourite beverage.

Photos on next page

Subcom.

BINFSOC's Subcommittee team has now been chosen. Having received many applications, the selection process was extremely competitive with a finite number of spaces to fill. As a result of the high calibre of responses, we had a very difficult interview process and we wish to extend our gratitude to all the candidates who applied.

For our recently chosen subcom team, we held an induction day on campus where our President and Vice President outlined the vision for BINFSOC's future, how we will achieve it, and the expectations and responsibilities of Subcom members. Our 11 new Subcom members span across our three portfolios of Digital Branding (4), Event Planning (4) and Human Resources (3), with one executive leading each team.

The induction was an invaluable way for the 2021 BINFSOC team to get to know each other whilst establishing an atmosphere where collaboration and new ideas are encouraged. We can't wait to see what's next for the 2021 BINFSOC team!

Photos on next page



BINFSOC

Peer Mentoring Program.

Social Session: Paint and Sip.



Subcom Induction. 2021 Team.



The 2021 BINFSOC team (minus three who couldn't make it to the day)



Digital Branding



Human Resources



Events Planning



BINFSOC

Bioinformatics Debugged.

Careers Panel.



THU 11 MARCH. 5PM - 6:30PM.

JOIN VIA ZOOM.

> unswbinfsoc.com/bioinformatics-debugged/

A LOT OF US PURSUE bioinformatics with a limited idea about what to expect. BINFSOC presents 'Bioinformatics Debugged', a career panel discussion for students and potential bioinformaticians. Attendees will have the opportunity to interact with a range of professionals who have studied bioinformatics and are now leading researchers in the field, have gone on to pursue a career in other areas such as software engineering and data analytics, and have experience from tech giants such as Microsoft, Google and Amazon.

Come along and answer all your burning questions about where your degree can take you whilst gaining insightful advice from industry leaders.

Note: no experience studying bioinformatics required, everyone is welcome.



speakers on next page





Introducing **Ajinkya Joshi.**



Current role?

Software Engineer @ Microsoft working on Azure Bot Service.

Favourite bioinformatics tool?

Hmm, I haven't used a bioinformatics tool in 12 years. So i'll say Sublime text editor.

Advice for students?

Even though Bioinformatics is a more structured degree than say CompSci or Biology, there is still flexibility in 3rd/4th year. Focus on what you enjoy, eliminate what you dislike!



Introducing **Jack Zhao.**



Current role?

I'm the co-founder and director of Small Multiples, a data consultancy based in Sydney. I create data visualisation projects to inform and educate people about societal trends and issues.

Favourite bioinformatics tool?

SciPy: Although not strictly for but widely used in bioinformatics, I use its core components, such as pandas, NumPy and Matplotlib regularly in my data visualisation work.

Advice for students?

The knowledge you acquire from bioinformatics will broaden your horizon. Look outside the field of bioinformatics, be a multidisciplinary generalist, this is a good way to figure out your calling.



Introducing **Sara Ballouz.**

Current role?

I'm a group leader in the Garvan-Weizmann Centre for Cellular Genomics at the Garvan.

Favourite bioinformatics tool?

My favorite bioinformatics software has to be the STAR alignment tool.

Advice for students?

R is awesome and that extra TB will never be enough.



Introducing **Sarah Kummerfeld.**

Current role?

Head Kinghorn Centre for Clinical Genomics (Scientific) at the Garvan Institute of Medical Research. My previous role was in industry as a Scientist at Genentech in South San Francisco.

Favourite bioinformatics tool?

Hmmm this is a hard one, there are so many good choices. As someone who has spends a lot of type analysing data to understand disease heterogeneity, I'd have to say all of Bioconductor (the R bioinformatics libraries). If you pin me down to a single software package, I find Salmon to be really useful and when released a new way of thinking about alignment.

Advice for students?

Seek out opportunities to apply what you are learning as early and often as possible even if it means cold calling research groups offering your time as a volunteer.

the mother of bioinformatics

IT IS TEMPTING TO BELIEVE THAT THE PRINCIPLES, METHODS, AND TOOLS USED FOR COMPUTATIONAL BIOLOGY THAT WE TAKE FOR GRANTED ARE AS MODERN AS THE COMPUTERS THAT WE USE THEM ON. BUT THEIR ORIGIN CAN BE TRACED BACK TO ONE PARTICULAR WOMAN WHO PIONEERED BIOINFORMATICS.

Writers Anthony Nguyen, Cam McMenamie

Women in BINF: Margaret Dayhoff

1925 - 1983



Each protein sequence that is established, each evolutionary mechanism that is illuminated, each major innovation in phylogenetic history that is revealed will improve our understanding of the history of life.

— Dr. Margaret Oakley Dayhoff,
Computer Analysis of Protein Evolution 1969

#ChooseToChallenge this year reminds us that we can choose to seek out and celebrate the achievements of women. In 1965, the very first database for protein sequences was compiled and printed by American physical chemist Dr. Margaret Oakley Dayhoff. Fast forward half a century and you are very likely to find databases based on her work being used day-to-day by both bioinformaticians and scientists alike.

Dayhoff, arguably the pioneer of the bioinformatics field, was responsible for interlacing biochemistry with the rapid

advances in computing present during the Cold War. In doing so, she paved the way for new methods and technologies to be developed in order to tackle tough biological puzzles -- methods that have evolved into the modern bioinformatics tools that we are familiar with today.

Born in Philadelphia in 1925, Dayhoff gained an education in mathematics while attending the Washington Square College in New York, and after three years of study, would go on to also pursue a doctorate in quantum chemistry at Columbia University.



Figure 1 Dayhoff with her daughter, Ruth.

This would place her at the very forefront of computing at the time, particularly IBM's Watson Scientific Computer Laboratory. The Watson Lab played an important role in the development of aerospace defence technology, and would later be involved with nuclear missile systems and used by NASA in the Apollo missions.

Driven by an interest in using computers to improve the speed and accuracy of the scientific calculations in her work, Dayhoff began an uphill journey to lay the seeds for the tools and techniques that we all recognise today in bioinformatics. In her position at the National Biomedical Research Foundation, Dayhoff addressed the problem of comparing and piecing together small peptide sequences to form a complete protein. Using IBM's new

7090 data processing machine, six times more powerful than previous models, she developed the first protein sequence alignment program named "COMPROTEIN", which she published in 1962.

Dayhoff also created the one-letter code used to describe amino acids as they appear in protein sequences. In the era of punch-cards and bulky vacuum-tube computers, the hardware used for piecing together these protein sequences was limited, and any opportunity to increase efficiency was valuable. In an effort to reduce file size out of necessity, Dayhoff developed the one-letter code to which biochemists still refer to this very day.

It is important to note that digital computing and statistics have not always been appreciated in biology. Dayhoff's work with her colleague, Robert Ledley, at the National Biomedical Research Foundation challenged many scientific "purists" to speak out against the move to incorporate new computer-based



Figure 2 IBM 7090 mainframe used by Dayhoff to develop COMPROTEIN. Courtesy of NASA.

methods into the realm of biology. Despite being an outsider to the challenging culture surrounding her, Dayhoff continued her push to show the true potential of computerised models in answering questions within the life sciences.

Her next step was compiling a list of all known protein sequences at the time, and documenting their variations across species. This work would be published in her “Atlas of Protein Sequence and Structure” in 1965. Later editions were moved to magnetic tape, and the legacy she started would continue through projects including the Protein Data Bank and GenBank. Notably, the reconstruction of phylogenetic trees to describe evolutionary history was made possible through her method of computational sequence comparison.

Dayhoff’s approach to computing and biology has allowed the expansion of

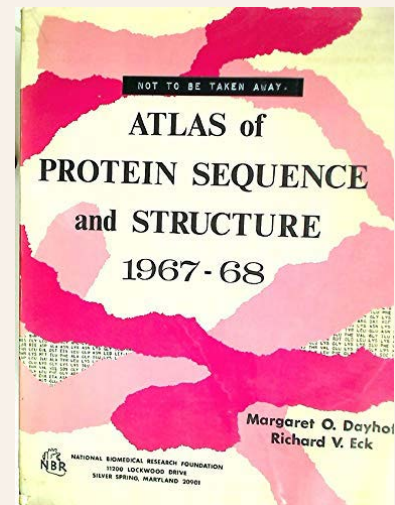
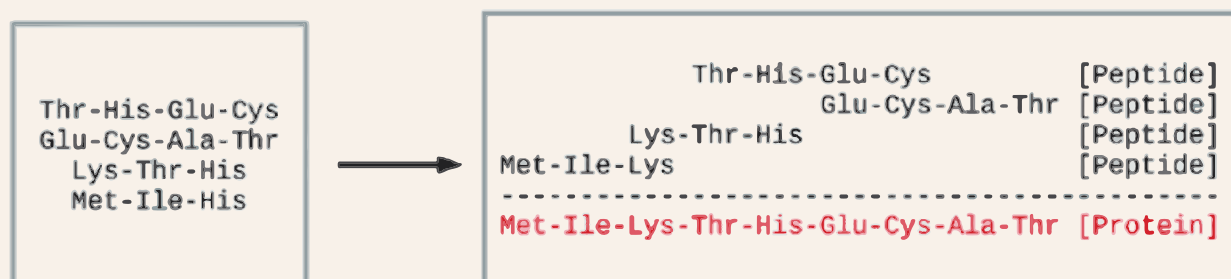


Figure 3 Dayhoff’s publication -- the first protein database.

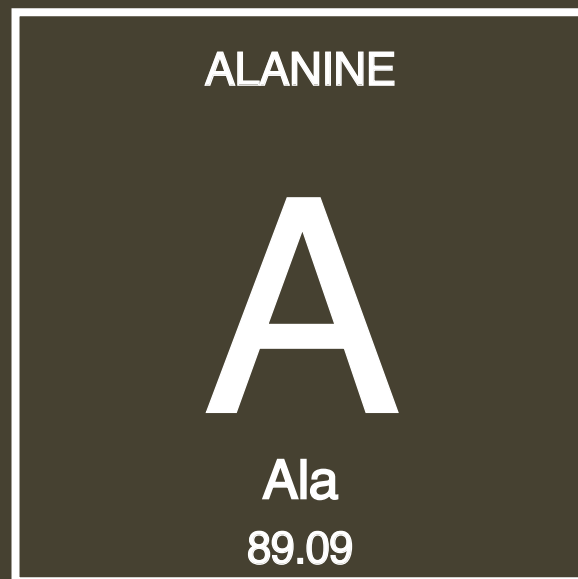
biomedical research to where it is today. Her contributions to bioinformatics are humbling, and are responsible for giving us the tools and methods to continue asking computational questions to answer biological questions -- from the proteins in our blood, peptide sequences in our atmosphere, and the origin of life itself.

COMPROTEIN

Sample input / output



AMINO ACID OF THE WEEK



DNA CODONS:

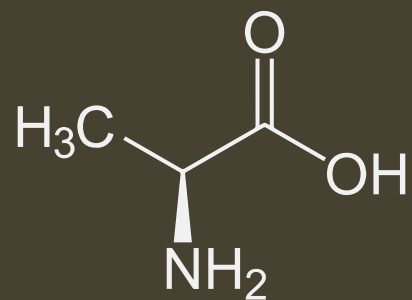
G C T

G C C

G C A

G C G

CHEMICAL STRUCTURE:



USES:

BUILDING BLOCK FOR PROTEINS
[ONLY L-STEREISOIMER USED]

DUE TO ITS LOW REACTIVITY, IT
CAN FORM SIMPLE STRUCTURES.
THIS ALLOWS IT TO FORM 30% OF
THE RESIDUE FOUND IN SILK.