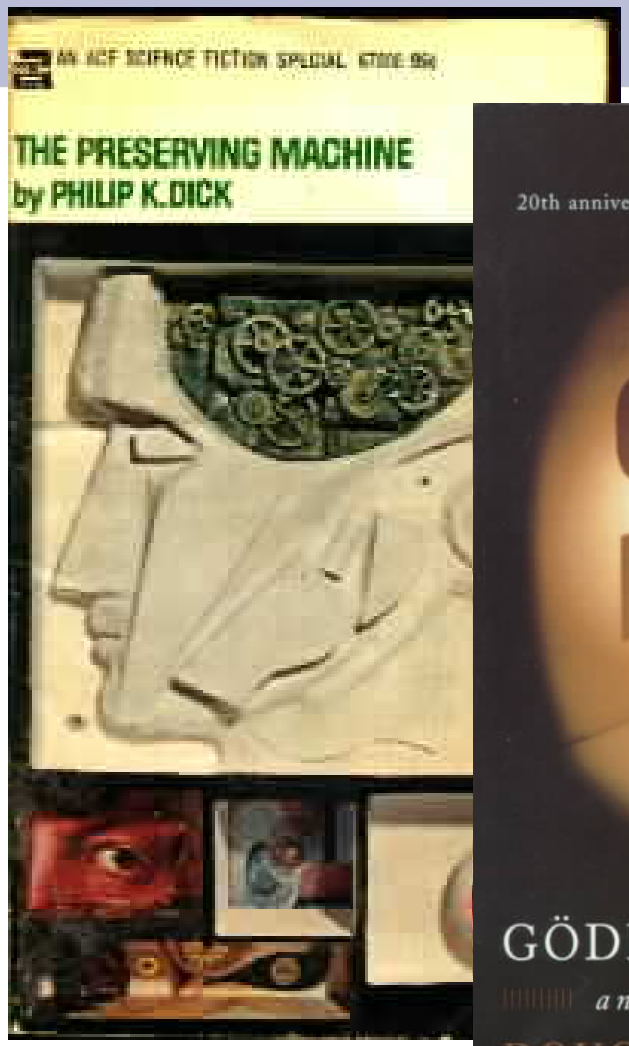
# Auralizing DNA

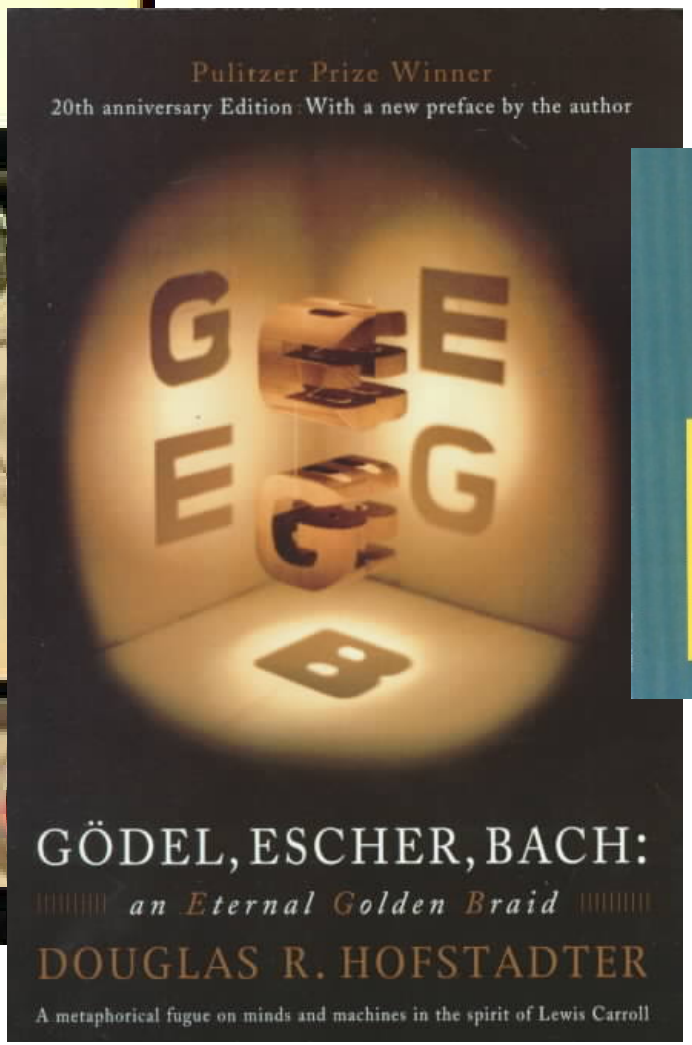
R. Mark Adams, Ph.D.

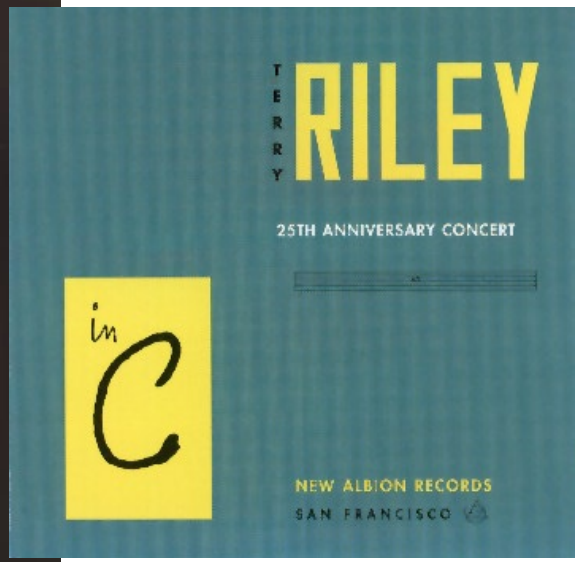
# Presentation Contents

- Inspirations
- Some facts about DNA to know and tell
- A (very) little bit of music theory
- Software and tools
- Putting it all together

# Inspirations



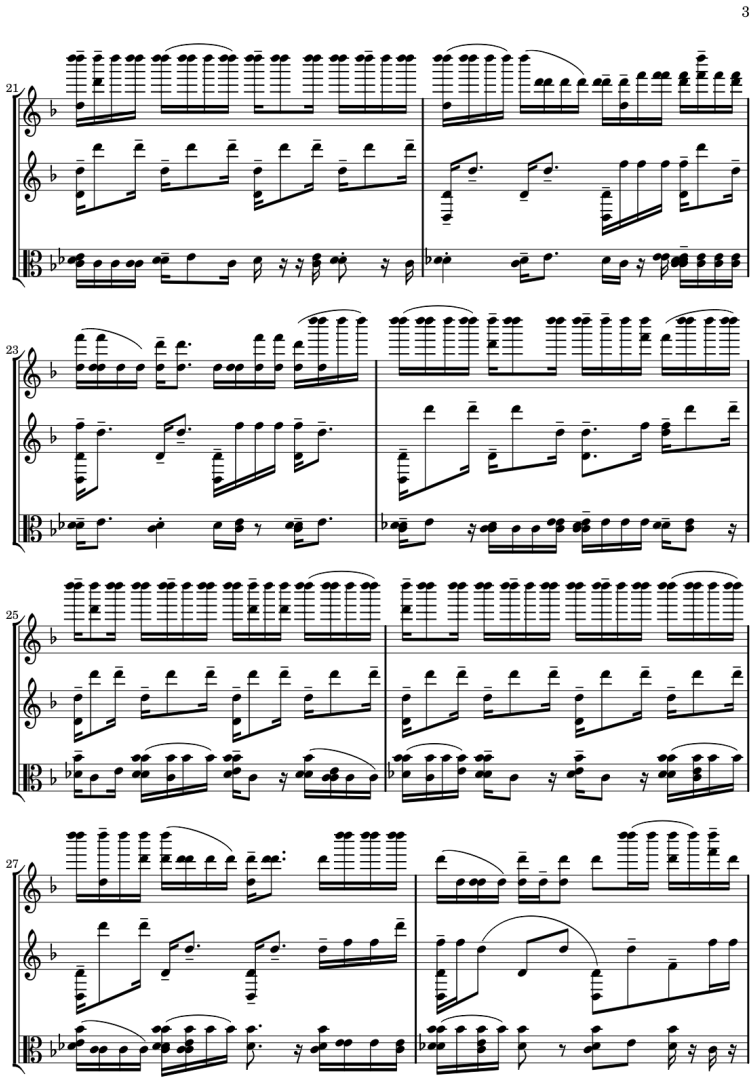

...and many more...

# Your DNA and you...

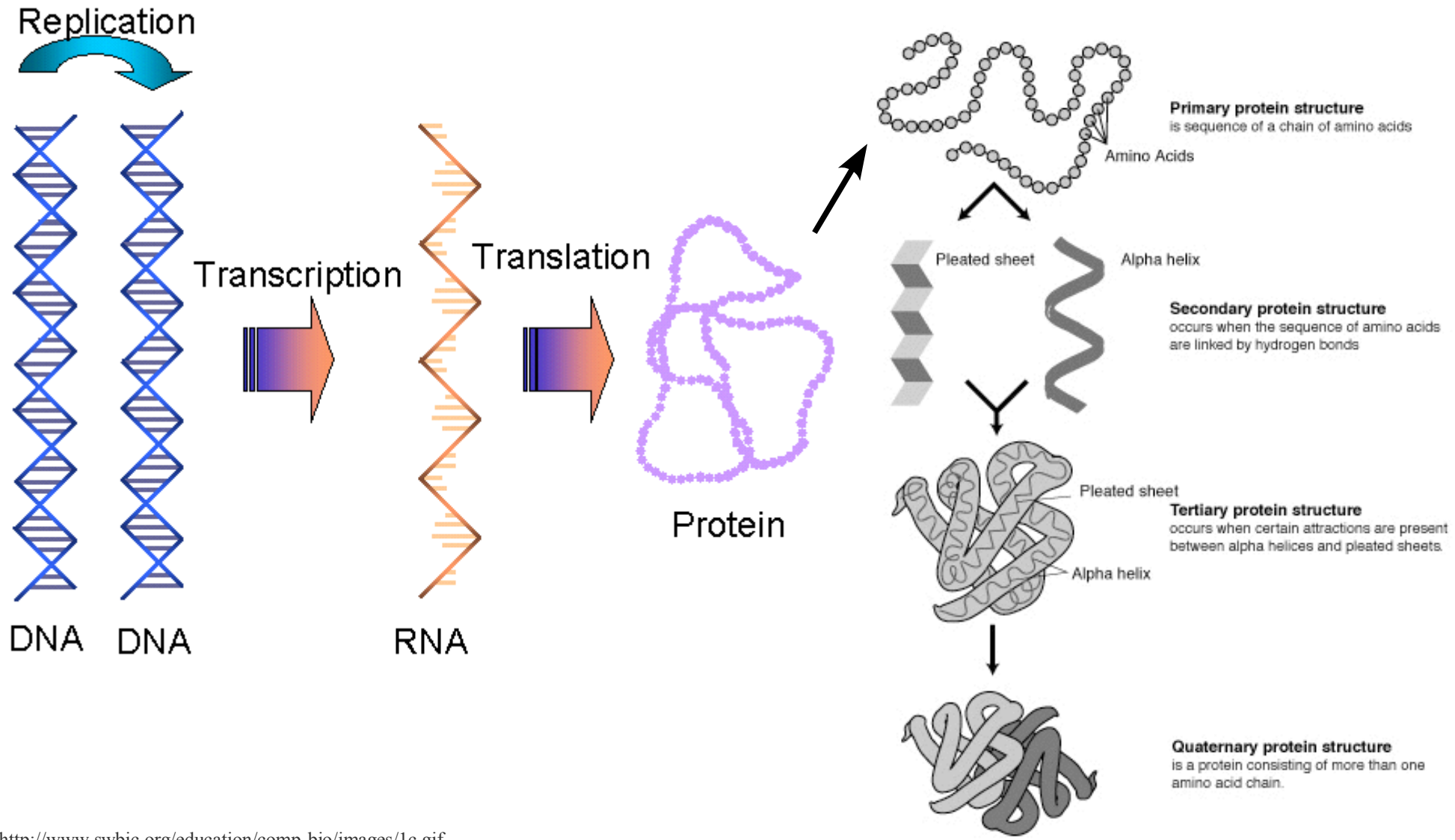

- Your DNA:
  - Contains the entire plan (more or less) for you
  - Compressed, about the same size as a DVD (3.2e9 bases, in 1 ASCII byte per base)
  - Differs about 1 every thousand bases from the next person

# Like music, DNA is a structured and layered string of data...

```
   1 aacatacctg ctttatgcac tcaagcagag aagaaatcca caagtactca ccagcctcct
  61 ggtctgcaga gaagacagaa tcaatatgag cacagcagga aaagtaagca aaaaatatat
 121 tactgttgat aatatattct cttcaatata acaaataaag gaatacagta ttaataatga
 181 ataatttaaa atttaaaatt catatctaaa ttagatgatg aataatttaa aattcatatc
 241 taaattagaa atgatacact ggaatatgat atgcaatata tgctataata tgtaatgtat
 301 actgaactac agtggaaata agctattcct aaataccttc aaaaagaata tatagaatct
 361 gtatctattg tctttatttc ctacattaaa taaatttgcc gtaaagtgat agtttattcc
 421 aagctaatca tgactgattt gtaaacctaa agttagaaaa gtctttaatc agaagctatc
 481 tttatattaa gaaagcataa tttaaatgtg ttattatttg tattcatatt ttttgtgaac
 541 acaagaagtc tgacaaaact tttatgagag ggatttgaga agattttgaa atatactttt
 601 aatcttacta taaaatatct tacaaaatac tcattgatct cacagcatta gaatcatcaa
 661 ggttaagcaa gacatcacat tcaaattccg tttaaagggg gcccattatg acacaattca
 721 ggcaatttcc acagaaatct tatggaacag tatctccccc atataaaagt caatatgatc
 781 ttacagaaaa ataataatgc aatttgaatc acttattagc actcagaaca caaatatttg
 841 ttttttcttc tataaattta tacttatttt tcaatgtgtt tacaggtgca cagaaatgca
 901 tgtggtcatt caatataatc aattgatatt attaattgcc taatttaaaa aaatctgtgc
 961 aactatttcc agccatttgt tgtgctagga gtgtatcaca caataaaaca cctcactatg
1021 ataattcagt ttaaaggttc tgaggcttac ctttatgctg tgcgacaaaa caggctcatg
1081 tcaataagac tggttggaaa tcacatgagt ggcccattgg tactgttctt acaccacttc
1141 actttacttt actttcattc attattgatt aatatttaca ttcctcatag aaaataatta
1201 gaaaaaagaa aatttaaatt taccatttac taaactcgac ttaaaagaaa taatgagttc
1261 atagagcaaa agtataaacc aatcattaat gaaaataata actgatgaaa tagataatcc
1321 tcccctcttg agtgcaacat caataactta gctttttgac agcatttcat ttatgtttac
1381 ccgtcctgca ttttattttc ctcaatccta aattgtgaca atactaatgt ctatttcata
1441 aggtagttgt gagaattcag taaattaata gtgaaaagca cttagaatag tacctggtaa
1501 ataaaaataa gtcaataaat attagccact gttattattg ttgctttata acttttgat
1561 atttactacc acggagtaca gaaaacgtga ggctacatta atttttttcat tcgttttttt
1621 gtttggagat ggagtatctt tctattgccc aggttggagt gcaatggtga tctcggctca
1681 ctgcaacctc tgcctcccgg gttcaagtga ttctcctgcc tcagcctcac aagtagctga
1741 gtttacaggt acacgccacc atgcttggct aattttagca ttttttagtag agacagggtt
1801 tcaccatgtt ggtcaggctg gtctccaact cttgacttca ggtgatccgc ccacctcggc
1861 ctcccaaagt gctgggatta aagggatggg caactgcacc cggcctgatc tttattctct
1921 ggacagccag ctttgagact tcaggaaaat tattcaatca ctgagtcagt tgatcctcaa
1981 ttatttcaga tgtagtaaga ccaataattc aatagtactg tcctggtagc atccgtttta
2041 ggtttaaaag taattcatat tgtttacagc agcacaattt gcaattgcaa aaatatggaa
2101 acttcctaaa tgcccatcaa ccaacgagtg gataaagaga atgtggtata tgtacaccat
2161 gagatactac tcagccataa aatggaacaa aataatggcc tttgcagcca cttggatgaa
2221 gctgaaggcc attattttaa gtgaagtaac tcaggaatgg aaaaccaaat accatatggt
```

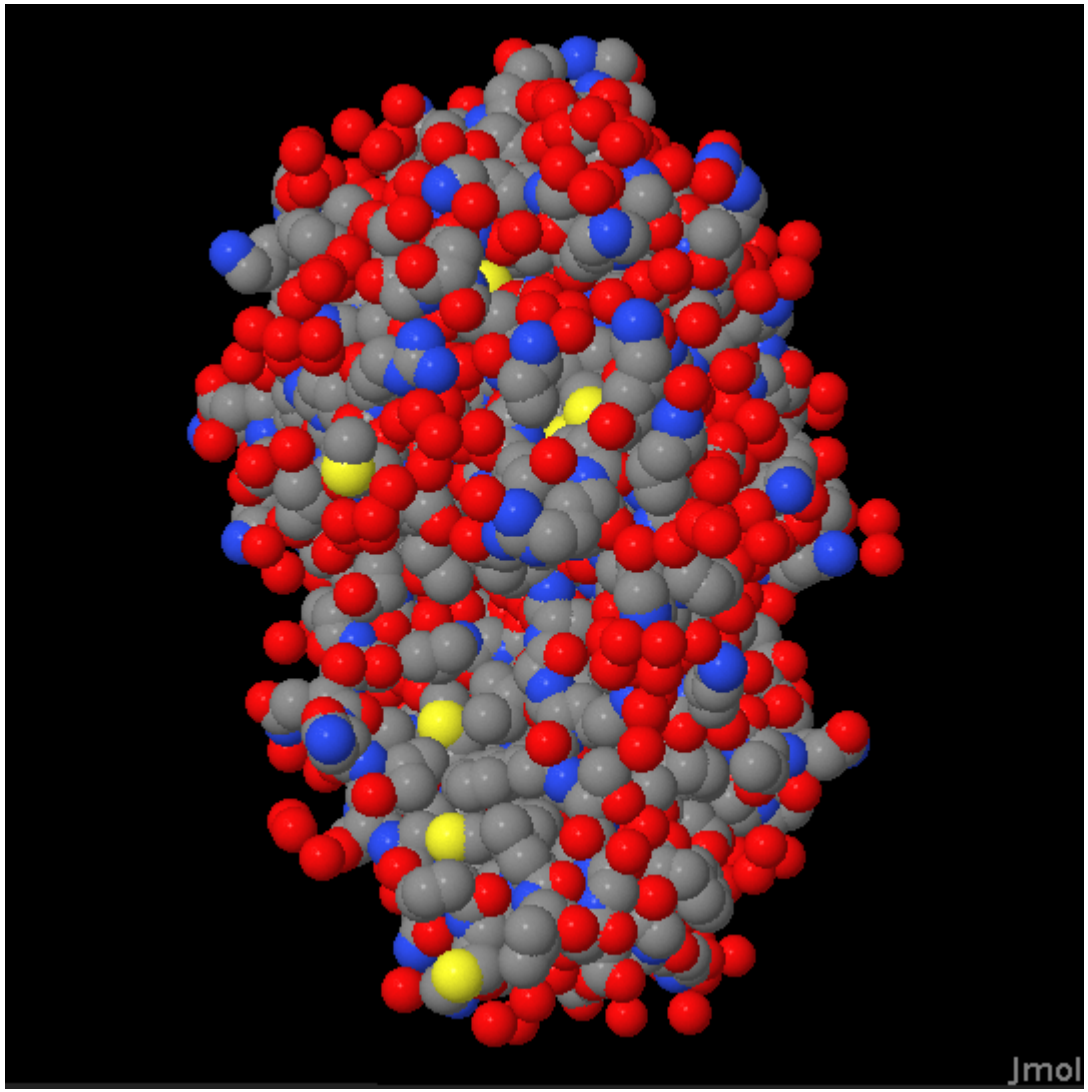


# DNA contains coded information at many layers...



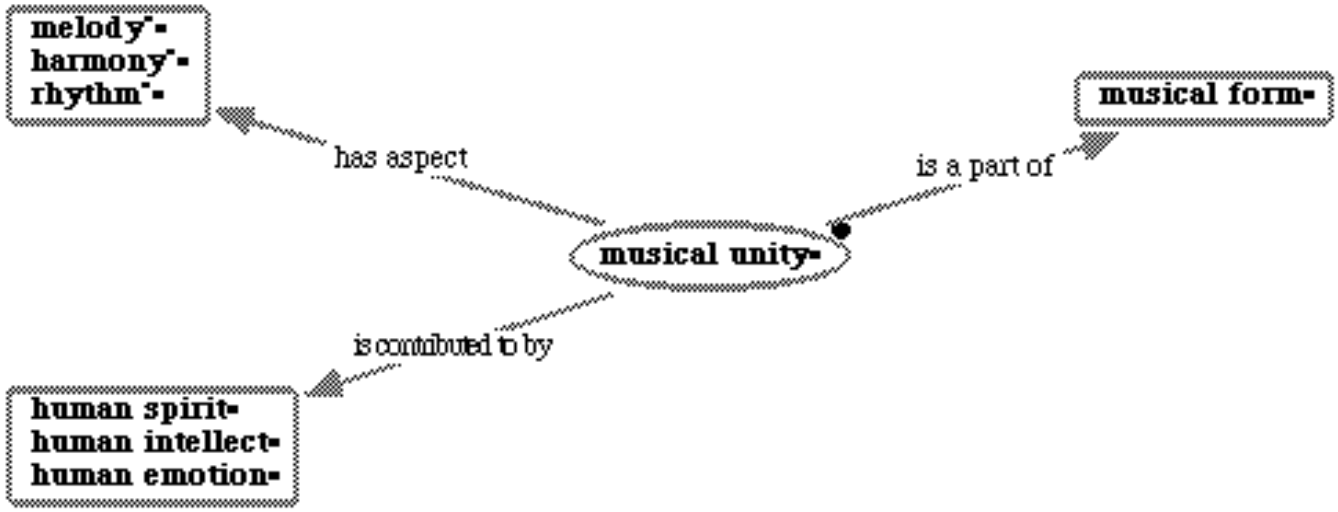

http://www.swbic.org/education/comp-bio/images/1c.gif
http://en.wikipedia.org/wiki/Protein_structure

# ... leading to fully formed and operational structures



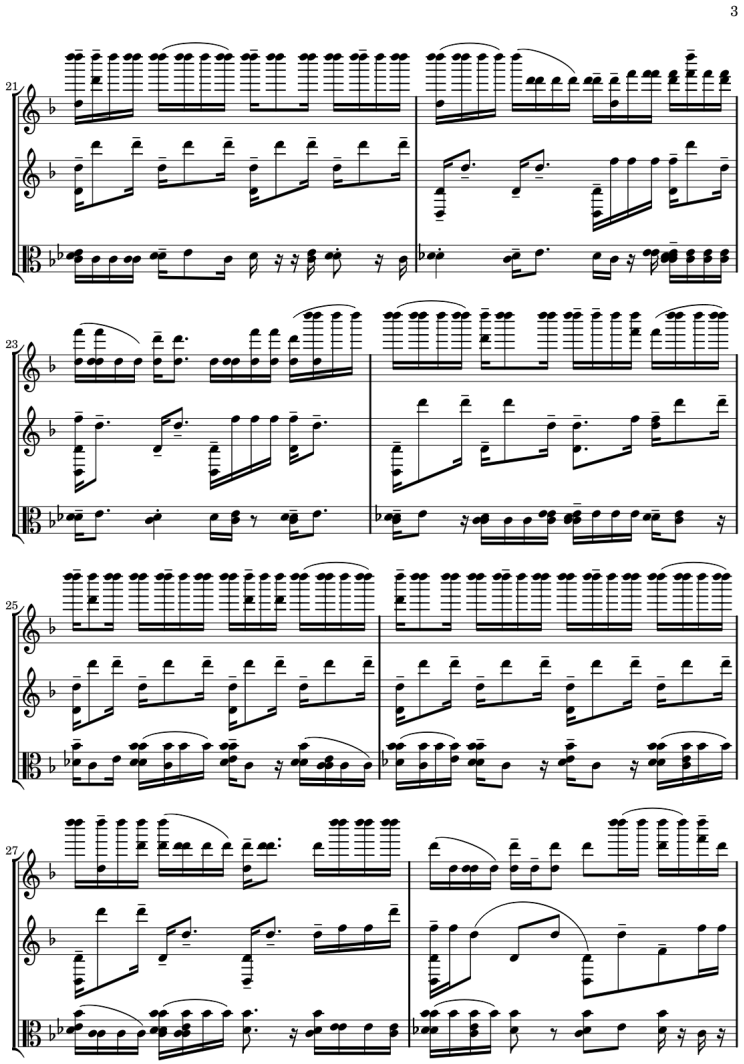
# Music is very similar, where information at many layers...

- Rhythm
- Melody
- Harmony
- *et al.*




http://trumpet.sdsu.edu/M151/SN_10musicalunity.GIF

# Combine to produce finished works...

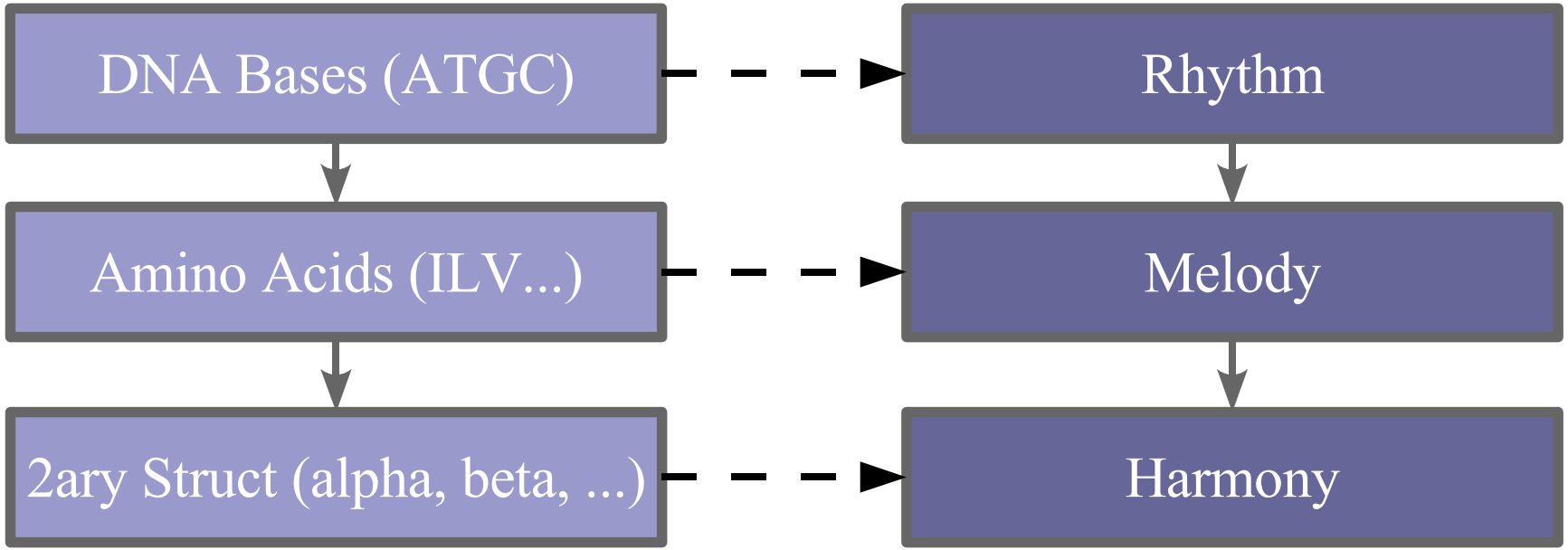


# A simple set of rules

- Almost the simplest possible set of rules yields interesting results
- I used transformation rules derived from “trance” style dance music
- Applied it to two sequences and compared the result


http://mrskingsbioweb.com/Genetics_files/image003.jpg

# Mapping the layers of information in DNA to musical

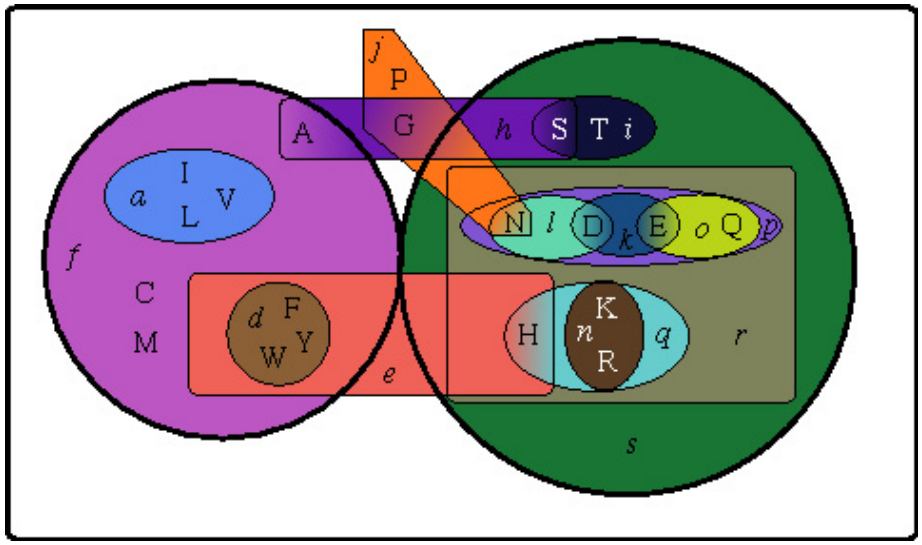


# Rhythm

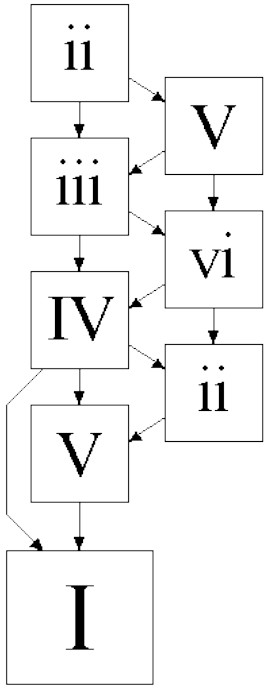
- Directly mapped each base of DNA to eighth notes in middle 'C'
- Added a “break” at starts and ends of protein translation, “*open reading frames*”


<<Play the “base track”>>

# Harmony




The Simple Map




- The harmony is drawn from the structural characteristics of the protein sequence
- Chord changes signaled at amino acid class changes

http://bmerc-www.bu.edu/description/aaclasses.html
http://chordmaps.com/part3.htm

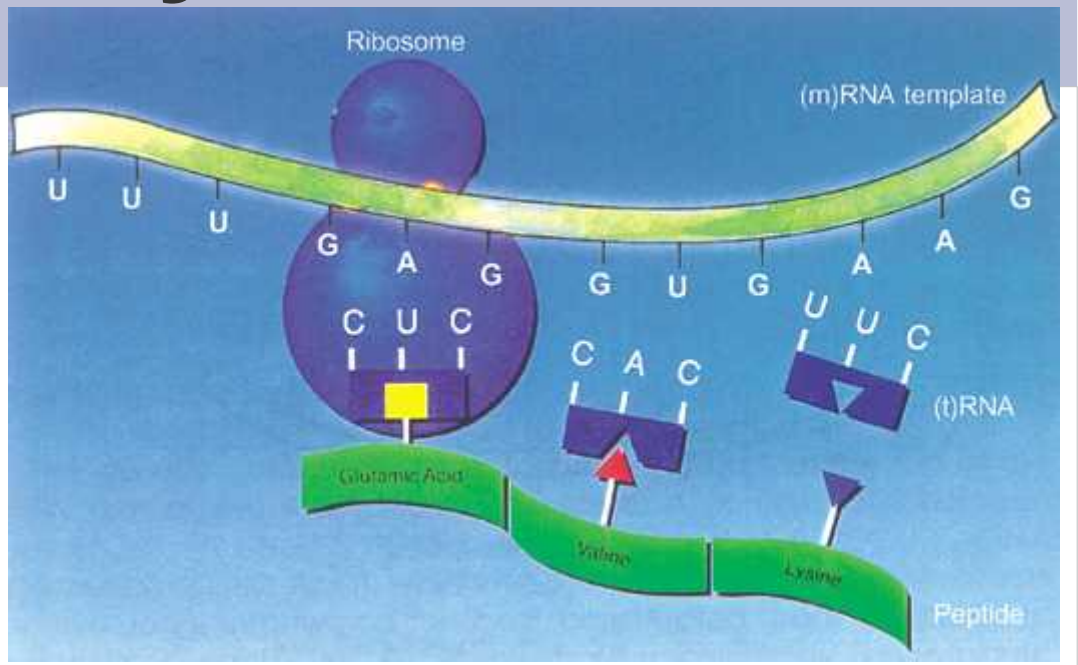
<<Play the "melody track">>

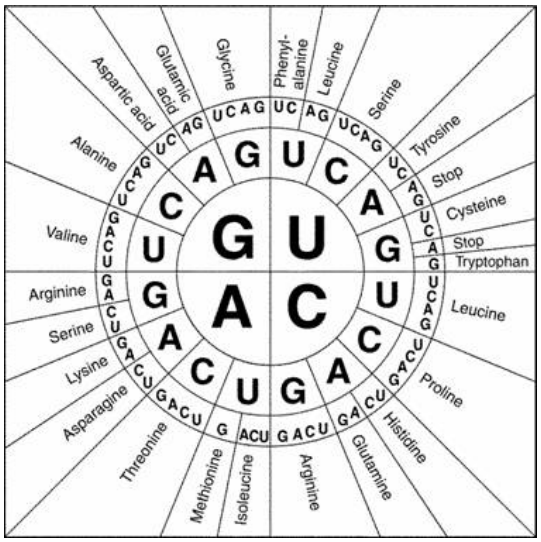
# Melody

- The melody line was derived from the actual DNA bases within each codon

- A>up an interval
- T>down an interval
- G>stay
- C>rest

- Intervals defined by current chord




http://www.littletree.com.au/images/dna10.jpg
http://mrskingsbioweb.com/Genetics_files/image003.jpg

# Results

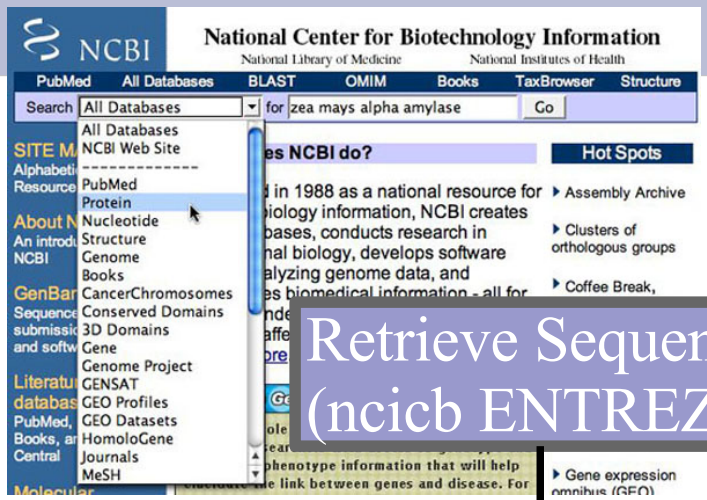
- After generating the musical parts, integrating and orchestrating them, listen to the results and see if they are distinct
- Obviously more complex rules can yield more complex peices

**Sequence #1**: A repetitive sequence that does not code for any proteins. (it is an ancient virus sequence integrated into the human genome)

**Sequence #2**: The coding sequence for Alcohol Dehydrogenase, a gene involved in alcohol detoxification.

<<Play segments of full pieces>>

# Technical Specifics




Retrieve Sequence (ncicb ENTREZ)

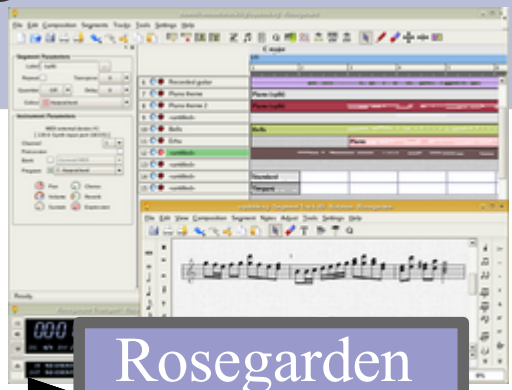
Custom python programs

Python-MIDI

Rhythm MIDI File

Melody MIDI File

Harmony MIDI File



Rosegarden Sequencer


Audacity Recorder

# Acknowledgments


- The tool developers
  - Python
  - Python MIDI
  - Rosegarden
  - Audacity
  - Linux
  - *et al.*
- And all the scientists of the genome project

- (and of course, Dorkbotters everywhere!)

# Resources

- Everything you ever wanted to know, plus the genome:
  - http://www.ncbi.nlm.nih.gov/entrez
- Python:
  - http://www.python.org
- Python MIDI:
  - http://www.mxm.dk/products/public/pythonmidi
- Rosegarden sequencer:
  - http://www.rosegardenmusic.com/
- Audacity audio workstation software:
  - http://audacity.sourceforge.net/
- EMU-Orbit9090 info:
  - http://www.planet-groove.com/e-mu/orbit.html

# Questions?