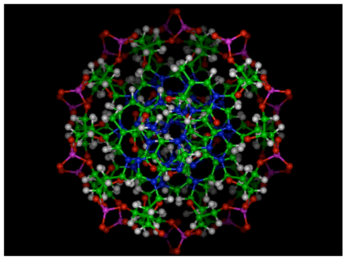
Class 34: Computing with Life (and the Chicken Flu)

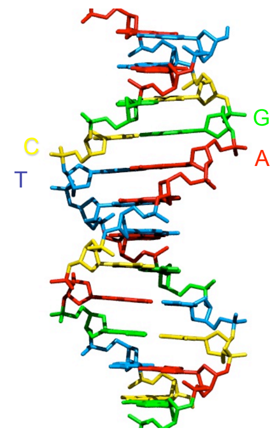


CS150: Computer Science
University of Virginia
Computer Science

Greg Humphreys
http://www.cs.virginia.edu/~humper

DNA

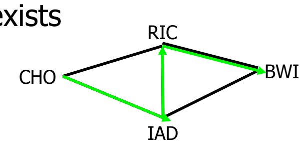
- Sequence of nucleotides: adenine (A), guanine (G), cytosine (C), and thymine (T)
- Two strands, A must attach to T and G must attach to C



Hamiltonian Path Problem

- Input: a graph, start vertex and end vertex
- Output: either a path from start to end that touches each vertex in the graph exactly once, or false indicating no such path exists

start: CHO
end: BWI



How hard is the Hamiltonian path problem?

Encoding The Graph

- Make up a two random 4-nucleotide sequences for each city:

CHO:	CHO_1 = ACTT	CHO_2 = gcag
RIC:	RIC_1 = TCGG	RIC_2 = actg
IAD:	IAD_1 = GGCT	IAD_2 = atgt
BWI:	BWI_1 = GATC	BWI_2 = tcca

- If there is a link between two cities (A→B), create a nucleotide sequence:

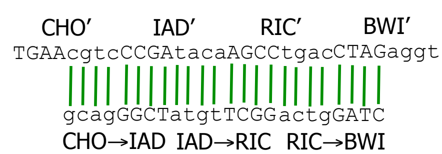
Based on Fred Hapgood's notes on Adelman's talk
http://www.mitre.org/research/nanotech/hapgood_on_dna.html

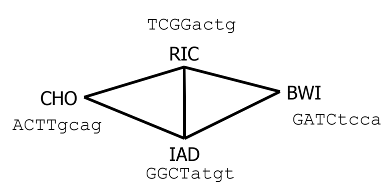
Encoding The Problem

- Each city nucleotide sequence binds with its complement (A ↔ T, G ↔ C) :

CHO:	CHO_1 = ACTT	CHO_2 = gcag
CHO':	TGAA	cgtc
RIC:	TCGGactg	
RIC':	AGCCtgac	
IAD:	GGCTatgt	IAD' = CCGAtaca
BWI:	GATCtcca	BWI' = CTAGaggt

Path Binding





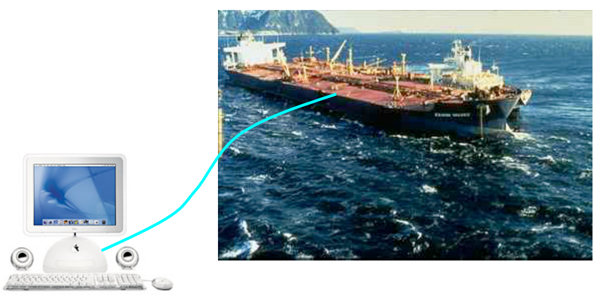
Getting the Solution

- Extract DNA strands starting with CHO and ending with BWI
  - Easy way is to remove all strands that do not start with CHO, and then remove all strands that do not end with BWI
- Measure remaining strands to find ones with the right weight (7 * 8 nucleotides)
- Read the sequence from one of these

Why don't we use DNA computers?

- Speed: shaking up the DNA strands does 10^{14} operations per second ($400M supercomputer does 10^{10})
- Memory: we can store information in DNA at 1 bit per cubic nanometer
- How much DNA would you need?
  - Volume of DNA needed grows exponentially with input size
  - To solve ~45 vertices, you need ~20M gallons

DNA-Enhanced PC



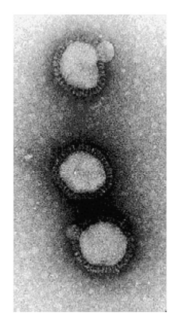
Complexity for DNA Computer

- Complexity classes are different than for regular computers, because a step is different
  - Scales with amount of DNA you have
- Turns time complexity into "DNA-amount" complexity
  - If you have an amount of DNA that is exponential in input size, you can solve NP complete problems in P steps!

Computability for Quantum and DNA computers

- DNA computers: no change to what is computable, only changes time it takes
- Quantum computers:
  - They are so strange they even change what is computable!
  - Quantum physics provides true randomness, something a Turing machine cannot do

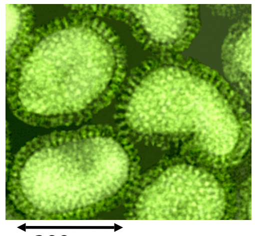
Chicken Flu



Some of these slides are from Dave's USENIX 2004 talk, What Biology Can (and Can't) Teach Us About Security http://www.cs.virginia.edu/evans/usenix04/

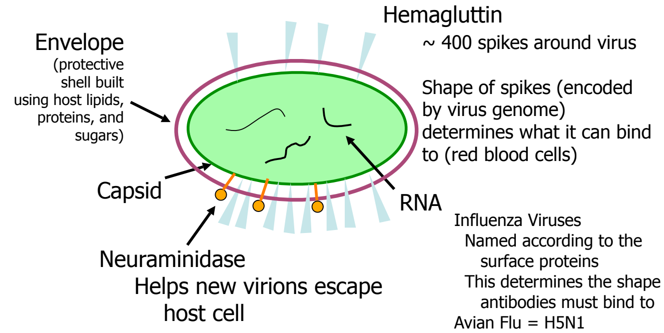
Viruses

- Genetic material (RNA) with protective coat
- Receptor binding proteins attach to cell
- Injects genetic material into cell nucleus
- Uses proteins in cell to reproduce
- Releases copies to infect more cells



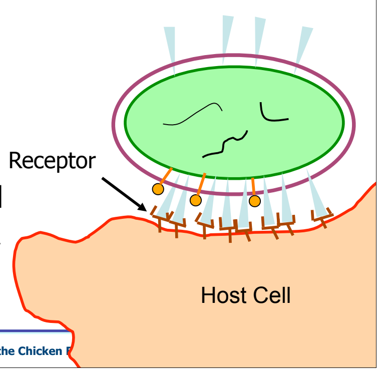
Influenza Virus

Influenza Virion



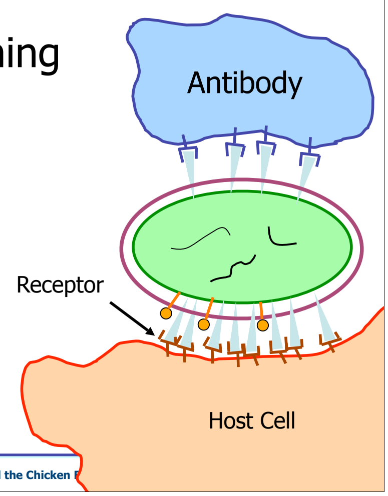
Virus Binding

- To bind to a cell, virus receptor binding proteins must match cell membrane receptors
- Virus is internalized by cell, injects RNA into nucleus



Virus Scanning

Virions have a specific shape to bind to hosts, so scanners (antibodies) can recognize that shape and block virions



Immune Systems vs. Virus Scanners

- Standard Anti-Virus software scans for **known** attacks: compare code against a library of already known attacks
- Approach doesn't work if new viruses emerge quicker than updates
  - Internet Worms: spread time ~ 20 minutes
  - Human: genome updates ~ 20 years

Need a way to detect and defeat previously unknown attacks

Pathogen Diversity

- Genetic drift: random point mutations
  - Some will be successful, and multiply
  - RNA-based viruses mutate very rapidly
- Genetic reassortment: mixing up
  - If two strains of influenza virus infect the same cell, they mix up their genes
  - Influenza A viruses have 8 segments
    - Can mix and match segments
    - If an organism is invaded by a strain that attacks a different species, segments can jump to the new species by combining with other viruses

Receptor Diversity

- Lymphocytes are white blood cells that have surface proteins to recognize intruders; when stimulated by antigen they make antibodies
- Need to recognize all foreign intruders, but DNA can't know about all (~10^{16}) possible intruders
- Gene segments are randomly combined to form different receptors
  - Create 10^7 new lymphocytes every day
  - Lymphocytes that match intruders reproduce quickly (build immunity)
- But, need to ensure lymphocytes don't match self

Recognizing Self

- Major Histocompatibility Complex
  - Surface molecules that are unique to individual on all cells (except red blood cells)
  - Authenticate cell as self
  - Diversity of MHC types protects a population
- Thymus gland
  - Lymphocytes that match self molecules are eliminated, others are mature and enter body

Immune System Disorders

- False negatives are immune deficiencies
- False positives are auto-immune diseases:
  - Reject organ transplants
  - Multiple Sclerosis – motor nerve cells are antigens
  - Rhumatoid Arthritis – connective tissue is antigen

Computer Immunology

- Forrest, Hofmeyr and colleagues, 1994
- Recognize computer intrusions
- Generate library bit-strings that encode patterns of normal behavior (system calls, network connections, etc.)
- Generate random detectors: keep ones that don't match the normal behavior
- Recognize behaviors that are abnormal as possible intrusions

Detecting intrusions is still undecidable... just a probabilistic technique

Racing Parasites

- Parasites evolve quickly:
  - E. Coli bacteria ~ 1 hour per generation
  - Influenza virus
- Offspring should be optimized for a different environment than their parents – parasites have evolved

Matt Ridley, The Red Queen

Achieving Diversity

- Natural selection reduces diversity
  - Will select against inferior genes for particular current environment
- Sex maintains diversity
  - Obtain multiple forms of a gene (AB blood type)
  - Retain currently unfavored genes
  - Opposites attract!
    - Wedekind and Füri found that men and women are attracted to odor of members of opposite sex that have MHC genes most different from themselves

Diversity in Computer Systems

- "A computing monoculture is a danger, a security danger, a national security danger. It is a danger on principle. It is a danger in practice."

  Dan Geer, USENIX Tech 2004
- Microsoft Bashing
  - Client OS (2002): Windows (93%)
  - Client Applications: Office, IE
  - Server OS (2002): Windows (55%), Linux (23%)

A more competitive marketplace might help...but not enough

Not All Bill's Fault

- Protocols: IP (100%), TCP (>90%), IEEE 802.11b/g, Bluetooth
- Firewall/VPN: ISS BlackICE/Real Secure (~20%)
  - Enough for Witty Worm (12,000 victim hosts in ~45 minutes)
- Image processing code: libPNG
  - Same vulnerability may be exploitable in IE and Mozilla on Mac, Windows, Solaris

Human-engineered diversity is not enough

Charge

- If you intend to procreate, make sure to find a partner who smells good
  - Diverse MHCs will give your offspring good immune systems
- Later class: how to make computing systems more diverse