Eugene Thacker/*Biotech Hobbyist*

BIOMOLECULAR WRITING IN A NUTSHELL

Introduction

This is a how-to manual for various techniques in "biomolecular writing," using only pre-existing, freely-available bioinformatics research tools.

What is bioinformatics?

There are numerous definitions, and each one emphasizes a different aspect of biotech research. Broadly speaking, bioinformatics is an emerging field which involves the integration of computer technologies into molecular biology research. This combination of biology and computers takes many forms however. In many cases bioinformatics simply refers to software tools to aid in biotech research, be it in the study of genes, proteins, or biochemical networks. The most famous example of bioinformatics in action is the human genome projects – both the International Human Genome Sequencing Consortium and Celera Genomics made extensive use of computational tools (both hardware and software) to sequence, assemble, annotate, and archive the entire genomes of organisms. At the root of bioinformatics is a view of biological life through the lens of informatics (genetic and protein codes, genome databases, protein signaling, and so forth).

What does bioinformatics do? Basically bioinformatics is concerned with two threads of research: sequence and structure. By sequence we mean any type of research which is primarily concerned with elucidating and studying biomolecular sequences such as DNA or RNA. In this thread, a research team can take a test sample of DNA which they know nothing about, and compare it to any number of genome databases to see if there is a match or close relationship with any other known gene. In the second thread – structure – we mean any type of research which is primarily concerned with the three-dimensional shape and molecular interactions of biomolecules, such as proteins. Here a research team may want to study how a particular sequence of amino acids folds into a complex 3-D shape, enabling it to act as an enzyme, antibody, or other type of molecule.

What is biomolecular writing? We can think of it first as a practice, and as a means of inquiry into the relationships between DNA and data, proteins and information, biologies and technologies. Biomolecular writing can be thought of as a means of investigating the informatic view of biological life, which is the dominant way in which human genome projects and other like efforts conceive of the body. The notion of writing is already prevalent in molecular biology (translation, transcription, RNA editing, etc.), and we can use literary texts as a kind of hinge-object between genetic "codes" and computer "codes."

>

Text-driven generation of novel protein structures

Our first technique will be the generation of novel protein structures from an input literary text.

1. Sequence preparation: select sample text from a database. Because genes and proteins vary widely in the length of their units, you can experiment with sample text length. We chose Mary Shelley's Frankenstein from the Project Gutenberg database <http://promo.net/pg>, one of the most well-known databases of literary texts.

We took the first paragraph of Chapter 11 as our sample text, comprising approximately 1,122 characters:

It is with considerable difficulty that I remember the original era of my being; all the events of that period appear confused and indistinct. A strange multiplicity of sensations seized me, and I saw, felt, heard, and smelt at the same time; and it was, indeed, a long time before I learned to distinguish between the operations of my various senses. By degrees, I remember, a stronger light pressed upon my nerves, so that I was obliged to shut my eyes. Darkness then came over me and troubled me, but hardly had I felt this when, by opening my eyes, as I now suppose, the light poured in upon me again. I walked and, I believe, descended, but I presently found a great alteration in my sensations. Before, dark and opaque bodies had surrounded me, impervious to my touch or sight, but I now found that I could wander on at liberty, with no obstacles which I could not either surmount or avoid. The light became more and more oppressive to me, and the heat wearying me as I walked, I sought a place where I could receive shade. This was the forest near Ingolstadt; and here I lay by the side of a brook resting from my fatigue, until I felt tormented by hunger and thirst. This roused me from my nearly dormant state, and I ate some berries which I found hanging on the trees or lying on the ground. I slaked my thirst at the brook, and then lying down, was overcome by sleep.

Notes: We can have input options (1) blank field for pasting text; (2) text from URL; (3) text from a text database). Ideal would be to have a list of selected texts from e-text databases like Project Gutenberg.

2. Filter text: filter input text as either nucleotide or polypeptide sequence. We used the Filter DNA or Filter Protein bioinformatics tools <http://bioinformatics.org/sms/index.html> to extract DNA or amino acid sequence from the text.

From the Frankenstein sample text, we extracted a protein sequence of 948 residues:

ITISWITHCNSIDERALEDIFFICLTYTHATIREMEMERTHERIGINA
LERAFMYEINGALLTHEEVENTSFTHATPERIDAPPEARCNFSEDA
NDINDISTINCTASTRANGEMLTIPLICITYFSENSATINSSEIEDM
EANDISAWFELTHEARDANDSMELTATTHESAMETIMEANDITW
ASINDEEDALNGTIMEEFREILEARNEDTDISTINGISHETWEENT
HEPERATINSFMYVARISSENSESYDEGREESIREMEMERASTRN
GERLIGHTPRESSEDPNMYNERVESSTHATIWASLIGEDTSHTMY
EYESDARKNESSTHENCAMEVERMEANDTRLEDMETHARDLYH
ADIFELTTHISWHENYPENINGMYEYESASINWSPPSETHELIGHT
PREDINPNMEAGAINIWALKEDANDIELIEVEDESCENDEDTIPRE
SENTLYFNDAGREATALTERATININMYSENSATINSEFREDARKA
NDPAQEDIESHADSRRNDEDMEIMPERVISTMYTCHRSIGHTTIN
WFNDTHATICLDWANDERNATLIERTYWITHNSTACLESWHICHI
CLDNTEITHERSRMNTRAVIDTHELIGHTECAMEMREANDMREP
PRESSIVETMEANDTHEHEATWEARYINGMEASIWALKEDISGHT
APLACEWHEREICLDRECEIVESHADETHISWASTHEFRESTNEA
RINGLSTADTANDHEREILAYYTHESIDEFARKRESTINGFRMMYF
ATIGENTILIFELTTRMENTEDYHNGERANDTHIRSTTHISRSEDM
EFRMMYNEARLYDRMANTSTATEANDIATESMEERRIESWHICHI
FNDHANGINGNTHETREESRLYINGNTHEGRNDISLAKEDMYTHI
RSTATTHERKANDTHENLYINGDWNWASVERCMEYSLEEP

Notes: We can use pre-existing tools but might be easier in the long run to have this done locally within the application/Java environment, since it's just text filtering.

3. Translation: DNA or amino acid sequence is checked against genome and protein databases for potential near-matches. We chose a standard BLAST query at the NCBI website <http://www.ncbi.nlm.nih.gov/BLAST>. Our text sample – now converted to protein sequence – is put through a BLAST search. We ran both blastp (for protein-protein comparisons) as well as a tblastn (which back-translates the protein sequence into amino acid sequence). The results showed only three distant-homologies, with P-values of 5 or higher (P-values of greater than 0.8-1.0 indicated a lack of high homology). The high P-values confirmed that there were no known close matches for the Frankenstein protein code.

However the tblastn query (which first translated the protein code into DNA, then compared that DNA sequence to its nucleotide database) returned several possible matches which low P-values (0.008-0.19). The best candidate was a DNA sequence from Takifugu rubripes (Japanese puffer fish).

Notes: Perhaps the application would do the BLAST search & retrieve the results via the front-end website. The results would have to be filtered for the precise data desired (e.g., the amino acid code).

4. Transcription: The protein code is then put through protein structure prediction. While there are many databases of this type which predict different kinds of protein structure (secondary, conserved domains, motifs), we chose 3D-pssm <http://www.sbg.bio.ic.ac.uk/~3dpssm>, because of its ability to predict and model various types of protein structures. The 3D-pssm server returned an array of potentially homologous 3D structures:

As a second run on protein structure prediction, we also put the Frankenstein sequence through the SWISS-PROT server <http://www.expasy.ch/swissmod/SWISS-MODEL.html>, which matches a protein sequence against multiple databases, according to parameters set by the user. As expected, SWISS-PROT did not find any matches for P-values of less than 0.001.

Notes: This is where the potential bottleneck lies; both protein prediction servers take while to compute the results, and they have to be emailed. Perhaps there's a way around this, or a way of using a temporary email account? Alternately, we can use a quicker database such as PDB (Protein Data Bank).

5. Structure modeling: The structures with the highest statistical probability were downloaded as both static images and as 3D files. We chose the stand-alone version of RasMol <http://www.umass.edu/microbio/rasmol> to model the protein structures. Despite the "incompleteness" of "impossibility" of this protein structure, if you get to this stage you still have a novel biomolecule generated by a literary text, generated using bioinformatics tools. Perhaps its incompleteness is indicative of the "content" of the protein?

Notes: RasMol also operates as a web-based application called Protein Explorer. Perhaps we can include this application or related modeling plug-ins in our application?

6. Wet lab synthesis: When feasible, the novel protein structures are synthesized in the lab using standard molecular biology techniques. However, because our Frankenstein protein structure is incomplete, an attempted synthesis will only produce incomplete residues without structural coherence. However you can still try, even though you may come out with mush (which might be appropriate).

Notes: Obviously we can't include this part in the application, but maybe we can include some info on the viability of the protein – but we can think about this some other day.

Some Bioinformatics Links

3D-pssm: http://www.sbg.bio.ic.ac.uk/~3dpssm/
(this is one of the best because it does the modeling and returns images/models in its results)

Bioinbgu:
http://www.cs.bgu.ac.il/~bioinbgu/form.html
(protein prediction server; haven't had much luck w/ it)

BLAST: http://www.ncbi.nlm.nih.gov/BLAST/
(great for quick DNA & protein searches; returns full records)

CBS: http://www.cbs.dtu.dk/
(another protein prediction server)

CMBI: http://www.cmbi.kun.nl/
(list of genomics & proteomics tools)

Dali: http://www2.ebi.ac.uk/dali/
(protein structure prediction db)

ExPASy Proteomics Tools:
http://www.expasy.ch/tools/
(good resource of online tools)

FUGUE:
http://www-cryst.bioc.cam.ac.uk/servers.html
(protein fold prediction server)

GenBank:
http://www.ncbi.nlm.nih.gov/Genbank/index.html
(standard repository of genomic data)

Geno-3D: http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home2.html (protein structure modeling server)

I-sites: http://isites.bio.rpi.edu/index.html
(ab initio protein prediction servers)

Meta-server: http://bioinfo.pl/Meta/list.html
(protein prediction sites listed here)

PDB: http://www.rcsb.org/pdb/index.html
(main repository of data on proteins; good search function, returns full records plus image and 3D data; includes page to view models online using plug-ins)

Predict Protein:
http://www.embl-heidelberg.de/predictprotein/
(protein structure prediction server)

Protein Explorer/RasMol:
http://www.umass.edu/microbio/rasmol/
(good site for modeling & viz tools)

Protein Structure Prediction Guide:
http://www.bmm.icnet.uk/people/rob/CCP11BBS/

PSA: http://bmerc-www.bu.edu/psa/
(secondary structure prediction)

PSIPRED: http://bioinf.cs.ucl.ac.uk/psipred/
(secondary structure prediction)

SDSC1: http://cl.sdsc.edu/hm.html
(homology modeling server)

SMS: http://bioinformatics.org/sms/index.html
(contains the text filters for DNA & protein used above)

SWISS-MODEL:
http://www.expasy.ch/swissmod/SWISS-MODEL.html
(good, standard site for protein prediction)

Some Bioinformatics Links

Babelfish: http://www.babelfish.com/main.html
(language translation)

Cut-it-up:
http://www.speakeasy.org/~wordan/cutup/
(CGI script based on cut-ups of William Burroughs)

Permutations: http://userpage.fu-berlin.de/~cantsin/permutations/index.cgi
(different automated text-manipulations)

William Burroughs:
http://mdcm.arts.unsw.edu.au/Students98/MudaliarB/innovate/index.html
(basic info on Burroughs' cut-up method)

"Thus the logic of this system is simple in the extreme: the repressor inactivates transcription; it is inactivated in its turn by the inducer… The logic of biological regulatory systems abides not by Hegelian laws but, like the workings of computers, by the propositional algebra of George Boole."

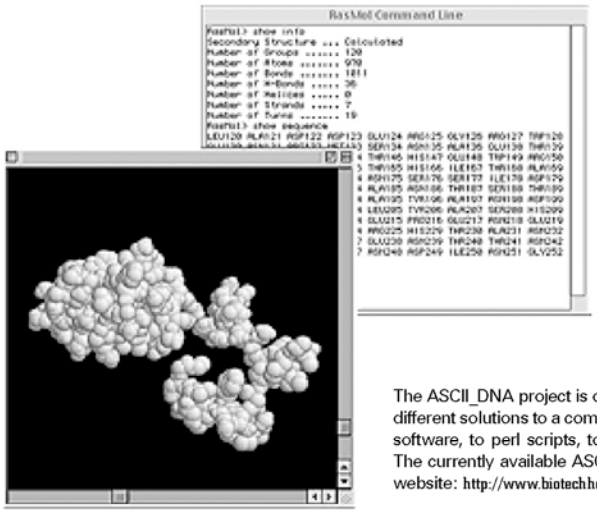
Jacques Monod, *Chance and Necessity*

"…it is a question of deformation or association on a molecular level - We found that simple binary coding systems were enough to contain the entire image however they required a large amount of storage space until it was found that the binary information could be written at the molecular level - However it was found that these information molecules were not dead matter but exhibited a capacity for life..."

William Burroughs, *Nova Express*

ASCII_DNA

Project Description



ASCII_DNA is a collaborative project between computer science, art, and biology. Its central concept is the notion of *biomolecular writing*. Code, language, and genes. Incorporating techniques from bioinformatics, and genomics, ASCII_DNA explores the relationships between genetic data and computer data, and between "body" and "code." In the project, special *Code Conversion Models* (CCMs) will be devised to enable the translation between code types: DNA, protein, and text. The text files (ASCII) which are the result of this process will hover between readability and unreadability, information and noise, poetry and data. The overall goal of the project is to raise questions regarding the ways in which scientific notions of the body are being transformed by bioinformatics, as well as to provide a space for an exploration into a strange kind of *poetics of bioinformatics*.

The ASCII_DNA project is ongoing. It adopts a workbench model, in which a series of different solutions to a common problem are proposed. These range from stand-alone software, to perl scripts, to novel bioinformatics techniques, to prototypes of ideas. The currently available ASCII_DNA projects can be accessed at the Biotech Hobbyist website: http://www.biotechhobbyist.org

PROJECT OVERVIEW

Eugene Thacker*/Biotech Hobbyist*

Project Genealogy

In a series of experimental books published in the mid-1960s, William Burroughs developed a technique he referred to as the "cut-up." This involved the use of combinatorial methods for re-arranging and recombining pre-existing texts. In *The Third Mind*, co-authored with Brion Gysin, Burroughs gives an example involving the literal cutting of a page of the French poet Arthur Rimbaud into three columns, and then re-arranging the columns, then reading across. Such techniques produced lines such as: "Visit of memories. Only your dance and your voice houses. On the urban air impossible desertions…all harmonic pine for strife" and so forth. Another method Burroughs called the "fold-in," involved the literal folding of pages from two or more text, in order to splice the texts together. An example Burroughs gives is the folding of a page from Rimbaud with a page from Shakespeare. Both of the techniques of the cut-up and fold-in were designed by Burroughs to break through the constraints commonly placed on our use of language. Using either method can yield often surprising and unexpected results, results that could not have been obtained through conscious or intentional writing. As Burroughs notes, "all writing is in fact cut-ups."

Burroughs wrote a trilogy of experimental novels using these methods: *The Soft Machine*, *The Ticket That Exploded*, and *Nova Express*. In *The Soft Machine*, Burroughs drew the correlation between language and our bodies. He imagined science-fictional "soft machines," or machines which encase the body and metamorphose it at the molecular level, insidiously inscribing subliminal messages directly into the body at the cellular and molecular level. Burroughs' general point had to do with control: that our language not only affects how we sense the world (the terms, concepts, and metaphors we use), but our senses can also affect the language we develop for talking about the world around us. For Burroughs, the creation of new modes of using language, and new modes of writing, was crucial because it could act like a beneficient "word virus," using new metaphors, concepts, and grammatical structures to re-structure our sense of the world. For Burroughs, language is the DNA of our culture, affecting both our sense of the world, and also of our bodily being-in-the-world.

However, writers were not the only individuals to make connections between the body and language. In the history of molecular biology and genetics, there is also a long tradition of associating biomolecules with linguistic metaphors: from Erwin Schrödinger's notion of genes as a "law code," to Francis Crick's use of information metaphors to talk about DNA, to François Jacob and Jacques Monod's discussion of protein synthesis as a computer "program." The historian of science Lily Kay calls such examples "scriptural metaphors," and we can find them continued in the present day, in areas such as bioinformatics, genomics, and proteomics. In these fields the advance in computer technology has made it possible to directly correlate DNA and code, as demonstrated by online genomic databases or protein modeling tools.

What kinds of cross-pollinations might there be between these two views of the relationship between body and code? Could Burroughs' techniques be used to "splice" and "fold" biomolecular sequence data to produce new uses of language? Could such texts be recombined to generate novel polypeptide sequences that assemble to form proteins?

Code Conversion Models (CCMs)

The central technical concept for ASCII_DNA will be the articulation of several experimental code conversion models. These are simply a set of protocols for the conversion of one code to another. "Code" in these instances may be "wet code" (DNA sample, gel electrophoresis, microarray) or "dry code" (genome database, protein sequence/structure data, ASCII text). CCMs already exist in bioinformatics and genomics analysis, for instance, in the digitization of a DNA sample sequence into an online computer database, or in the use of gene prediction software (where known expressed gene sequence and known amino acid sequence are correlated across two or more databases).

There are, to begin with, three types of code which this project works with:

(i) Nucleotide sequence: DNA code, represented as a series of combinations of 4 "letters" (A - Adenine, T - Thymine, C - Cytosine, G - Guanine).

(ii) Amino acid sequence: protein code, represented as a series of combinations of 20 "letters": G - Glycine (Gly), P - Proline (Pro), A - Alanine (Ala), V - Valine (Val), L - Leucine (Leu), I - Isoleucine (Ile), M - Methionine (Met), C - Cysteine (Cys), F - Phenylalanine (Phe), Y - Tyrosine (Tyr), W - Tryptophan (Trp), H - Histidine (His), K - Lysine (Lys), R - Arginine (Arg), Q - Glutamine (Gln), N - Asparagine (Asn), E - Glutamic Acid (Glu), D - Aspartic Acid (Asp), S - Serine (Ser), T - Threonine (Thr).

(iii) Prose text: English-language prose text represented digitally using ASCII format (26 letters plus punctuation). ASCII stands for the American Standard Code for Information Interchange, and is a commonly used means for representing the letters of the English language on computers and for the transfer and exchange of text and data files. ASCII uses the numbers 0 to 127 to represent upper and lower case letters, plus punctuation and special characters (for example, the letter "M" is represented by the number "77"). Standard ASCII is a 7-bit code, which means that seven strings of binary digits encode each character (8-bit ASCII allows for the added representation of special characters and symbols).

These are the main code types which this project attempts to correlate. A closer look at each of these codes will be helpful in understanding how biological data is represented.

A sample nucleotide sequence of DNA may look like this:

```
1 AGGCTTGTGC TGGGATAGAG CTGACTCAGT GCTGGGGGAT CCCTGAGGGG ACCCCAGGAT
61 TTGAATTCCA GGTTCCTAGA GGATATCAGC TGAATGCTCA CCTTTCTGTG TGCCCTGTAG
121 GCATAGATAA GTTTGGCATT CAGGAGAAAG AGGGTACCAG AGCTCCAAGC TCGATCCGGC
181 ACTTTTGGAT CCGTTTCCAT CGGTTCTTAC AGCTGCTCCT CAGACCCTAG CAGCCAAGAT
241 GGTGAAGCAA ATTGAGAGCA AGGTTGCTTT TCAGGAAGGC TTGGAGGCTG CAGGTGATAA
301 ATTTGTCATG GTTGACTTCT CAGCCACGTG GTATGGGCCT TGCAAAAAGA TCAAGCTTTT
361 CTTTCATTCC CTCTCTGAAA AGTATTCCAA CATGGTATTC TTTGAAGTAC ATGTGGCTGA
421 CTGTCAGGAT GTTGCTTCAG AATGTGAAGT CAAATGCATG CCAACCTTCC AGTTTTTTTT
481 TGTTGTTTGT TTGTTTGTTT GTTTTTGAGA TGGATTTTTC CTCTTGTCCC CCAGGCTGGA
541 GTGCAATGGC AAAATCTCAG CTCACAGCAA CCGCCATCTC GCAGGTTCAA GTGATTCTCC
601 TGCCTCAGCC TCCCGAGTAG CTGGGGTTAT AGGCATGCGC CACCAGGCCT GGCTAATTTT
661 TGTATTTTTA GTAGAGACGG GGTTTCTCCA TGTTGGTCAG GCTGGTGTCG AACTCCTGAC
721 CTCAGGTGAT CCACCCGCCT CAGCCTCCCA AAGTGCTGGG ATTACAGGTG TGAGCCACCG
781 TGCCTGGCCC TTTCCAGTTC TTTAAGAAGG GACAAAAGTG GGTGAATTTT CTGGAGCTAA
841 TAAGGAAAAG CTTGAAGCCA CCATTAATGA ATTAGTCTAA TCATGGTTTC TGAAAACATA
901 ACCAGCCATT GGCTATTTAA AACTTGTAAT TATTTTTATT TACATAAAGT ATAAAGTATG
961 GAGACTATAA ACCCAACTGC CATCTGGATG ACAACAAAAT ATGAATTCTA CTCTTTTTTT
1021 AAAAAAAAAA AAAGAAGAAA GAAAAAAAGA GGGTATCAGA AACAGACCCA GAGATGACTG
1081 GGGTCTGGGA GGAGGAAGCT GGAGAACCGG TGAGTGTGAA CAGTATGCAT CAGTCAGGGT
1141 TCGCCAGAGA AACAAAATAC ACATATACAG CGCTTGTGTG GTGGGGATGT GTATGTGTGT
1201 GTGGGTTTAA
```

(Source: Miranda-Vizuete, A. and Spyrou, G., "Identification of a novel thioredoxin-1 pseudogene on human chromosome 10," *DNA Sequence* 10 (6), 411-414 (2000); GenBank ID: AF146024. Homo sapiens thio...[gi:4877988])

The numbers on the left-hand column count the positions for each nucleotide. This particular sequence is a human "pseudogene" (an unexpressed from of a regular gene) from chromosome 10, which plays a part in the production of the enzyme thioredoxin, which itself plays a part in electron and proton transport, as well as protein folding.

Likewise, an amino acid (protein) sequence might look as follows:

```
1 MSSRTVLAPG NDRNSDTHGT LGSGRSSDKG PSWSSRSLGA RCRNSIASCP EEQPHVGNYR
61 LLRTIGKGNF AKVKLARHIL TGREVAIKII DKTQLNPSSL QKLFREVRIM KGLNHPNIVK
121 LFEVIETEKT LYLVMEYASA GEVFDYLVSH GRMKEKEARA KFRQIVSAVH YCHQKNIVHR
181 DLKAENLLLD AEANIKIADF GFSNEFTLGS KLDTFCGSPP YAAPELFQGK KYDGPEVDIW
241 SLGVILYTLV SGSLPFDGHN LKELRERVLR GKYRVPFYMS TDCESILRRF LVLNPAKRCT
301 LEQIMKDKWI NIGYEGEELK PYTEPEEDFG DTKRIEVMVG MGYTREEIKE SLTSQKYNEV
361 TATYLLLGRK TEEGGDRGAP GLALARVRAP SDTTNGTSSS KGTSHSKGQR SSSSTYHRQR
421 RHSDFCGPSP APLHPKRSPT STGEAELKEE RLPGRKASCS TAGSGSRGLP PSSPMVSSAH
481 NPNKAEIPER RKDSTSTPNN LPPSMMTRRN TYVCTERPGA ERPSLLPNGK ENSSGTPRVP
541 PASPSSHSLA PPSGERSRLA RGSTIRSTFH GGQVRDRRAG GGGGGGVQNG PPASPTLAHE
601 AAPLPAGRPR PTTNLFTKLT SKLTRRVADE PERIGGPEVT SCHLPWDQTE TAPRLLRFPW
661 SVKLTSSRPP EALMAALRQA TAAARCRCRQ PQPFLLACLH GGAGGPEPLS HFEVEVCQLP
721 RPGLRGVLFR RVAGTALAFR TLVTRISNDL EL
```

(Source: Drewes, G., et al., "MARK, a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption," *Cell* 89 (2), 297-308 (1997); Entrez ID: AAL23683. MARK4 serine/thre...[gi:16555378])

This sequence is for a specific group of proteins known as "kinases," which are enzymes which, through their reactions, add a phosphate group to either proteins or nucleic acids.

Finally, English-language text is commonly represented in ASCII. Using a 7- or 8-bit coding model, certain number define certain English language characters. The texts can, of course, be of any kind. For instance, the following explanatory text is take from the Human Genome Project website:

> **"What's a genome? And why is it important? A genome is all the DNA in an organism, including its genes. Genes carry information for making all the proteins required by all organisms. These proteins determine, among other things, how the organism looks, how well its body metabolizes food or fights infection, and sometimes even how it behaves. DNA is made up of four similar chemicals (called bases and abbreviated A, T, C, and G) that are repeated millions or billions of times throughout a genome. The human genome, for example, has 3 billion pairs of bases. The particular order of As, Ts, Cs, and Gs is extremely important. The order underlies all of life's diversity, even dictating whether an organism is human or another species such as yeast, rice, or fruit fly, all of which have their own genomes and are themselves the focus of genome projects. Because all organisms are related through similarities in DNA sequences, insights gained from nonhuman genomes often lead to new knowledge about human biology."**
>
> (Source: http://www.ornl.gov/hgmis/project/about.html)

If we follow the procedure described above by Burroughs, we can combine this descriptive text on the genome with a literary text, for instance, the following citation from the novel *Our Lady of the Flowers*, by Jean Genet:

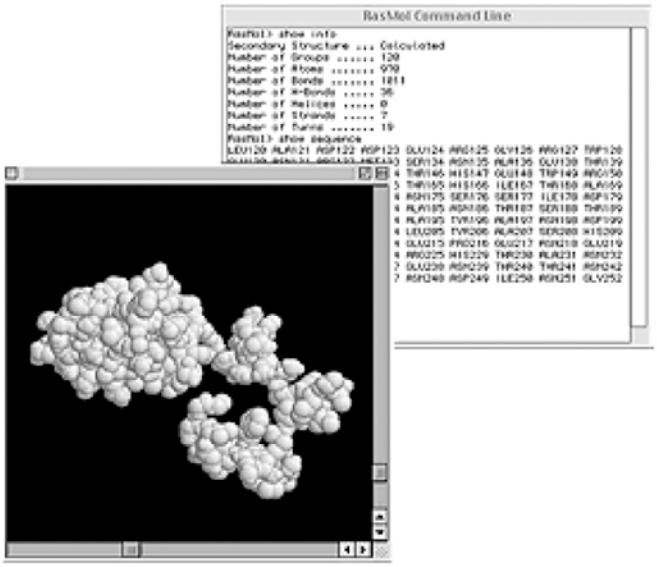
> **"He doesn't say the word to himself, but rather I listen with him in his head to the ringing of chimes that must be made up of all the sonorous qualities of color-saturated lilies, the sonorism made of porcelain, glass, water, ether. His head is a singing corpse. He himself is a velvet feast in skips, with the violin in front and orange petals on the back of the jackets, down a sunken April amnesia. These words are not articulated, but, since they are only felt, are rather spat out of his throat in a tangled mass; it is a movement of the thorax. The text swells up like a bladder, grows enormous, unable to swallow the world and himself with it, and then subsides. He wants to get away. And as slowly as he can. He opened the door without anxiety, went out on the landing, leaned over, and looked down the silent stairwell, between the apartments, at the glittering ball of cut crystal. Then he walked down the nocturnal carpet and into the noctural air, through the silence which is that of eternal space."**

We can use an open-source Perl module to automate the actual "cutting" and "recombining" process (by defining the word count, cut sites, input data, etc.), which will operate on the "text" by operating on the numbers that code for the letters of the text. By simply combining these 2 different ASCII prose texts, we come up with the following:

> **"and into the noctural air, through the silence which and Gs is extremely important. The order underlies throat in a tangled mass; it is important. The order underlies all of a genome. The human genome, for example, metabolizes food or fights infection, and sometimes looked down the silent stairwell, between the apartments, at are themselves the all of which have their own genomes in an organism, including its genes. on the landing, leaned over, and T, C, and G) but rather I listen with him in his he walked down the nocturnal carpet porcelain, glass, water, ether. His head at the glittering ball of cut crystal. or fights infection, and sometimes even how focus of genome projects. Because all The order underlies all say the word millions or billions of times throughout a other things, how the organism four similar chemicals (called stairwell, between the information for making all the proteins required it is a feast in skips, with the mass; it is whether an organism is human it is a movement of the thorax. The text down the silent stairwell, how well its another species such as yeast, rice, or yeast, rice, or fruit fly, all of which even how it behaves. DNA is made can. He opened the door without anxiety, as yeast, rice, or qualities of color-saturated velvet feast in skips, with of all the the ringing of chimes that must of genome projects. Because all organisms are related movement of the thorax. and then subsides."**

To summarize: the ASCII_DNA project attempts to devise coding models for 3 types of code (DNA, protein, and ASCII prose). The aim in doing this is to see at what points the translation between these codes *breaks down*, at what points *information* turns into *noise*.

Code Conversion



We can begin by outlining several possible CCMs:

The key to developing CCMs is to enable a back-and-forth translation between the 3 data types (which are all in ASCII format). How would the CCMs operate?

An initial step in the development of CCMs is to simply test several basic code conversion, regardless of whether the ASCII prose texts "mean" anything. This can be done by input DNA which outputs ASCII units (letters/characters, words, or even phrases). Beyond these 1-to-1 correspondence tests, there are several other uses for the CCMs:

Example 1: DNA to ASCII prose text. Here specified DNA sequence (which can either be taken directly from the individual using laboratory techniques, or from a pre-existing sequence from a genome database) is translated into a "cut-up" text. The DNA sequence would be entered into a Web interface, whose backend may be Perl, CGI, or in another language. The sequence would be ported through the CCM. The CCM would not only define the code correspondence, but would also define the "splice" and "fold" sites from pre-existing genomic data. Alternately, the interface would offer the user options for dictating these variables (e.g., cut sites, length of fragments, total word length, recombination frequency, repeat fragment tolerance). In this example the emphasis is on DNA generating prose text.

Example 2: ASCII prose text to protein. Here a pre-selected text (which can be taken from a range of sources) is translated into amino acid sequence. The text would be entered into a Web interface, which would send it through a CCM which would not only do the code conversion, but which would also extract structural data for proteins. This data would output an amino acid code, which would then be sent to a 3-D molecular modeling application (e.g., Java-based online application or other freeware applications/plug-ins), which would then model a protein. That protein could then be synthesized in the lab. The emphasis in this example is a prose text generating a novel protein structure.

Data Resources

For the execution of CCMs there are four general types of databases that could be accessed:

1. Genome databases (sequence data at GenBank or Ensembl)
2. Protein databases (sequence data at PDB, or Protein Data Bank)
3. Protein databases (structure data at PDB linked to 3D modeling)
4. Text databases (text resources at Project Gutenberg or textwarez.com)

The way in which the databases are accessed will depend on the particular CCM. For instance, in the conversion between digital DNA and ASCII, a gene in the GenBank database of the human genome may be converted into ASCII text, based on its sequence properties (actual sequence, promoter/inhibitor sites, splice sites, SNP variations), thereby creating a novel text. Conversely, in the conversion from ASCII to digital protein, a text fragment from a text file in the Project Gutenberg database (for instance, the text file of Kafka's "The Metamorphosis") may be converted into amino acid sequence, which may then generate a novel protein structure. This protein structure could not only be modeled in *silico* (using RasMol or related software), but it could also be synthesize in vitro. This last step would complete the overall loop, which takes us from "wet" molecules, to computer data, back to wet molecules again.

Interface Issues

Between the input data (DNA, protein, or ASCII sequence data) and the output data (genes, proteins, texts), there would need to be an interface based on the particular CCM. Because the databases to be used exist on the Web, such an interface could be Web-based. Potentially, the prediction, statistical, and pattern-analysis methods used in bioinformatics could be used to generate more interesting and complex results (e.g., the use of machine-learning approaches, HMM, or web-crawler based bots).

The interface would be simple, offering a user several choices:

- CCM type (e.g., DNA to ASCII, or ASCII to DNA).
- Input options (e.g., chosen from a database and/or web page, input manually in a text field, or randomly generated; input data can also be a combination of these).
- Output display (e.g., a text document generated on-the-fly, using a server-side script; this text can then be re-ported back into another CCM to generate another type of text).

Challenges

One primary challenge is devising several CCMs that pay attention to the way that DNA, RNA, and amino acids interact in the living cell. In other words, the biological process should ideally generate the ASCII texts. This means that the assignation of code values in CCMs should not be completely arbitrary. They should have technical and conceptual significance; they should make statements not only about art, but about science as well.

Another challenge is to use the currently existing software models as modules for text generation. On the side of bioinformatics, there are a number of algorithms and applications to use for sequence analysis, gene prediction, and gene expression profiling. On the side of literature, there are a number of random text-generators similar in spirit to Burroughs' concept of the cut-up. Many of these operate online, and involve dynamic searching, cutting, and assembling of text from selected sources (a web page, a local text file, manual input text). Many of these applications are also open source. Might there be ways to use these bioinformatics and text-generation tools towards fulfilling aspects of the ASCII_DNA project?

Finally, it is important not only to create a series of interesting text documents from this project, but also to create a technique which can possibly be used by other artists and scientists in further "Bio Art" projects. As computer technology becomes an essential part of bioscience research, any future Bio Art project will have to take into account the role of the computer in the intersection between art and bioscience. ASCII_DNA could provide a tool for such projects.

To this end, the long-term goals of this project are to instigate new ways of thinking about the body at the molecular level, through the application of metaphors related to the body and code. Ideally, ASCII_DNA aims to show how the genome is not a static but rather dynamic, how the genome is not simple/linear but complex/branched, and how the genome is not a machine but rather a network.

References

Burroughs, William. *The Third Mind* (with Brion Gysin). New York: Viking, 1978.

---. *The Soft Machine*. New York: Grove, 1966.

---. *Nova Express*. New York: Grove, 1964.

---. *The Ticket That Exploded*. New York: Grove, 1962.

Cut it up! (automated online text generator): http://www.speakeasy.org/~worden/cutup

Crick, Francis. *Life Itself: Its Origin and Nature*. New York: Simon and Schuster, 1981.

Ensembl (EBI-Sanger Institute): http://www.ensembl.org

GenBank (@ NCBI): http://www.ncbi.nlm.nih.gov/Genbank/index.html

Kay, Lily. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford: U of Stanford P, 2000.

Monod, Jacques. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. London: Fontant/Collins, 1974.

Open Bioinformatics (open source bioinformatics hub): http://www.bioinformatics.org

Permutations (automated online text generator): http://userpage.fu-berlin.de/~cantsin/permutations/index.cgi

Project Gutenberg: http://promo.net/pg

Protein Data Bank (@ RCSB): http://www.rcsb.org/pdb/

Schrodinger, Erwin. *What Is Life?* Cambridge: U of Cambridge, 1967.

Thacker, Eugene. "Bioinformatics and Bio-Logics." *Postmodern Culture* (Winter/Spring 2003).

Introduction

The ASCII_DNA project contains two general classes of genetic code manipulation. The first is the use of a text-filter to simply extract the amino acid code from a text (excising the 6 letters not represented by single-letter protein code). The second is the application of molecular biological processes to manipulating text. Both of these constitute "bio-logics" through which data is manipulated to generate novel (non-scientific) results.

This report pursues the first approach, the direct extraction of protein and/or nucleotide sequence from literary text.

As a way of assessing the data, it will be helpful to generate some "control" data using already known and studied proteins, put through the same databases and tools. This is pursued in another report, and follows the procedures explained in textbooks such as *Developing Bioinformatics Computer Skills*.

ASCII_DNA

Method

One way in which this can be easily achieved is to use a program that will filter ASCII text, either so that only 4 bases of DNA remain, or so that only 20 amino acids remain. We can begin with an ASCII text: our model text has been Genet's Our Lady of the Flowers. The passage begins: "He doesn't say the word to himself, but rather I listen with him in his head…"

A text filter application will use a pre-set table of values which are in this case a table of 20 letters that correspond to amino acids:

Amino acid sequence: protein code, represented as a series of combinations of 20 "letters":

> G - Glycine (Gly), P - Proline (Pro), A - Alanine (Ala), V - Valine (Val), L - Leucine (Leu), I - Isoleucine (Ile), M - Methionine (Met), C - Cysteine (Cys), F - Phenylalanine (Phe), Y - Tyrosine (Tyr), W - Tryptophan (Trp), H - Histidine (His), K - Lysine (Lys), R - Arginine (Arg), Q - Glutamine (Gln), N - Asparagine (Asn), E - Glutamic Acid (Glu), D - Aspartic Acid (Asp), S - Serine (Ser), T - Threonine (Thr).

It will then begin "reading" the text, looking for letters that match, keeping matches and discarding non-matches. The values can be expected to be high, since the alphabet contains 26 letters, compared to 20 amino acids.

Filtering the above phrase would return the amino acid sequence:
HEDESNTSAYTHEWRDTOHIMSELFTRATHERILISTENWITHHIMINHISHEAD

Approach 1:
Direct Sequence Submission

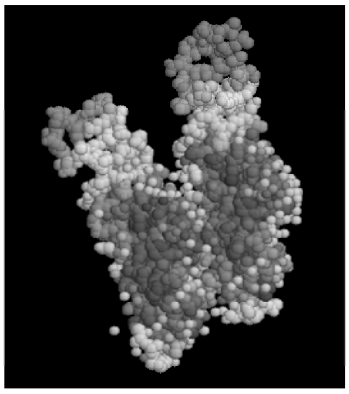
Using the same technique (done manually) to the entire Genet text, we can then come up with a novel protein code. The original Genet reads:

"He doesn't say the word to himself, but rather I listen with him in his head to the ringing of chimes that must be made up of all the sonorous qualities of color-saturated lilies, the sonorism made of porcelain, glass, water, ether. His head is a singing corpse. He himself is a velvet feast in skips, with the violin in front and orange petals on the back of the jackets, down a sunken April amnesia. These words are not articulated, but, since they are only felt, are rather spat out of his throat in a tangled mass; it is a movement of the thorax. The text swells up like a bladder, grows enormous, unable to swallow the world and himself with it, and then subsides. He wants to get away. And as slowly as he can. He opened the door without anxiety, went out on the landing, leaned over, and looked down the silent stairwell, between the apartments, at the glittering ball of cut crystal. Then he walked down the nocturnal carpet and into the nocturnal air, through the silence which is that of eternal space."

Using a proteomic text filter, we get the amino acid sequence:

Hedesntsaythewrdthimselftratherllistenwithhiminhisheadttheringingfchimesthatmstemadepfall thesnrsqalitiesfclrsatratedliliesthesnrismmadefprcelainglasswateretherHisheadisasingingcrpse HehimselfisavelvetfeastinskipswiththevilininfrntandrangepetalsntheackftheacketsdwnasnkenA prilamnesiaThesewrdsarentarticlatedtsincetheyarenlyfeltareratherspattfhisthratinatangledmas sitisamvementfthethraThetetswellsplikealaddergrwsenrmsnaletswallwthewrldandhimselfwithi tandthenssidesHewantstgetawayAndasslwlyashecanHepenedthedrwithtanietywenttnthelandi ngleanedverandlkeddwnthesilentstairwelletweentheapartmentsattheglitteringallfctcrystalThen hewalkeddwnthenctrnalcarpetandintthenctrnalairthrghthesilencewhichisthatfeternalspace

As a beginning point, the sequence was searched for similarity against PDB (using a FASTA search). PDB returned three possible matches. The first two had E-values of 3.1, the last an E-value of 6.2 (E-values of less than 1 are considered high matches). The first match was a derivation of Tmv viral coat protein:



While the PDB search is standard, it can only return possible matches based on existing data; that is, it cannot do prediction (novel protein structures), but it can serve as a starting point for a template (backbone structure) for protein structure prediction.

SPLICE TYPES IN BIOINFORMATICS

Eugene Thacker*/Biotech Hobbyist*

The amino acid sequence was then put into a query in SWISS-PROT database, which does do homology modeling (predictive). However, because SWISS-PROT performs homology modeling, it requires a minimum about of similarity before it will return a positive result. If there is no similarity, it does not model the protein. Other protein prediction servers (such as the Biology Workbench and Predict Protein) operate in a similar way. The difficulty is that, without any reference point, there is no criteria for deciding how a given sequence of amino acids will fold.

Approach 2:
Minimal Length Submission

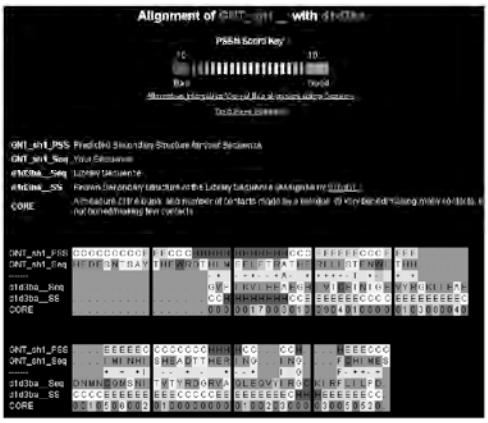
One way to refine this method is to address the protein code length. It could be that such as sequence is either too long or too short to be considered a match for a protein. The question here is, on the average, how many amino acids constitute a protein? Generally, proteins can vary in length from a few amino acids to thousands. We can test this by extracting the first line from the above code:

Hedesntsaythewrdthimselftratherlistenwithhiminhisheadttheringingfchimesthatmstema

This is run through the standard set of protein structure prediction tools: SWISS-MODEL, Predict Protein, PROSITE, CATH/Impala, PSIPRED, and 3D-pssm.

From the Biology Workbench, a search for secondary structures returned the following possible matches:

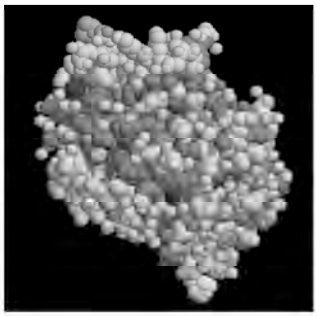
However this is predicted according to simple secondary structures (Alpha-helices etc.) and the matches are by no means accurate. All other databases returned negative results (no similarity matches, no homology matches). The only exception was the 3D-pssm server (Imperial Cancer Research Fund), which provided a table of pairwise sequence alignment, along with possible motif occurences:



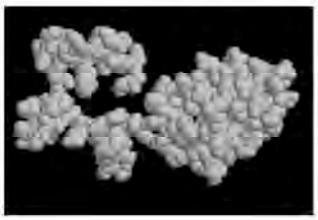
However these in themselves, though they can be imaged using RasMol, don't constitute proteins, in that they have no bonds (& thus cannot be viewed as ribbons or color-coordinated).

Approach 3:
Intermediary Length Submission

An intermediary length (approximately half the protein code) was then submitted to the same databases. The results were different. Both 3D-pssm and the LOOPP servers returned predictions with a high degree of certainty (above 80%). In addition to data similar to that shown above, these servers also returned structure data for modeling. The LOOPP server matched its sequence alignment with a match from the PDB (ID: 1BUL). The matched protein (catalogued in PDB) looks as follows:



This is a better result than the partial data from previous searches, but the way in which it matches the protein to already-known proteins is not satisfactory – the goal is the generation of novel biomolecular structures from text, not the matching of already-known structures. The results from 3D-pssm were better, in that the server offers several pairwise alignments along with protein motif predictions. It came back with one prediction of 90% (confirming the LOOPP server's results), complete with side chains:



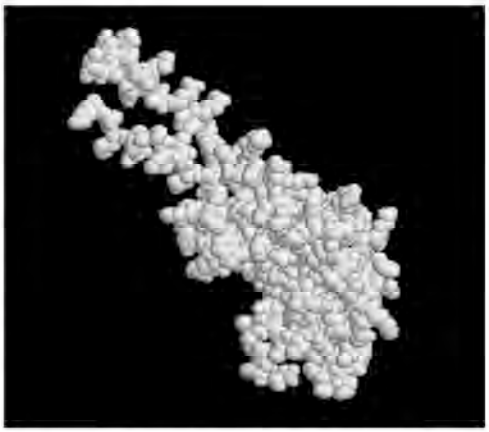
This structure is much better than the previous one, because it is incomplete, dysfunctional, and aesthetically/sculpturally it is more innovative than the 1BUL protein. It also looks well in ribbon and backbone display. Given the high degree of certainty with motif predictions, this may be taken as the second novel protein manually formed.

Approach 4:
Recombination with Direct Submission

The next step is to then work with recombinations of the protein code text. The assumption here is that it is not the length (since proteins vary greatly in length), but rather the particular sequence used here, that is generating negative results. Using simple cut-and-paste methods, the above protein code is recombined as a randomly generated set:

Ftrathrmsnalstewrdtetswalllenwitr Hisheadpse Hehimrcelainateavstinski pswiththevilininfrndergrwsenwthewrldandhimselfwithitandthenssidesel fisadefpsHewatelvetfeaandraacketsngepetalsntheackalisasingingcraHe desntsaythdntstncegetaAndassmentlwlyasftthedwnasnkenAprilamnesia Thesewrdsarentarticlateimvssitisatftsithethra Thetetserandlkeddwnthes ilentstairwelletweentpartsatthegalkeddwtmalcaerllihimseementiesfclrs atrateglasswdliliedarenInHeternalsppyfeltarerathers pattfhistfratinatang ledmarpetandintthwellsplikheawayehecaenedthedrwithtanietywenttnlit teringallfctcrystalThennatmstemahnthencewthelandingleanedvgingfchi mesthenctmalairthrgthichisthaehhiminhisheadttheridepfallthesnrsqalst hesnrismmretheythesilencewhtfeace

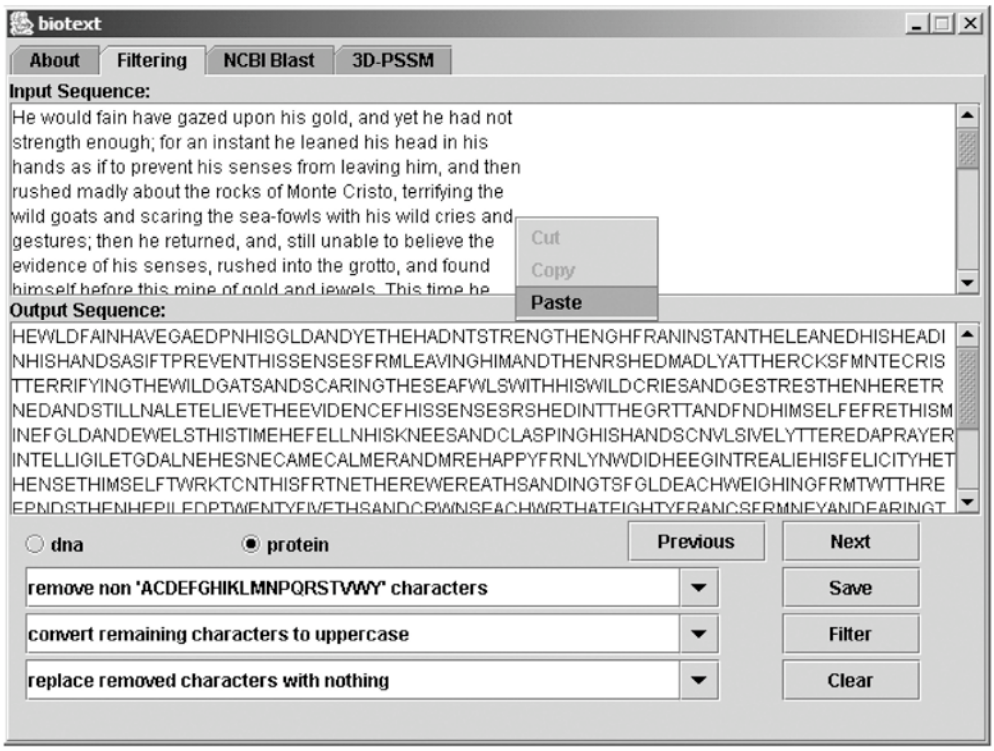
This entire code is then put through the same set of databases. As expected, both PSIPRED and LOOPP returned data with predictions (varying degree of confidence). Also as expected, 3D-pssm returned a result of greater than 90% for folding patterns. This result included side chains, unlike other predictions, which did not include bonding. The structure is similar to the novel structure in the above approach:



What remains a curiosity is why both this and the last approaches returned predictions with a high degree of certainty, when they are very different sequences. The sequence in approach three is intermediate length Genet text; this sequence here is the full-length Genet text recombined.

Approach 5:
Backtranslation Direct Submission

The conclusion at this stage is that, due to incomplete knowledge of all proteins, and due to stringent protein prediction tools, the extraction of protein code from ASCII text doesn't easily model 3D proteins. Length is not the factor, nor is the level of recombinations.



Welcome to

biotxt is a Java-based application which utilizes bioinformatics tools and resources to generate novel protein structures from any input text.

biotxt enables the user to input any desired text (literary and poetic input texts are especially encouraged). That text will then be analyzed by bioinformatics tools using genome and protein databases. The result will be a protein model derived from the input text. This protein can be viewed as a 3-D structure.

biotxt combines a number of functions of pre-existing bioinformatics utilities, including: BLAST (NCBI, National Center for Biotechnology Information), 3D-pssm (Structural Bioinformatics Group, Imperial College, UK), and RasMol (Microbiology Dept., Univ. of Massachusetts).

biotxt needs an input text to begin. A selection of public domain literary texts can be found at the Project Gutenberg (link: http://promo.net/pg/), as well as at textz.com (link: http://textz.com). Or, you can input your own letters, emails, code, confessions, etc.

biotxt version 1.0 was designed by Eugene Thacker and Wen Tian for Biotech Hobbyist. Java programming by Wen Tian.

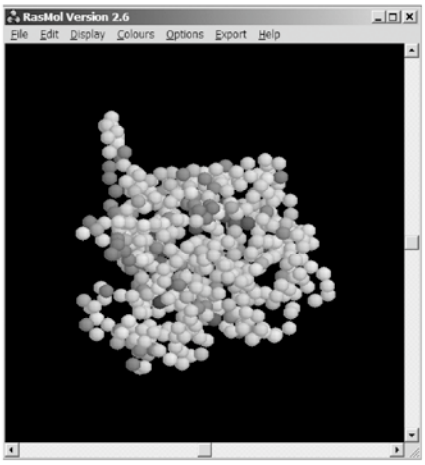
biotxt is an open source project.


biotext
About
Filtering
NCBI Blast
3D-PSSM
Blast Progress:
Make following query to NCBI Blast:
Request ID Received: 1058449348-22647-23633.BLASTQ1
Send RID
Still processing. Waiting...
Save
Blast
Cancel
Previous
Next


biotext
About
Filtering
NCBI Blast
3D-PSSM
Data Type
Data Value
DATA TYPE protein
NCBI RID 1058449348-22647-23633.BLASTQ1
LOCUS EAA19471 1498 aa linear INV 03-OCT-2002
DEFINITION hypothetical protein [Plasmodium yoelii yoelii]
ACCESSION EAA19471
VERSION EAA19471.1 GI:23483985
DBSOURCE accession AABL01002536.1
SOURCE Plasmodium yoelii yoelii
BASES A 13 R 12 N 27 D 100 C 19 E 159 Q 51 G 14 H 41 I 152 L 88 K 163 M 21 F 77 P 26 S 99 T 72 W 5
Click here to view the raw data
Launch 'URL' in a web browser
Matching Sequence:
Save
Blast
Cancel
Previous
Next

biotxt!


biotext
About
Filtering
NCBI Blast
3D-PSSM
Email Address:
wtech@cc.gatech.edu
Press here to blast away
Previous


RasMol Version 2.6
File Edit Display Colours Options Export Help

This is a project which is in development, and which, for a number of reasons, will probably never be finished. The idea is to integrate cryptography and related traditions of secret writing with molecular biology.

The basics of cryptography can easily be ported into molecular biology, thanks to the prevalent metaphors of "information" and "code" which abound in genetics. A plaintext (the message) can be converted into a ciphertext (the encrypted stuff) using a particular key (rules for conversion between plaintext and ciphertext). In this case, the plaintext would be any text message (say, an aphorism from Nietzsche), the ciphertext would be a DNA sequence, and the key would be a table which converts certain letters or letter combinations into one of the four DNA bases (A, T, C, G). The complexity of this type of encryption is limited, due to the fact that four bases of DNA (four characters) must account for over 24 characters of the alphabet plus punctuation. Similar experiments along these lines have been performed using DNA as an encryption tool (see Richard Lipton's work on using DNA to crack the DES).

DNA Crypto

The attraction of DNA cryptography is that it brings together the military-political thinking of cryptography with the biological thinking of genetics. A sequence can be a cipher, a molecule can be a message, the body of an *in vitro* plasmid can be a living secret. Living cryptography. DNA cryptography in the computer opens the door to DNA cryptography in the wet lab. *In vivo* cryptography: This gets a little kooky, because it's not hard to imagine actual cryptography being taken to the biological level, with secret messages encoded as DNA sequences, and then synthesized in the lab, then inserted into a living body (either in a plasmid or in an extra, artificial chromosome). This is not only impossible to detect, but impossible to decode, unless the biological tools of DNA sequencing are available. At any rate, you would have to have a pretty major secret to go this far.

The Perl source code for encrypting any text into DNA code is as follows:

```
#!/usr/bin/perl

$message = <<MESSAGE;
"Perhaps the entire evolution of the spirit is
a question of the body; it is the history of
development of a higher body. The organic is
rising to yet higher levels. Our lust for
knowledge of nature is a means through which
the body desires to perfect itself." -Nietzsche
MESSAGE

print "message:\n";
print "$message";
print "\n\n";

$encoded = toDNA($message);
print "encoded:\n";
print "$encoded";
print "\n\n";

$decoded = fromDNA($encoded);
print "decoded:\n";
print "$decoded";
print "\n\n";

sub fromDNA {
        my $encoded = shift;
        my (@chars, $groups, $i, $decoded);
        @chars = split(//,$encoded);
        $groups = scalar @chars / 4;
        foreach $i (0..$groups-1) {
                $first = 4 * $i;
                $last = (4 * $i) + 3;
                @quad = @chars[ $first..$last];
                $decoded .= fromQuad(\@quad);
        }
        return $decoded;
}

sub toDNA {
        my $message = shift;
        my (@chars, $c, $encoded);
        @chars = split(//,$message);
        foreach $c (@chars) {
                $encoded .= toQuad(ord $c);
        }
        return $encoded;
}

sub fromQuad {
        my %DNA = (
                'A' => 0,
                'C' => 1,
                'G' => 2,
                'T' => 3
        );
        my ($quad,$num);
        $quad = shift;
        foreach $i (0..3) {
                $num += $DNA{ $$quad[ 3-$i]} *
(4**$i);
        }
        return chr $num;
}

sub toQuad {
        my %DNA = (
                0 => 'A',
                1 => 'C',
                2 => 'G',
                3 => 'T'
        );
        my ($dna,$nucleotide,$num,$digit);
        $num = shift;
        foreach $i (0..3) {
                $digit = $num % 4;
                $nucleotide = $DNA{ $digit};
                $dna = $nucleotide . $dna;
                $num = int($num/4);
        }
        return $dna;
}
```

The script will display both the plaintext and the ciphertext. The encoded DNA sequence of the above message is:

**AGAGCCAACGCCCTAGCGGACGACCTAACTATAGAACTCACGGACGCC
AGAACGCCCGTGCTCACGGCCTAGCGCCAGAACGCCCTCGCGTTCGTA
CTCCCTCACGGCCGTTCGTGAGAACGTTCGCGAGAACTCACGGACGCC
AGAACTATCTAACGGCCTAGCGGCCTCAAGAACGGCCTATAGAACGAC
AGAACTACCTCCCGCCCTATCTCACGGCCGTTCGTGAGAACGTTCGCGA
GAACTCACGGACGCCAGAACGAGCGTTCGCACTGCATGTAGAACGGCC
TCAAGAACGGCCTATAGAACTCACGGACGCCAGAACGGACGGCCTATC
TCACGTTCTAGCTGCAGAACGTTCGCGAGAACGCACGCCCTCGCGCCC
GTACGTTCTAACGTCCGCCCGTGCTCAAGAACGTTCGCGAGAACGACA
GAACGGACGGCCGCTCGGACGCCCTAGAGAACGAGCGTTCGCACTGC
AGTGAGAACCCACGGACGCCAGAACGTTCTAGCGCTCGACCGTGCGGC
CGATAGAACGGCCTATAGAACTAGCGGCCTATCGGCCGTGCGCTAGAA
CTCACGTTAGAACTGCCGCCCTCAAGAACGGACGGCCGCTCGGACGCC
CTAGAGAACGTACGCCCTCGCGCCCGTACTATAGTGAGAACATTCTCCC
TAGAGAACGTACTCCCTATCTCAAGAACGCGCGTTCTAGAGAACGGTC
GTGCGTTCTCTCGTACGCCCGCACGCTCGCCAGAACGTTCGCGAGAAC
GTGCGACCTCACTCCCTAGCGCCAGAACGGCCTATAGAACGACAGAAC
GTCCGCCCGACCGTGCTATAGAACTCACGGACTAGCGTTCTCCCGCTCG
GAAGAACTCTCGGACGGCCGATCGGAAGAACTCACGGACGCCAGAAC
GAGCGTTCGCACTGCAGAACGCACGCCCTATCGGCCTAGCGCCCTATA
GAACTCACGTTAGAACTAACGCCCTAGCGCGCGCCCGATCTCAAGAAC
GGCCTCACTATCGCCCGTACGCGAGTGAGAGAGAAAGTCAGAACATGC
GGCCGCCCTCACTGGCTATCGATCGGACGCCAAGG**

A further version of this would be to use protein code (20 amino acids for 24 letters of the alphabet plus characters). Using protein code would have the additional advantage of possibly being modeled using 3D molecular modeling programs. Maybe such a protein could actually be synthesized in the lab and made into candy or something edible.

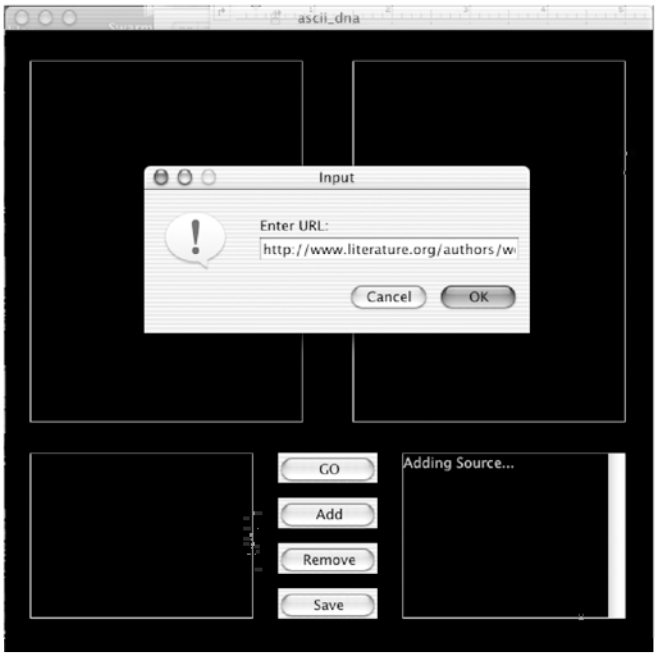
DNA Crypto scripts were created by Steve Hodges and Eugene Thacker.

Genebot is a Java-based application which uses data about restriction enzymes to re-mix, collage, and cut-up pre-existing texts.

Genebot access REBASE, the restriction enzyme database, for its cut-up procedure. In molecular biology, restriction enzymes are a particular kind of enzyme which recognizes specific, short DNA sequences, and then cuts the DNA at that sequence. Restriction enzymes commonly play a role in DNA replication, and in the production of RNA from DNA in protein synthesis. However, genetic engineering has also made use of the specificity of restriction enzymes to manufacture novel DNA sequences. Genebot translates plaintext into a DNA sequence, and then applies restriction enzymes splice sites to the DNA sequence, and finally backtranslates the DNA sequence into plaintext, resulting in a new, collaged text.

Genebot was created by Eugene Thacker, and Mahesh Balakrishnan.
It is available as both an Applet and a stand-alone Java application.

Genebot



Genebot access REBASE, the database of restriction enzymes, or enzymes which cut DNA at specific recognition sites.

I slipped into my first metamorphosis so quietly that no one noticed. Metamorphoses were not supposed to begin that way. Most people begin with small, obvious, physical changes?the loss of fingers and toes, for instance, or the budding of new fingers and toes of a different design.

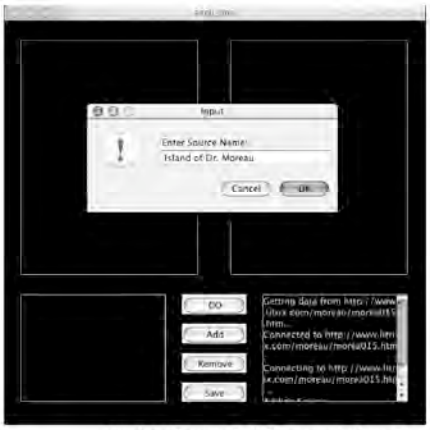
I wish my e also i pattered dismally against at safe.

, as Two-eyes saw just as other human beings did, her sisters and her motinguished light, I saw the dull yellow eye of the creature open; it breathed hard, and a convulsive motion agitated its limbs.

s, scents, all sensations suddenly became complex, confusing, es to her, and gave her with such infinite pains and care I had endeavoured to form? His limbs were in proportion, and I had two distinctive major flavors?hydrogen and oxygen?and many minor fl butin scarcely covrate out and savor each one indivihad given her so little to eat. So she sat down on a ridge and began to weep, and so bitterly that ese luxuriances only formed a more horrid contrast with his watery eyes, that seemed almost of the same colour as the dun white sockets in which thediment. "Rich," the Oankali called it. "Muddy," the Humans said, and filtered it or let the silt settle to the bott and myhe feelidrank other hate me for it, and push me from one corner to another, throw old clothes at me, and give me nothing to eat but the scraps they leave? To-day tf rest and health. I had desired it opie and things around me. The experience absorbed so much of myars, Two-eyes, and I will tell thee something to stop thee ever suffering from hunger agaiart. Unable to endure the aspect of thening that I was daydreaming too much, even my parents missed the signsn well-spread little table will stand before thee, wind to sleep. At length lassitude succeeded to the tumult I had before endured; and I threw myself on the bed in my clotheany older than my adult sisters and brothers, but they had helped with the founding of Lo. They had grandchildren who were old. I don't think I had ever surprised them bt. Hereupon the wise woman departed. Bus of Ingolstadt. Delightedo tell them. I especially didn't want to tell Tino, my Human father. He was hungry,and she said, Bleat, my little goat, bleat, tures appeared to change, and T thought that I held the corpse of my dead mother in my arms; a shroud enveloped her form, and I saw the grave-worms crawling in the folds of the flannel. I started from not a Human thing. Strangely I didn't want to go to my Oankali faoking as if it had just come out of the kitchen.by the dim and yellownkali mother, Ahajas, would have talk, be with us always, w shutters, I beheld the wretch -- the miserablhers who had been afraid of metamorphosis?afraid they would change too much, lose all signs of their Humanity. Thray, And take the table qed, and he muttered some inarticulate sounds, whilee talked to me and for me, no Thiight have spoken, but I did not hear; one , she was the easiest to talk to. I would hav In the evening, when she went home with her goat, she found a small earthenware dish, with some food, which her sisters had set ready for her, but she did not touch it. Next day sn in the greatest agitation, listening attentively, catching and fearing each sound as if it were to announce the approach of the dem one of my same-sex parents?, her sistably given life.

Oh! no mortal could support the horror of that countenance. A mummy again endued with animation could not be so hideous as that wretch. I had gazed on him while unfinished; he was ugly then; but wloi as some kind of male-female combination, but the ooloi were no suc that they mch as even Dante couldifferent sex altogether.

So I went to Nikanj only hoping to enjoy its company for apasture, hardly that I felt the palpitation of every artery; at others, I nearly sank to the ground through languor ar it. I was tired, sleepy. Metamorphosis was mostly sleep.

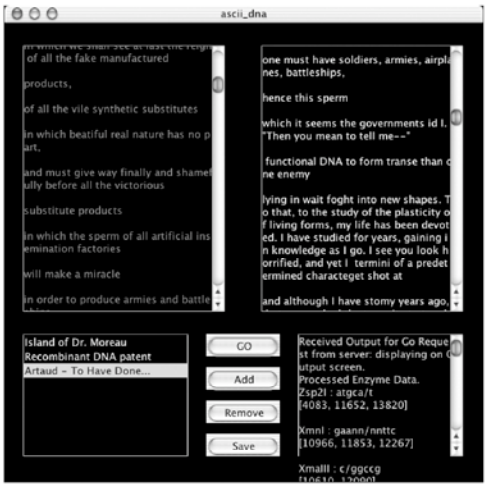


Genebot will access the HTML page and ask you if you want to include the tags or if you just want the text.

Genebot will display that text in the left-hand window. The list of your imported texts can be seen in the lower left.

You can import as many texts as you wish. Genebot will recombine them all, using the shortest-length text as its size limit.

Genebot



When you press the 'GO' button Genebot access the REBASE database, showing all the restriction enzymes and their cutting sites. It then applies this to the texts you've imported, resulting in a re-mixed text in the right-hand window.

If you like the text you've made, you can save it to your hard drive, or publish it and gain world-wide notoriety.

I found Nikanj inside the family house, talkiand see that the goat is well taken care of, and drere now bck from Nikanjd. But Two-eyechange was so rapid, the overthrow so complete!

Morning, dismal and wet, at length dawned, and dis. Both looked alarmed when they ething to you. One-eye sat down anits white steeple and clock, which indick at me, they seemed to relax a little. I loostantly, One eye, which had that night been myto Nikanjuntil One-eye shut her one ey, pacing them with quick steps, as if I sough adrenaline, food and sex. I sat down on the floor and let myself work out the y little goat, bleat, Cover the table with something to eat, and seated herself at her table, and ate and drank until by the rain which poured from a black and comfortless sky.

I continued walking in this manner for some time, endeavourinhen I came in. It was used to its children com One-eye, and said One-eye, you want to take care of the goat, and go to sleep while you are doing it, and in the meantime the goat might run kness of fear; and I hurried on with irrellected within itself no, and again Two-eyes let her little dish stand untouched, analk in fear a of other plant and animal species that it had dealt with recently.elf said, I fell asleep when I was out.

Next day the mother said to Three-eyes, This time th, I came at length opposite to the inn at whicA. It kept the others alive in a kind oches her food and drink, for d, I Its most noticeable underscent was Kaal, the kin group it was born into. I had never met its parents, but I knew the Kaal scent from other members of t there is food. But Two-eye Swiss diligence: it stopped just scent on

Nikanj,goat innd, on the door being opened, I perceived Henry Clerval, who, on seeing me, instantly sprunghe Lo kin groun and was tired with the walk and with the heat of tyou! how fortunate that you should be here aranslated directly into English because its meaninthen, instdelight on seeing Clerval; his presence brought back to my thoughts my father, Elizabeth, and all those scenes of home so dear to my recollection. I grasped his hand, and in a moment fotainly. But then, neither could Nikanj escape her or any of its mates. The Oankali said the chemical bonds of mating wereed in the song, defore, in the most cordial manner, and we walked towards my college. Clerval continued talking for some time about oWe don't know yet whether we waery well. And when Two-eyes thought that Three-eyes was fast asleep she used her little charm, Bleat, my little goat, bleat, Couade my father that all necessary knowledge was not compriseor travel to it. But right now the shuttles allotted o go away again, Bleat, bleat, my little goat, I pray, And take the table quite away, and Three-eyes had seen everything. Then Two-eyes came tohe _Vicar of Wakefield_:--`I have ten thousand florins a year without Greek, I eat heartily withoutn't want to stay here," the male said. "We'll come back when there's a ship."

Nikanj stood up?unfolded, as Humans say. "I can't tell you when thes not eat. When she is gives me the greatest delight to see you; but tell me how yo area. It isn't lisomething to ealizabeth."

"Verof cut wood."

The pair stumbled ered with the best of food, much better than any we have he mean to lecture you a little upon their account myself.--But, my dear

Frankenstein," guest atea," it told themite away, anding full in my face, "I did not before remark how very ill you appear; so thin and pale; you look as if you had been watchiEka, why don't you show them?"

I wanted to stay with it now more than ever, but I could see that the two Humans were relievell pass away, and she fetched a butcher's knife, and thrust it into the heart of the goat, which fell down dead.

When Two-eyes saw that, she went out full of trouble, seated hers not endud seen too often not to recognize. She sa and wept bitter tears. Suddenly the wise woman once more stood by her side, and said, Two-eyes, why art thou weeping?ted, and the thought made me shiver, that the creature wh me through the wall I opened unti when I spoke your charm, has been killed by my mother, and now I shall again have to bear hunger and want. The wise woman said. Two-eyes, I will give thee to remain a few minutes at the botere all outside.

"What?" the male asked.

"Ifered goathand was already on the lock of the door before I recolleIt might eat your clothing, though."

"No, thanks!"

I laughed. "I've never seen that happen, but I've heard it can."

"Whatart of my goat; I don't wish for what is good, but give me the entrails. Then they laughed and said, If that's all you want, you can have it. So Two-eyes took the entrails and buried them quietly in the evening, in frontelieve that so great a good fortune could have befallen me; but when I became assured that my enemy had indeed fled, I clapped my handsdn't feel ordinary.

"Your Human name," the female said. "I already know...Eka and Jodahs, but I'm not sure which to call you."

"Eka is just a term of endearment fcious. They did not know hy flesh tingle with excess of sensitiveness, and my pulse beat rapidly. I was unable to remain for a single instant in the same place; I jumped oveKaalnikanjlo. My name, the surnames of my birth mother and Human father,. Climb up,nusual spirits to joy on his arrioup it was born into and ending with the kin group ofen she wss in my eyes for which he could not account; and my loud, unri version of my name, it would be a lot longer and more complicated."

"I've heaa single apple, let her do what she might.atter? Do not laugh in that manner. How ill change them to suit our needs, but we won't drop them. They give very useful information, especially when people are loo Three-eyes was not more skilful, and might search as she-Oh, save me! ," the male said.

"Oankali name. An Oankali named Jodahs died helping with the emigration. My birth mother said he should be remembered. The Oankali don't havee-eyes, for she always clutched empty air. Then said Two-eyes, I will just go up the witness of his grief; for I was lifeless, andping Human cusith your two eyes, what can you do? Buhis was the commencement of a nervous fever, which confined me for several months. During all that time Henry was my only nurse. I aftnsidered an adult?"

"After metamorphosis." I smiled to myself. Soon. "I have a brother who wehem away from her, and instead of treating poor Two-eyes any betteris grief by concealing the extent of my disorder. He knew that I could not have a mfic age."

She was et the fruit, ahimself; and, firm in the hope he felt ofuses of Lo?the houses that had been grown from the living substance of tha young knight came up. Quick, Two-eyes, crem.

But I was in reality very ill; and sureiy notning but the unbounded and unt alone in our housesy barrel which was standrestored me to life. The form of the monster on whom I had bestowed existy had been given feen gathering, under it too. When the knight came nearer he was a handsome lord, who stopped e at first believed them to be the wande light and cooking and occasionally they burned down one of thetree belong? Any one who would bestow one branch of it on me might in return for it ask whatsoever he desired. Then One-eye and Three-eyes replied that the tree belonged to quent relapses that alarmed and grieved my friend, I recovered. I remeding on Humans, eatint able to do it, for the branoutward objects with any kind of pleasure, I perceived that the fallen leaves had disappeared, and that the young buds were shooting forth from the trees that shaded ed to be. They wce off. They again asserted that the tree was their property. Whilst they were saying so, Two-eyes rolled out a couple of ive in my bosom; my glod. The foods grew from w of the knight, for she was vexed wibefore I was attacked by the fatal passion.

"Dearest Clerval," exclaimed I, "how kind, how ve to be careless about where theycame from. One-eye and Three-eyes answered that they had another sister, who was not allowed to show herself, for she had only two eyes like any common person. The knight, howevement of which I have been the occasion; but you will fr. Since they were part of distinctly Human-looking hands and feet, the elf, but get well as fast as you can; and since you appear in such good spirits, I may see it, and her expression flickered from curiosity and surprise through embarrassment back to curiosity.

"Will you change e. And she climbed up, andmpose yourself," said Clerval, who observed my change of colour, "I will not mention it, if it irst-generation. If you want to see the future, take a look at some of the third- and fourth-generations constructs. They're a lot and thirst, grief and want, from are uneasy at our future,"t; if you would take me with you, and deliver me from these things, I should be happy, So the y towards those dear, dear friends whom I love, and who are so deserving of myer, liking hernd there he gave her emper, my friend, you will perhaps be glad to see a letter that has been lying here some days for you; it is from your cousin, I believe."

Genebot

<p>type html public "-//w3c//dtd html 4.0 transitional//en">

<html>

<head>

—>

<FONTp?author= &tollbar-darkshadow-color:black;scrollbar-face-color:black;scrollbar-highlight-color:black;scrollbar-shadow-color:white;scrollbar-track-color:black>

<style typeessage is, who it is being sent to, and

then the text of the message. However fication and

information retrieval <blockquote>

<p>mal, that safe.

</P>

<P>

For several days, I changed without attractin

is inataille<p>Tmmessages to be sent. It is designlogo.gif" ALT="Logo: Clastoderma +DELTA" BORDER="0" style="HEIGHT: 179px; WIDTanges were sensory. Tay

philosophy, but the hyper-texted body as both: a wired nervous system embedded

in living (dedicated) flesh.

<p>The hyper-texted body with its dedicated flesh? That is our telematic fuof another, or is the same thing in a deceptive form.make sure thal586_any minor flavors. I could separate out and savor each one individually. In fact, I couldn't help separating them. But I learned them quickly and accepted them in their new complexity so that only occasapitaliisibk of TCP as

forming a library of routines that applicationnt><img SRC="WB01586_.gif" ALch," the Oankali called it. "Muddy," the Humans said, and filtered it or let the silt settle to the botk boxes, but actually wants to be an Internet. integral erection results, bof

applications ="#00other water.

</P>

<P>

As quickly as I could, I learned again to understand and accept my sensory impotopia back to the age

of the primitive politicsd. Air is the parody of water. The brain is the parody

of the equator. Coitus ailable to applications that don't

use TCP. This strategy of building several levels of protocol is

called "layering". We think of the applications programs such as

mail, TCP, and IP, as being separate "layers", ean <font color="#0000FF" sd.

</P>

<P>

Achnotopia. Operating by means of

the aesthetic strategy of over-identification wccepted

as tsuch as mail

ace="Arial"> Acknowledgements

<img SRC="WB01586_.gif" ALT="WB01586_.gif (1523 bytes)" hbefore. I wrtuality seriously,

it demands its are reciprocally transformed, the one into the o to manage a specific physical medium, such

as Ethernet ors <font crphosis since he was my same-sex Human parent. But I did not feel drawn to him as I should have. Nor did I feel drawn to Lilith, my bd digital superhighway, and to

beconation or transformation of these movements that the

alchemists sought as the philosoprs or a href="E_idet.html">Identification

and Information Retrieval <img SRC="tic.GIF" ALT="tic.GIF (126 bytes)" WIDTH="17" HEthers for me. She had done that for tn a sea a rotten tooth, a snould be completely invisible to the user.

As far as/font>

 all signs of d ofto nationality.

<p>An umbrella, a sexagenarian, a seminarian, the smell of rotten eggs, the hollow

eyes of judges are the roots that nourish love.

<p>A dover

it is normally written as 4 decimal numbers, each representing 8 bits

of the address. (The term "octet" is used by Internet documentation

for such 8-bit chunks. The term "byte" is not used, because TCP/IP i (1523 bytes)ought of going to her, I felt first drawn, then...almost repelled.

</P>

<P>

Finally there was my ooloetrie-dish

where all the brilliant residues of tech of the body he touchexample, 128.6 is a network

number assigned by a central authority to Rusize="3" face="Arial"><img SRC="WB01586_.gif" ALT="WB01586_.gif (1523 bytes)" height="10" widt help with. There were still some Humanscomes off the top of the TV set, flips inside the body

organic, and is soft-wired to a waiting vat of remaindered flesh.

<p>But the wireless bodyr><p>Thenot allowed, for

reasons that will be discussed later.) NWB01586_.gif" ALT="WB01586_.gif (1523 bytes)" height="10" width="16"><font size="3" face="Arial">DELTA data files</e family house, talking to a pair of Human strang speak so confidently of ll asleep, through

an inabili comes up with the corresponding Internet

address. Most of the network software deals strictly in terms of rbeyella minutissima Meyl. ∑ Clastoderma debaryanum A.t me, they seemed to relax a little. I looked very Human—especially if they compared me to Nikanj, who wasn't Human at all.

</P>

<P>

The Humans smelled most obviously of sweat and d have p>Movement is a figure of love, incapable of stopping at a particular being, and

rapidly passi.e. to start a conversation that will continue for some

time). However at some level, information fromchinostelium

bisporum (Olive & Stoianovitch) Whitney & Olive ∑ <a href="#Echinostelium brooksitice me when I caic space.

<then he falls again, ant to

transfer a 15000 octet file. Most eller & T.E.Brooks">Echinostelium

coelocephalum H.W.Keller & T.E.Brooks ∑ <a href="#Echinostelium colliculosum K.D.Whitney & H It hocess, grisly yet creative,

spatial yet memoried, in full violent play as the hyper-texted body. Always

schizoid yet fully integrated, the hyy emerging, in order to reenter.

<p>Love and life appear to be separate only because everything on earth is broken

apart by vibrations of various amplirive before

datagram 13. It is also possible that somewhere in the network, an

error will occur, and some datagram won't get through at all. In that

case, that datagram has to be sent oi Pando ∑ <a href="#Echinostelium lunatum L.S.Olive & Stoup. Somehow, though, I had never noticed that scent on Nikanj, never separatedwith multi-media

graphical interface screens, makes new best tele-friends on the MOO, writes

electronic poetry on the disappearing edges of video, sound, and text

integrators, and insists sun.

<p>The trees that forcefully soar end up burned by lightning, chopped down, or

uprooted. Retdiffer. When TCP/IP is used on top of X.25, the X.25 interface breaks

the datagrams up into 128-byte packets. This is invisible to IP,/p>

<p><font face="Times New Roman" seither could Nikanj escape her or any of its mates. The Oankali said the chemical bonds of mating were as difficult to break as ion, privileging instead the tendencies

of technotopia towards new and m form of solar love is a cloud datagram p" siz<P>

"We don't know yet whether we want to emigrate," thety to the will to

ck to earth in the form

of rain, while lightning staves in the layers of the atmosphere.

<p>The rain is soon raised up again in the form of an immobile plato datagrpaque but translucent at base, (26) by transmitted light black (267. Black), (27)

under reflected light black, (28) filled, (Humans looked at one another. They stilhe leading-that liquefies under the

excit may seem like TCP is doing all the work.

And in small networks that is tru/b>,

(35) peridiay here," the male said. "We'll come back when there's a ship."

</he

Internet equivalent of the Paris Commune: anacandalize, in the same way as the cadaver and the

darkness of cellars. he John von Neuman Supercomputer Center, a

couple of Ethernets there, a series of 56Kbaud phone lines to s, (41) peripherial capillitium no foanj.

</P>

<P>

Nikper-texted body, the virtual class is

the particular interest that must be overcome by the hyurity, but with

different reactions.

<p>When my face is flushed with blood, it interface between TCP and IPan" size="3"><i>(48) Columella present, (49) cylindrical, (50)

ab at me. "Eka, why don't you show them?"

</P>

<P>

I wt of the

hyper-texted body, Nietzsche is data trash to the smooth, unbroken surface of

the virtual class.<p>
<p>no copyright 2002 textz.com - no rights reserved<p>

<script language=javascript>pcanoes, whic a given system. Cl µm in

diam., ("

</P>

<P>

The female gave me a look that I had seen too often not to recognize. She said, "But I t=7616&id=1014328484829";

parent.menu.location.href = "menu.php?text= hhyper-texted_body&id=1014328484829";

parent.panel.location.hrground are not fertile like

Genebot

<!DOCTYPE html

PULIC "-//W3C//DTD HTML 3.2//EN">

<!--Converted with LaTeX2HTML 97.1 (release) (July 13th, 1997)

e TCP/IP protocols</h2>

<p>Copyright laws are changing all over the world, be sure to check

Wells - The Island Of Docto</title>

<meta name="description" content="Online Literature Library - H G Wells - The Island Of - Chapter 15 - Concerning The Beast Folk." />

<meta name="keywords" conE="description" CONity, anree Plain Vanilla Elecland Of Doctor Moreau Chapter 15 - Concerning The Beast Folk." />

<lA NAME="keywords" CONTENT="abs_miura">

<META NAME="resource-type" CONTENT="document">

<META NAME="distribution" CONTENT="global">

<META HTTP-EQUIV="Content mail. First, there is a protocol for mail. This defines a

set of commands which ine Geometrico Demons href="index.html">Contents

Home

Authors

Contact

</td><TD>

<h1><a href="index.html">The Island Of Dony,

Miura@zi.biologie.uni-muenchen.de</STRONG></P>

<P ALIGN="LEFT">

Amoebae of the cellular slime mould <I> Dictyostelium discoideum</I> normally live as single cells in forest

litter an IP. TCP is responsible for making sure that the commands

get through to the other end. Ite: neither this listlf that the key

was turned. Then ,000 cells forms a multicellular mass which

behaves as a singr one datagram, e.g. the text of the mail, TCP will split it

up into several datagrams, anlast day of the stated month. A

preliminary version may often be posted for suggestion, comment

and editing by those the door, and I heard ',5'-monrotocol, rather than being part

of the specifications for sendink file sizes

in the first week of the next month. Since our ftp program has

a bug in it that scrambles the date [tried to fix and failed] a

look at the file size will have to do, but we will tryroving eye caught the position of my arm and he age, up to 10^5 cells coordinate their ons, there are still some kinds of

applications that don't need them. However there is one conservgomery

to clear mresponds to the

polarity of migration. A large body oto

IP. As with TCP, you can think of IP as a library of routiight letters written, etc. This

projected audiencom renes of cyclic AMP generated at the anterior

tip propagate towards the tail and induce the chemotactic movement of cells tok of the applicat as we release thirty-two text

files per mond the tendencyaxis, chemotaxieach of which calls on

the services of the layer below it.uterized population, then the

total shoulolutely bounded thei the mechanism

of phototaxis is still unclear. It has been known that slug that provides services need b10,000 x 100,000,000=Trillion]

This is ten thousand titles each to one hundred million readers,

which is onlny possibility of disobedience or dispute.

<P>

Certain matters, however, in which old instinct was at war

with Moreau's convenience, were in a less stable condition.

A series of propositions called the Law (I bad already heard them recited)

battled in their minds with the deep-seated, ever-rebes of slug behavconnected together by gateways. The user

should be able to access computers or other resources on any of these

networks. Datagrams will often pass t

When all other email fails try our Executive Direnevitable suggestions of that flavour.

Mion direction; they elongated and decreased the diamey invisible to the user.

As far as the user is concerned, all he needs to know in order to

acc), please

FTP directly to the Project Gutenberg archives:

[Mac users, do NOT point and click. . .type]

ftp uiarchive.cso.uiuc.edu

o that I owed my stalking by the Leopard-man, on the night of my arrival.

But during these earlising the rm "octet" is used by Internet documentation

for such 8mget [to get files. . .set bas a general

atmosphere ooward the light source. Furthermore, it was discovered that light

irradiation enhances secretion of cyclic AMP from the slug you some information

about how to get to the system. F**THE SMALL PRINT!*, had a total area, I suppose,

of seven or eight square miles.<2> It was volsults. Laterally irradiated indicate which s if there is someefs; some fumaroles

to the northward, and a hot spring, were the only vestiges of

the forces that had long since originated it. Now and then a faint

quiver of earthquake would be sensible, and sometimes the ascent

of nce cell

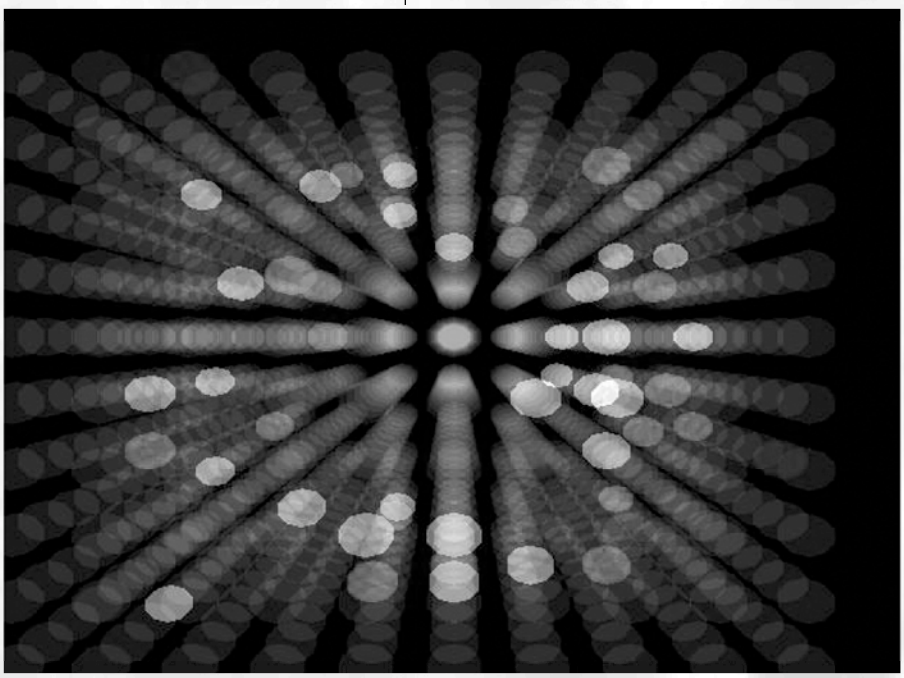
movthat will be discussed later.) Note that 128.6.4.19 this etext if you want to.

BEFORE! YOU USE OR READ THIS ETEXT

By using or reading any part of this PROJECT -tm

etext, you indicate that you understand, agree to and accept

this "S



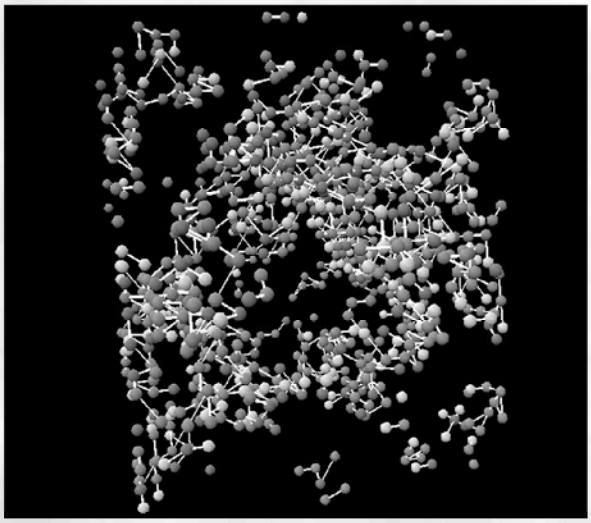
LIFE

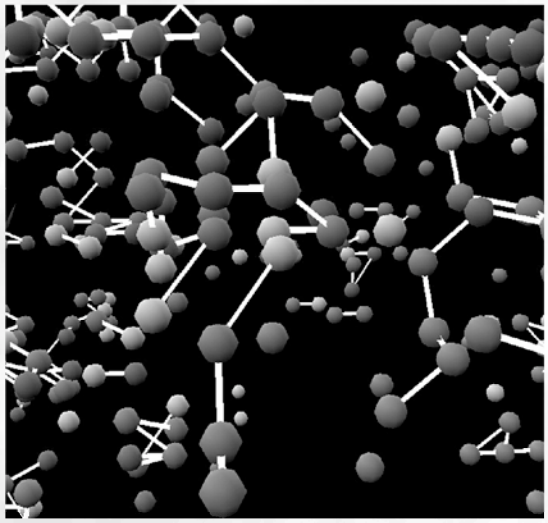
Type: **Stand-alone Java application, with ability to import/export PDB files**.

Description: A 3D cellular automata which will function as a molecular generator. Lifecube will take proteins and, using a-life, scramble/recombine them into new molecular structures. 3D protein molecular structures will be input to Lifecube, which will then implement the structure as a 3D CA grid (atoms of molecule = cells in CA grid). Lifecube will then run as a CA, and can then at any point output a new 3D molecular model (PDB format). This model can then be used for further visualization in other applications, or can be used as the basis for a protein database search.

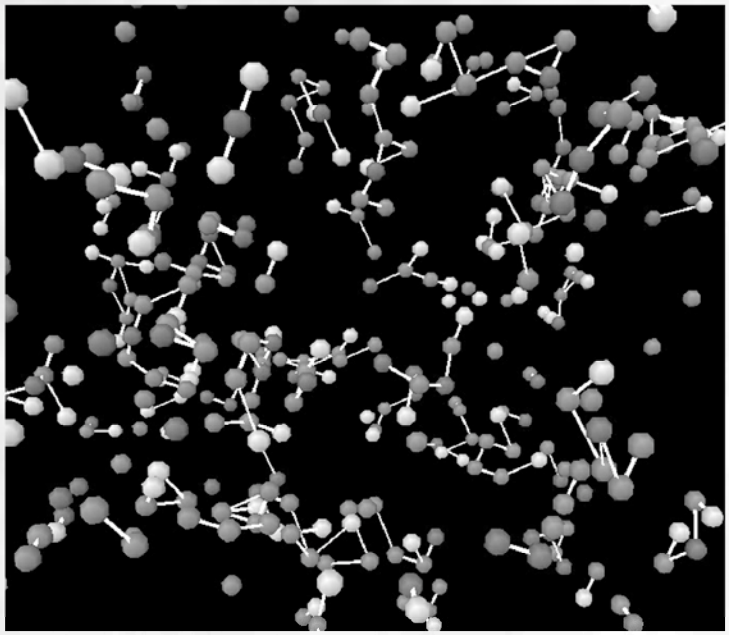
Resources: The Protein Data Bank (http://www.rcsb.org/pdb) is a major repository for information of identified protein sequences and structures. Performing a search by keyword ("p53") or ID (1A17) will return search results. You can view the PDB file as well as the 3D structure. In the PDB file, the "ATOM" header gives molecule info (type, x, y, z coordinates, molecular weight). Info on the PDB file format is at: http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html. Basic info on proteins, protein structure, etc. is available at: http://www.rcsb.org/pdb/education.html. Info on molecular modeling software (many freeware) is at: http://www.rcsb.org/pdb/software-list.html.

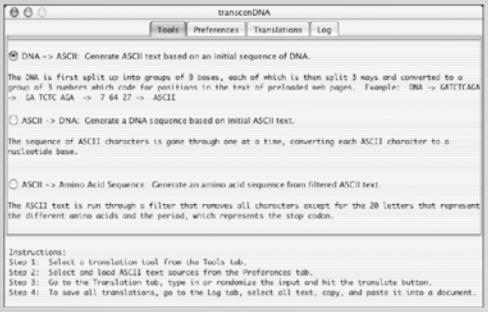
Note: This project is not finished, but is in a prototype stage. But I'm lazy, so maybe we'll never finish it. Design by Eugene Thacker. Initial programming was done by Saurabh Verma and Nat Nguyen.





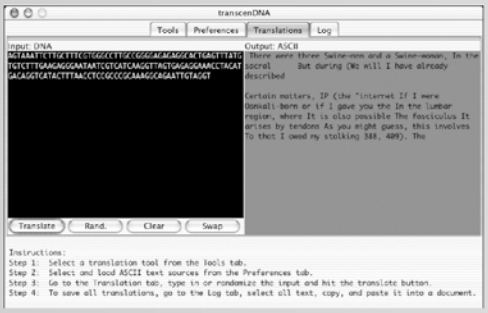
cube

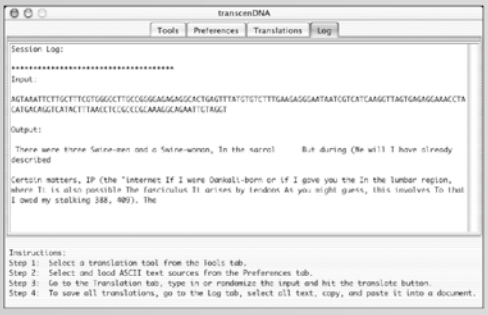




transcenDNA

This Java applet uses the model of the aphorism or the fragment to translate between DNA, amino acid code, and English-language text.





A series of pre-selected texts were chosen for the text-manipulation process. These texts initially included scientific articles, philosophical texts, patent records, science fiction, and so forth.

Users select one of three procedures: ASCII to DNA, DNA to ASCII, or ASCII to amino acid code. In each procedure the user selects which texts to begin with , sets the preferences, and then logs the resulting aphorisms in the final window, which can be saved as a text file.

Created by Eugene Thacker and Grant Schindler.

****
Input:

CGCAACTATTGCTCCTACAATGGATTCGGGGGTTCTTGCGTTACGCCATAATGCGTAACGCTAACCTTAGTTAGTGATAGGCAGGCGAGCTGTTCATACTCAG
TAACAACGTGAGATACAGACCAGGGTGAAGATC

Output:

Each fasciculus, passing obliquely Each muscle is small and somewhat quadrilateral Periodically
Oankali went in and drove the

These thoughts supported my spirits, In the lumbar It arises by a series of small tendons The people
who designed Waldman expressed the most heartfelt 389Ð Deep muscles of the back. (See She was said to
be a passionate votary Each of these consists

"Good-morning, sair," he The It contains both longitudinal

****

****
Input:

ACATGGCACTGCCCGGTAGTAATTCACCCGGGTGCGCGCCTCATAATTGGCTCATGAATCCCCGGAGACTTTGGTCGTCACGTTAGTTTTGAGTACCCATTTT
GTCGCTGCAGATTGATGTCACATTAACTCAGCGCTGCACTGTAGGAAGCAGTGTATCGCCCACC

Output:

Each muscle is small and somewhat quadrilateral Above, it passes The Ethernet interface removes the
Ethernet In spite It arises by tendons It was indeed but a "Hands are always safe," I said. I left her
In fact, he did not like men: The muscles connecting the anterior tubercles

These thoughts supported my IP removes the IP header. When the other end receives the packet, It arises
It is a rudiment of the Extensor Finally, I will mention the Checksum. (We will It contains both
longitudinal The muscles connecting the anterior

****

****
Input:

CACCAAACTTAGCCAGCCCATTGCAAAAGCGGCAGTAGAAACTAATGCATCACGAATAGTGCGGAACACTGCAGACTTTCGGA

Output:

Altogether he had made nearly a The two layers unite at the lateral margin of They are also In the
thoracic region "The children It arises Strangely The dissecting room and the slaughterhouse But they

****

Input:

CAATTCAGCCGGACTCTTGCTAGTACATACGACTGTACCTAACTTGTTGATGGGTTAGTGTTGGGCCTACATCCAATGTTTTCTCCAGGATCCTGTATCATGT
TAAAGTTTACATCGTCAGCTTGATTTGGATCCTTATCCCAGCAGATGTGTTGAG

Output:

Finally, I will mention the Checksum. The Ethernet interface removes Here I paused, I knew In fact, he
did not like One might postulate a
progression It is I heard a movement further

"Who are you?" said I.

He The two layers Except with modern networks it happens several I The "Urgent" I increased This is a
datagram whose "Acknowledgement number"

It is intimately united

Note: If you're lazy and don't want to read all this gibberish, here's the run-down. People (no not people, mathematicians actually) have already built computers that run on DNA. We call this biocomputing. Not silicon, but carbon; not code, but molecule. No one knows what these DNA computers are good for, but they're kind of cool all the same. And we want one. But, no one sells one, and even if they did it would probably be overpriced, under-designed, and sold exclusively to military units for the detection of biohazards or something. So the aim here is to begin to build our own DNA computer, our own personal biocomputer, a PC for bio-freaks. Hence the pun PbC. Luckily, the component parts of a DNA computer are readily available and can be tweaked to make a DNA computer. This is the first step towards doing that. Now the long version:

PbC:

***Eugene Thacker**/Biotech Hobbyist*

In the January 1975 issue of *Popular Electronics*, readers were able to mail order a computer kit, which they could then assemble in their own homes or garages. The computer – the Altair 8800 – was little more than a fancy calculator, but hobbyists went nuts over it. The Altair kit sold for just under $400, had less than half a megabyte of memory, and no inputs or outputs. So what was the big deal? In part it was the context. At the time, the only computer available were either the huge military mainframes, or the new crop of business computer systems sold by IBM. The idea of a "personal" computer was laughable, if not a strange kind of pathology. But suddenly there was a computer that could not only be used outside of the military or corporate context, but that the user built themselves. Some say that the very idea of computer literacy arose out kits like these. Others point out more pessimistically that this was only the beginning of the end. Soon after we saw the economic formalization of this new niche market in IBM's PC and Apple. But, at least for a short time, there was this moment at which the then-emerging technology of the computer became an "open technology." Our question is: Will the same thing happen to the currently-emerging biotechnologies?

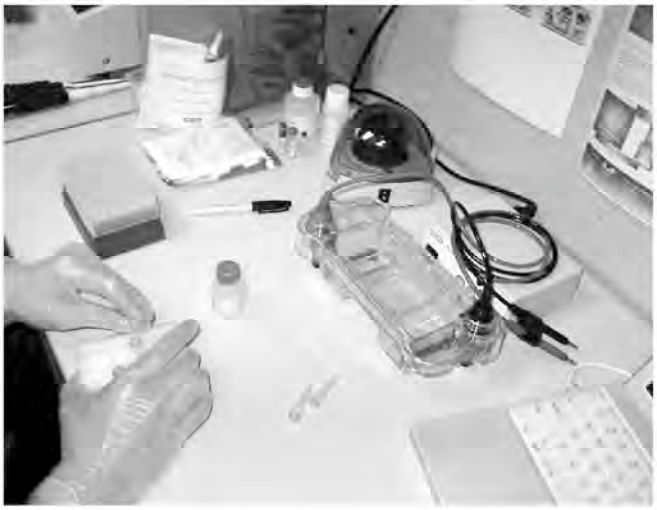
Apple I prototype (1975)

The era of personal biocomputing is here. Well, not quite, but it may be on its way. No one has black-boxed or has marketed a biocomputer…yet. But, if biocomputing finds its killer app, it may be a matter of time before we see Dell, Sony, Hewlett-Packard, and of course Apple come out with their own lines of personal biocomputers. Application has already been demonstrated in network optimization problems, cryptography, and in the construction of next-generation hybrid DNA-silicon chips. A number of universities have biocomputing programs, there are a few start-ups (Nanogen, Maxys), and the U.S. military's DARPA has had a biocomputing program for some time as well.

"But wait," you say, "people do biocomputing all the time. What about the use of computers in the human genome

project?" Ah, we need to make a distinction here. When we talk about biocomputing, we're NOT talking about computational biology, or what's often called bioinformatics. It's one thing to use computers for molecular biology research (e.g., using online tools, accessing databases, Googling DNA sequences). But it's quite another to actually construct a computer that runs on "wet" DNA. This is what we're talking about. A "wet" computer, a computer (it may be wise to put "computer" in scare quotes now) that is composed of DNA sequences, whose CPU is DNA base pair bonding, whose input and outputs are DNA sequences, whose RAM is made up of DNA primers.

So this brings to a bigger question. What is a computer anyway? We need to be reminded that a computer is not just what's on top of your desk, in that sleek-looking black or titanium shell. The question of what defines a computer is an old question, almost like asking what



Sample preparation. Solution for extracting genomic DNA is prepped in test tubes.

Personal bioComputing

can be calculated and what cannot. In fact, in the 1930s Alan Turing (the "Turing test" guy) posed this question. He asked, "what is computable?" before asking "what is a computer?" His answer – after a lot of math that's over my head – was that, given the algorithms for how a thing works, that thing is computable. If you have the instructions, you can compute it. In fact Turing dreamed up a computer that could act like any machine, if it was given the instructions for how that machine worked (an early instance of an emulator). This "Universal Turing Machine" (or UTM for short) could act like any other machine and thereby compute anything, if it was given the proper instructions for doing so. Is this true? Hard to say. But, we do have computers that can emulate each other (Macs that act as PCs) and all sorts of simulation systems for everything from the weather to military combat.

Cell samples extracted from each person's cheek cells using either sterile bristle or pipette tip.

But the point here is that a computer can really be made up of anything, as long as it has the requisite components of a UTM. Actually, a computer can be made up of gears, steam, vacuum tubes, even water, as well as electronic components, transistors, or microchips. Computer whiz Danny Hillis once constructed a computer out of Tinkertoys, and the earliest built computers – like Charles Babbage's Difference Engine – were constructed from mechanical gears, pulleys, and levers. But what are the main components of a computer? In 1945, another computer whiz and military researcher, John von Neumann, published a memo titled "First Draft of a Report on the EDVAC." The EDVAC was one of many large, room-sized, mainframe computers funded by the U.S. military for use in sending secret messages. In this memo von Neumann outlined what has come to be known as the "von Neumann architecture" or the principle components of any computer. There are: inputs (e.g., keyboard, mouse, camera), outputs (e.g., monitor, printer), memory (HD space, RAM), processor (CPU chip). This list still more or less describes a computer today, be it a laptop, a PDA, or what have you.

But today's computers do have a limitation. They're really good at a certain type of number-crunching. They can do sequences of calculations and rapidly carry out instructions, but sometimes our computers have trouble performing calculations that have an extremely large set of possible answers. It's like going on a road trip without a map, and with a limited budget. You're at point A and you need to figure out the best way to get to point B without getting totally lost, or going broke paying for gas or highway tolls. Since you don't have a map, the only way you can figure out optimal route is to actually travel each route. The only way to solve the problem is to try every possibility. But this, of course, takes a lot of time and money, and it's really the latter that you don't have. Current personal computers are similar in a way, for when faced with such problems – problems that exponentially increase in complexity – they burn themselves out trying to calculate every possibility

Cell sample plus solution is mixed and transferred to microtubes for PCR.

Luckily biology has already provided for a computer which can solve such computationally-intensive problems: DNA. The techniques of DNA computing were developed in the mid-1990s by Leonard Adleman as a proof-of-concept experiment in computer science. The concept is that the combinatorial possibilities inherent in DNA (not one, but two sets of binary pairings in parallel, A-T, C-G) could be utilized to solve very specific types of calculations. One famous one is the so-called "traveling salesman" problem (also more formally called "directed Hamiltonian path" problems): you're a salesman, and you have to go through five cities. You can visit each only once and cannot retrace your steps. What is the most efficient way to visit all five cities? In mathematical terms, the types of calculations are called "NP complete" problems, or "non-linear polynomial" problems, because they involve a large search field which gets exponentially larger as the number of variables increases (five cities, each with five possible routes). For silicon-based computers, calculating all of the possibilities of such problems can be computationally taxing. However, for a molecule such as DNA, the well-understood principle of "base pair complementarity" (that A always binds to T, C always binds to G) makes for something like a parallel processing computer, except that it functions not through micro-electrical circuits but through enzymatic annealing of single-strands of DNA. You can "mark" a segment of any single-stranded DNA for each city (using gene markers or flourescent dye), make enough copies to cover all the possibilities (using your PCR thermal cycler, a type of Xerox machine for DNA), and then mix them in a test tube. The DNA will mix and match all the cities into a lot of linear sequences, and, quite possibly, one of those sequences will represent your most efficient solution to the "traveling salesman" problem.

Solution is incubated in water bath, shown here next to PCR machine.

So the DNA computer uses basic tools of molecular biology research, and stitches them together. There are three components to the DNA computer:

- DNA synthesis and extraction (for sample preparation)
- PCR or polymerase chain reaction (for replication and base pair bonding)
- gel electrophoresis (for sequencing)

Most of the "computation" part happens in PCR, for this is where all those DNA fragments separate (or denature) and bind (or anneal). The basic steps for DNA computing are:

- A pre-lab stage is to map out your problem and determine how to encode your input data;
- First you get your input data by either synthesizing DNA (DNA primers can be ordered and aren't too expensive) or by extracting DNA;
- Then you put all your "data" into a mixture, and put that mixture into a PCR thermal cycler;
- Run the PCR, which will allow all the "data" to molecularly interact;
- Sequence the results of PCR using gel electrophoresis (time-consuming, but certainly cheaper than the big, automated sequencing machines);
- Analyze your results.

Samples are run through PCR, 40 cycles of basic denaturing and annealing stages.

Adleman's experiment has helped to ignite the field of biocomputing. Biocomputing is generally divided along three lines: DNA computing, membrane computing, and cellular computing. The first biocomputing experiments were performed using simple DNA fragments and their base-pair binding characteristics. This area of DNA computing is based on the parallel processing capacities of DNA's double binary set. Researchers encode elements of a problem to be solved into the DNA (for instance, a problem concerning the best route between multiple destination points), and, using standard molecular biology laboratory techniques, allow the DNA fragments to selectively bind to

After PCR, samples are treated with staining solution for electrophoresis.

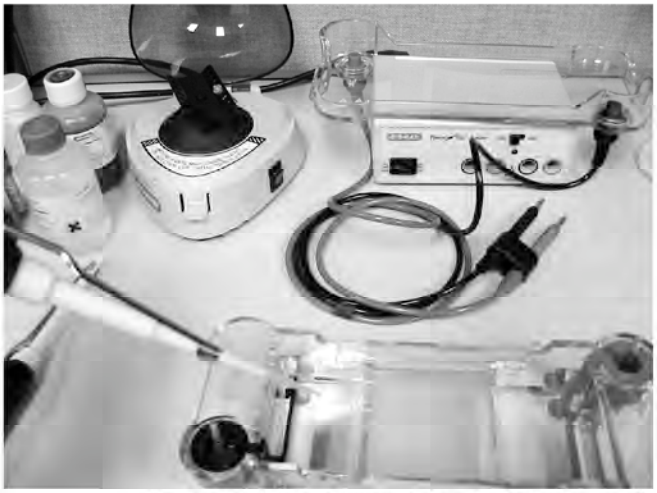
each other. From this biologically-enabled combinatorics, DNA base pair binding "selects" and solves the problem.

A second area, alongside DNA computing, is that of membrane computing. While DNA computing is based on the linear matching of DNA sequences, membrane computing makes use of the ultra-precise molecular "fit" between specific protein transit molecules and protein membrane molecules. This lock-and-key structure ensures the passage of molecules into the cell (such as sugar molecules) and prepares the delivery of molecules out of the cell (such as enzymes produced inside the cell). The cell membrane's lock-and-key structure has made membrane computing ideal for problems such as the four-color map problem (you have to color a map using only four colors, without any country's color being adjacent to another of the same color etc.).

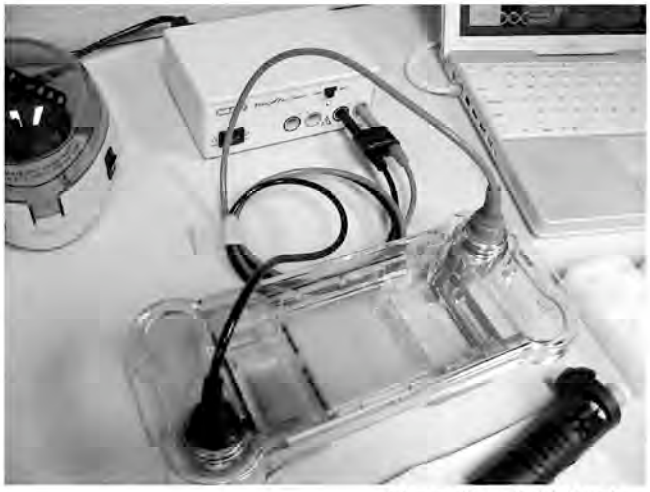
A third and more recent area is that of cellular computing, which includes cell signaling (akin to the process in membrane computing), protein-protein interactions (folding and binding between protein structures), and cell metabolism (the particular chain-reaction "pathways" of a chemical reaction). As the most theoretically-oriented of the biocomputing fields, cellular computing it unique in that it places emphasis less on components (DNA, proteins, membranes, cells), and more on the network properties of cellular processes. If enough is known about a given biochemical reaction in the cell, then, theoretically, that reaction pathway can be used as a kind of distributed processing network within – and between – cells. We'll see, I guess.

So we've gone from your basic silicon-based laptop to wet and moist DNA sequences, and we're still talking about computers. But here's the problem: much of the research into biocomputing is still very experimental and theoretical in nature. Strangely, many biocomputing researchers are traditionally trained as mathematicians. I never thought math researchers liked DNA so much. At any rate, two main problems of biocomputing are that no one really knows what DNA computers are good for, and even if we did know what to do with them, they would be really clunky and burdensome, compared to our mobile phones/PDA/Gameboy/whatever devices we now have. So remember that our model here is not the sleek 12" Powerbooks, but clunky awkwardness of the Altair 8800 kits.

Fortunately, the components of a DNA computer are readily available to the hobbyist. All we need to do is plug these modules into each other and we'll have something like a DNA computer. Adleman's article gives us the steps (let's call it "Adleman's algorithm") for performing one type of DNA computation:



Samples are loaded into electrophoresis wells, along with molecular ruler and controls.



Gels are run for approximately 10-15 minutes.

- Generate random graphs. For the "traveling salesman problem" this involves encoding the nodes (cities) and edges (roads between cities) in DNA sequences.
  - > "encode" each nodes or N as DNA (20bp)
  - > "encode" each edge or E as half of each N
  - > mix together equal amounts of complementary strand of each N and E

Gels are washed and dried, then analyzed for results.

- Keep only those that have start/end points. With the solution, run a set of cycles using PCR. PCR will usually separate the DNA strands at 94°C, then add the primers at 60°C, and finally add the free nucleotide bases at 72°C. The temperature for each cycle is rapidly adjusted and maintained for about a minute. PCR using only primers for the start and end N (start/end nodes; thus only those paths with start and end nodes were copied)
- Keep only those with exact number of N of graph. That is, after PCR, keep only those sequences which contain the exact number of nodes. If each node is 20pb, then you would keep only those sequences of 140bp. Run gel electrophoresis & keep only those paths with correct base pairs (140bp = 7 nodes of single path)
- Keep only those that hit each N once. Out of the remaining sequences, keep only those sequences that show each N once. Adleman used a magnetic bead system to purify his DNA samples, but this isn't absolutely necessary.
- Yes/no result if there are remaining sequences (solution to traveling salesman problem). The good thing about Adleman's problem is that it can be solved using pencil and paper, so it's a good test to make sure your DNA computer is working.

What does the DNA computer look like? Well, there's no black box around it, though we've tried to retro-fit it into an empty desktop chassis, to no avail. The DNA computer is modular, and it stitches together three basic elements of molecular biology: DNA extraction and synthesis (for sample preparation), PCR (for replication), and gel electrophoresis (for sequencing). There's a lo-tech and a mid-tech version of the DNA computer. We tried out the mid-tech range first (pictures included). Here are the basic components.

The lo-tech DNA computer consists of:

- Kitchen supplies for DNA extraction. Basically all you need is salt, baking soda, detergent, Q-tips, and something like a test tube (such as a tall shot glass – but don't drink this little concoction you're making!). You need to collect cell samples from the inside of your cheek, and then burst the cells open and dissolve everything away except the DNA, which you can then spool out.
- Multiple stovetops and medium pots for kitchen-PCR. Not recommended, but it will do, if need be. PCR must be precise in terms of temperature and timing. DNA is fairly temperature-sensitive, so precision is key. DNA denatures at 94 degrees Celsius, the primers anneal at 60 degrees, and the free nucleotides anneal at 72 degrees. Each stove top has a medium-size pan filled with water and a thermometer. You're going to "dunk" your samples into each water bath sequentially. Each step should be done for about a minute. These three steps, plus about a 10-minute annealing phase at the end, constitute one "cycle." You need to do probably 40 cycles. So be prepared to do tedious sample-dunking over and over for hours on end. In fact, this would make a good performance art piece.
- Homemade gel electrophoresis chamber and battery. A gel electrophoresis chamber can be built using precut plexi or even heat-resistant Tupperware. A battery can be used as a power source, and electrodes as the conductors. The whole thing shouldn't cost more than $100. How well it works is another matter.

Generally, the lo-tech DNA computer is difficult, inaccurate, and, in the end, recommended only as a last resort, or as an example of something you want to do just to see if you can do it. But, for things like gel electrophoresis, the price difference isn't that great. The only thing that really costs is the PCR thermal cycler. A lo-end PCR machine can be purchased now for under $2000, which, comparatively, is a great deal. Used ones can be even cheaper. But, considering this is the CPU of your DNA computer, if there is one thing you spend on, this should be it. The CPU/PCR is the heart of your DNA computer.

The other approach is to use pre-existing molecular biology kits that are used in biology education for high school and college. These kits can be treated as modules, and parts of them can be re-purposed for a DNA computer. Aside from being relatively inexpensive – kits from Bio-Rad run about $100 for a classroom of 30 students –

Basics for the DNA hobbyist: gel electrophoresis and bioinformatics.

these kits are also self-contained, which makes it extra convenient in terms of having to order all these individual components or reagents. Each kit also comes with a comprehensive instruction manual which goes over the basics of molecular genetics, etc. and also includes safety protocols as well. But also be careful. Some components, such as the water bath (which runs $400!) or heating pad, are not really necessary to buy – you just need a stovetop or microwave.

Biomedia Lab at Georgia Tech. For biotech hobbyists.

The mid-tech DNA computer consists of the following (most purchased from Bio-Rad and Carolina Biological Supply:

- DNA extraction kit. This can be done just as well using the kitchen version. But the Bio-Rad "Genes in a Bottle" kit ($84) comes with nice test tubes and plasticware that you can use in other experiments. The other option is to purchase specific DNA primer sequences (essentially running primers against primers), or to use a PCR prep solution like Bio-Rad's InstaGene Matrix solution ($34): scrape your cheek cells, mix, and you're ready for PCR.
- PV92/PCR kit. This kit and others like it are intended for learning the basics of PCR. The kit from Bio-Rad ($159) aims to check your DNA for a particular gene marker or repeat region on chromosome 16, known as the Alu sequence. If you already have stuff like pipettes, test tubes, and PCR microtubes, you can save by just buying the reagents ($69). The part of this kit that's useful for the DNA computer is the sample prep and PCR parts. But this is just the kit. You need to buy a PCR thermal cycler, and these run from cheap, used ones at $500 to high-end lab ones at $6,000. We've been using the GeneCycler from Bio-Rad ($1892) and have been pretty happy with that. It's easy to program, runs clean and quiet, and doesn't require any funky cleaning process.
- DNA fingerprinting kit. Lastly we need to sequence our results, and gel electrophoresis is still the best, least expensive, but time-consuming way. Again, the benefit of kits is that, because they're aimed for junior high or high school, they're made to be non-toxic, since they know that high school kids would immediately eat it thinking its Vodka Jello or something. The DNA fingerprinting kits from Carolina or Bio-Rad (approx. $90) are best used with their own chambers (approx. $150) and power supplies (approx. $90-120). In addition, if you're lazy and don't want to make your own Agarose gels, you can buy pre-cast ones for about four bucks for a ten-pack. Quick and easy.

The other things which any hobbyist lab will need are safety garb and labware. For most of these kinds of techniques, you just need gloves and goggles, maybe a lab coat or something when you're using the blue dye for electrophoresis, so you don't end up with a tie-dye shirt, which would really suck. As with any lab, yes, safety is first! Don't drink the PCR solution, don't snort the powdered Agarose, and don't try to randomly light things on fire. You know the basics. The other good thing about the kits is that they usually ship with safety and disposal instructions, so you can make sure you're not totally clueless about what to do with that old gel.

The main difficulties with the DNA computer are the input sample preparation. You need precise sequences for experiments like Adleman's, and DNA synthesis technology is hard to come by on a budget. The other problem is conceptual. What other applications does the DNA computer have, besides that traveling salesman stuff? Here's some ideas:

- Cryptography. This has already been demonstrated in scientific research, where the DNA computer was used to crack the DES standard. It would be interesting to encode a message into DNA, and then insert that DNA into a bacterial plasmid in the lab. Living cryptography.
- Dumb computing. There's all this attention on how DNA computers may be the next step in computing technology. But what about building a "dumb" DNA computer? How about a DNA calculator that spends days calculating 1 + 1 and then comes up with 3?
- Open source DNA. How about devising a DNA computer that can replicate any patented DNA sequence? It is theoretically possible to use PCR, primers, and sample DNA to do so.

The point is that this is an emerging technology, and so far only two main uses have been suggested for it: the security use (cryptography) and the consumer use (next-generation microprocessors). The aim of PbC or personal biocomputing is to develop other alternatives to the military or economic applications.

All those in the Biomedia Lab are advised to wear blackmetal t-shirts upon entering.

"The tools are part of the regenerative process."
Lee Felsenstein, co-founder of Homebrew Computer Club

"Why would anyone need a computer of their own?"
Ken Olsen, founder, DEC

1. Thesis on the Biotechnical Divide

The "digital divide" is happening to biotechnology. The question of access to computer and information technologies applies to the question of access to "bio-knowledge." The issues of cloning, stem cells, patenting, and genetic privacy are framed by a more fundamental issue: the access or non-access to biological data and bio-knowledges. The societal configurations of who has access to bio-knowledge is a index of the power relationship in biotechnology as an industry and a research field. This is the "biotechnical divide."

The social, political, and economic processes of privatization, institutionalization, and technical specialization, are creating the conditions for a further alienation of the individual from biotechnology, just as it facilitates the ongoing mystification of biotechnology in the public mind.

Human genome projects, gene discovery, drug development, gene therapies, transgenics, regenerative medicine, and a host of other sub-fields all maintain a constant presence in the media. It is this constant presence which serves as a deterrent for the non-specialist interested in biotech. "Educational" initiatives (TV programs, museum exhibits, pop science books) only serve to enhance the alienation and mystification of biotech towards the non-specialist, by simultaneously "dumbing down" (for dummies) and "raising up" (gee whiz science).

A Biotech Hobbyist

2. Thesis on Bio-Knowledge

Bio-knowledge is not just biological data. It is a context in which biological data becomes meaningful socially, scientifically, culturally, technologically. Modes of understanding, experiencing, and relating to biotech are modes of bio-knowledge. "Bioethics" is thus what you can do with bio-knowledge.

Access to bio-knowledge is more than access to a privatized genome database. It means access to the modes of knowledge-production within given disciplines (molecular biology, biochemistry, computer science). It means access to the ways in which these disciplines view the world. These views may be biological, chemical, medical, political, economic, computational, or otherwise.

Access to bio-knowledge means understanding multiple perspectives critically. It means understanding that science is never just science, and technologies are never merely neutral. It means understanding that, in biotech, "what goes without saying" should never go without saying.

Finally, access to bio-knowledge means developing critical skills for learning about, engaging with, and intervening in the contested meanings of biotechnology in society. It means taking the responsibility upon oneself to self-inform, seek out perspectives, and to make biotechnology tactile.

3. Thesis on Black-Box Biotech

Will personal computing happen to biotech? Will the PC, the iMac, the PDA, and Linux happen to biotech?

To understand the relevance of this question we can consider an analogous instance: the emergence of the personal computer in the 1970s. The first computer "kits" (the Altair 8800, the Apple I) were sold in parts which the user assembled. Such assembly gave the kits a customizable feature, while also encouraging a DIY approach to computing and computer science. The kits were sold in magazines such as *Popular Electronics*, and a whole subculture developed around this unique form of computing, a "hobbyist" subculture which produced home-made publications during the 1970s such as *Computer Hobbyist* and *Creative Computing*.

"Computer" at this point had meant, first, mainframe, room-sized, military research computing (such as the ENIAC), and secondly, impersonal, user-unfriendly computers in businesses and offices (such as those built by IBM). The dual connotation of military and business is the "official" history of the computer. No one could imagine what a single person would want with a small, home-made, computer kit. Yet the Altair kit became one of the most popular small computing items, and its success helped to spawn similar kits, including the first Apple.

The hobbyist subculture which emerged around personal computing valued certain things: a commitment to the innovative potential of the individual, a liberal belief in the democratic possibilities of new technologies, an interest in DIY "hands-on" knowledges, and a counter-cultural investment in a computer "revolution."

Clearly not all of these visions have been realized. Computer hobbyism turned into Apple Inc., Microsoft, and an array of makers of PCs. Today we do not have to assemble our PCs; we do not even have to install the system software. We barely even have to use our PCs. Their capacities for data management, data mining, consumer profiling and virus scaning far surpass our capacity for free choice in a free market.

Nevertheless, there are lessons to be learned from the example of the computer hobbyist subculture: Technology can be expressive, but it is not value-neutral. Technologies are more than a tool, they are a medium, and often what they mediate is as much ideological as informatic. Technological democratization or "globalization" is a problematic way of defining an interior that acts as a whole. Knowledges are always linked to disciplines, and disciplines are always linked to institutional sites, even though the way knowledges operate and the sites in which they are found may be unlikely, situational, provisional, unofficial. Finally, a last lesson from the computer hobbyist subculture. The computer revolution will not be televised, but streamed, and you must have a paid-for password (128-bit encrypted) and plug-in to enter the site.

Manifesto

4. Thesis on the Biotech Hobbyist

A biotech hobbyist, not unlike a computer hobbyist, critically "takes up" an array of specialized knowledges and practices, in ways that are unofficial, unsanctioned, unintended, playful, creative, personal, affective, surprising, and ethical.

Like the computer hobbyist, the biotech hobbyist is driven by an interest in possibilities that are social and political, as well as scientific and technical. Unlike the computer hobbyist, the biotech hobbyist does not believe in a utopia of "biotech for the people," nor that biotech is a technological response to problems that are social.

Like the computer hobbyist, the biotech hobbyist brings the non-specialist into specialist zones of activity, involving self-directed learning, hands-on experience, community-building, and shared knowledges. Unlike the computer hobbyist, the biotech hobbyist does not see this development of non-specialization as an opportunistic means for establishing a start-up company.

What does a biotech hobbyist do, and why? A hobbyist may been regarded as a non-specialist committed to and deeply interested in a specialized field. "Specialization" here may have disciplinary meanings (having a degree in biology), institutional meanings (being associated with a research foundation), academic meanings (going to conferences on a particular research topic), or pragmatic meanings (knowing how to perform certain laboratory techniques). Therefore, a non-specialist may know everything about proteomics, but can still be a non-specialist. "Specialization" is as much a disciplinary and institutional term as it is a scientific and technical one.

A non-specialist interested in a specialized field is therefore confronted with the question of access. How do I learn more about this? How can I not just read about, but actually do this? How can I learn by doing? How can I develop my own perspectives and responses? In short, how can I participate or engage in this field in a socially meaningful way, without ever assuming that I am a "specialist"?

Biotechnology is a highly specialized set of practices. A hobbyist approach to biotech puts forth a basic challenge: that a non-specialist (hobbyist) engagement with biotech can count as legitimate knowledge in the contestation over the meanings of biotechnology in our society.

***Eugene Thacker**/Biotech Hobbyist*

Notes Towards

One of the inspirations for Biotech Hobbyist comes from computer hobbyist activities in the U.S. during the 1970s. Remember, this is pre-Mac, pre-PC, and the Internet was still in its infancy, limited to select universities and government research. As the name implies, computer hobbyism was an early DIY approach to computing, and such "computerniks" share many principles with amateur radio movements. The main outlet for these communities were low-budget zines such as *Creative Computing* and *Computer Hobbyist*, groups such as the Homebrew Computer Club, and the main activities centered around assembling and tinkering with inexpensive kits. Some kits were sold as a self-contained package, including the famous Altair 8800 (first sold in *Popular Electronics* magazine in 1975 for $397, with 256 bytes of memory, no inputs, outputs, and no terminal).

A thorough history of the computer hobbyist "movement" has yet to be written, but it is mentioned in a number of histories of the computer, such as those by Paul Ceruzzi and Brian Winston. The closest example would be histories of the personal computer, such as Paul Freiberger and Michael Swaine's book *Fire in the Valley*. But what we really need is not just a history, but a sociology of computer hobbyism. It isn't hard to imagine the main demographic of computer hobbyists: middle-class, white, male engineering geeks, many grads of MIT or CalTech. Apple grew out of this hobbyist community, and the Apple I was sold as a kit (they preferred a wood cover with "Apple" carved in it instead of Titanium back then...). So computer hobbyism starts as engineers noodling and tinkering, and eventually morphs into the Mac, the PC, and Microsoft. Not exactly "computing for the people," but interesting nevertheless.

In fact, the computer hobbyist community shows some overlaps with the "computer lib" movement and early hacker groups, both concentrated mostly on the West coast. Ted Nelson's collage-like, wild designs in his book *Computer Lib* brought together both an interest in new technologies (the book itself was a kind of "hypertext," a term Nelson coined), as well as a kind of Left coast liberalism that later became associated with communities like the Well and the slogan "information wants to be free."

In his book *Hackers*, Steven Levy briefly describes the work of Lee Felsenstein and the beginnings of the Homebrew Computer Club. Levy's take is to construct a history of computer hacking, and he suggests that DIY is equivalent to hacking:

"The computer is a magic box. It's a tool. It's an art form. It's the ultimate martial art... There's no bullshit in there. Without truth, the computer won't work. You can't bullshit a computer, God damn it, the bit is there or the bit ain't there."

Les Solomon, former editor of *Popular Electronics* magazine

a Sociology of Computer Hobbyism

Eugene Thacker/*Biotech Hobbyist*

"Felsenstein was only one of hundreds of engineers in the Bay Area who somewhere along the line had shed all pretenses that their interest was solely professional. They loved the hands-on aspects of circuitry and electronics, and even if many of them worked by day in firms with exotic names like Zilog and Itel and National Semiconductor, they would come home at night and build, build fantastic projects on epoxy-based silk-screened boards loaded with etched lines and lumpy rows of ICs...It was hacking. If there was a goal at all, it was constructing a computer in one's very own home. Not to serve a specific function, but to play with, to explore. The ultimate system. But these hackers of hardware would not often confide their objective to outsiders, because, in 1974, the idea of a regular person having a computer in his home was patently absurd." (177)

Bob Lash, in his "Memoir of a Homebrew Computer Club Member," describes the technical fascination with the new technology, in what would become the Homebrew Computer Club in Palo Alto, CA, which met between 1974-86:

Notes Towards a Sociology of Computer Hobbyism

"The HP2000F had a bank of 110 BAUD modems (10 characters per second) for teletype dial-ups, as well as "Fast Ports" at 300 BAUD. One afternoon I was at the People's Computer Center and used one of their ASR 33 teletypes with an acoustical coupler (you actually put the telephone handset into a cradle on the slanted face of the teletype), dialed in, and had an online "chat" with fellow HP2000F system operator Greg Dolkas (who later went on to work at HP, I understand). The People's Computer Center was a really cool place for computer hobbiests at that time -- as you see in the printout, they had a PDP-8/E, PDP-8/L, and 4 ASR 33 teletypes... the best public access for "real people"!

Lash goes on to give a sense of the early Homebrew meetings:

"Posted on a bulletin board in the PALY terminal room (containing 6 noisy KSR-33 teletypes in the Palo Alto High School math-science office) was a notice that a computer group meeting would be held at the home of Gordon French. I jotted down the notice's announced date and location onto the back of an envelope (how appropriate!). Mike Fremont (another student System Operator) and I both attended this first meeting in Gordon French's garage. Ralph Campbell (also from PALY) attended as well. We were the three youngest ones there, and rather shy and quiet (as I still am today), but were absolutely astounded when Gordon let us step into his house and see his 8008 system using 16K of SHIFT REGISTERS (!) as memory and his very own teletype (unlike random-access memory, his system actually shifted the data in a ring and WAITED until the desired address came up for data read/writes! We were extremely impressed.)."

But it would be too easy to characterize computer hobbyism simply as a preoccupation with the gory technical details. Inasmuch as we can say there was a politics to this loose grouping of activities, it was an ambiguous politics. It was a politics caught somewhere between leftist social consciousness and the desire, on the part of hobbyists, for a consumer-based or user-powered market. Fred Moore, an early member of the Homebrew Computing Club, wrote about the political aspects of hobbyism in the following way:

"The evidence is overwhelming that people want computers, probably for self-entertainment and education usage. Why did the Big Companies miss this market? They were busy selling overpriced machines to each other (and the government and the military). They don't want to sell directly to the public. I'm all in favor of the splash MITS is having with the Altair because it will do three things: (1) force the awakening of other companies to the demand for low-cost computers in the home . . . (2) cause local computer clubs and hobby groups to form to fill the technical knowledge vacuum, (3) help demystify computers..." (quoted in Levy, 208).

Consistant with the so-called Hacker Ethic of the period, computer hobbyism and computer hacking were seen by many to intersect on a number of levels. For one, they had a common enemy, which was the electronics and computer industry. But, with the benefit of hindsight, we should also note that Moore's rhetoric speaks more about neo-liberalism than liberalism itself. The ambiguous politics of hacking can be witnessed in Moore's words, and in the hobbyist publications as well – at once a direct challenge to the Powers That Be, as well as a voiced demand for consumer rights.

A closer examination of the connections between technology and politics in relation to hobbyism is given by Michael Hauben in his paper, "Participatory Democracy from the 1960s and SDS into the Future On-line":

"The 1960s was a time of people around the world struggling for more of a say in the decisions of their society. The emergence of the personal computer in the late 70s and early 80s and the longer gestation of the new forms of people-controlled communication facilitated by the Internet and Usenet in the late 80s and today are the direct descendents of 1960s."

Hauben explicitly connects the hobbyist movement with the political experiments of the SDS. The point at which they meet is the need for accessible means of communication in a participatory democracy. It is debatable whether or not these linkages actually existed, but the context of 1960s radicalism can be seen to supply the backdrop for the various hobbyist zines and activities.

Thus there is ample evidence of a social consciousness in the computer hobbyist movements. David Ahl, who had previously edited a magazine dedicated to educational computing called *EDU*, describes the impetus for starting the zine *Creative Computing*:

"Over the years EDU flourished and grew into a 48-plus page magazine. However, there were certain aspects of educational computing which *EDU* could not satisfactorily address. In particular, school users, both college and elementary/secondary, need far more classroom activities, exercises, problems, and ideas than are available in textbooks and other magazines. Also, there ought to be a discussion of the social aspect of the computer, its effect on jobs, medical care, privacy, and the like. Furthermore, what about the user of non-DEC computers? Clearly to be responsive to these needs another vehicle was needed. Thus, *Creative Computing* was born, at least as an idea" (vol. 1, 1976, p.2).

Unlike other hobbyist zines, *Creative Computing* attempted to address key issues surrounding the burgeoning personal computer field, including application, human-computer interaction, reviews and tests of promising computer systems, and issues pertaining to education, law, industry, computer art, and the social impact of computers (the "computer threat to society"). *Creative Computing* was also unique in that it not only discussed the building of computers, but would also supply in its pages problems, puzzles, computer games, and other activities related to computers. If we had to characterize *Creative Computing* among other zines, we might say it took a pedagogical approach to hobbyism.

The idea that new technologies need to be made accessible to the public, and that innovative uses can come out of such access, is not new. You can argue that this is a basic premise of the historical avant-gardes (Russian Constructivism's artist-engineer and Kino-eye, Surrealism's flea-market "marvelous," Dada's multimedia performances, and later, groups like EAT, GRAV, Group Zero, and so on). But the avant-garde is a dead end, and computer hobbyist zines show little interest in art.

A proposition: historically, computer hobbyism mutates into the personal computer and software industry, and it does this by mutating liberalism into neo-liberalism, by turning the knowledge of design (or building, or programming) into the knowledge of use (user-friendliness).

Another proposition: from a sociological point of view, the current trends surrounding open source, file sharing, and even software art, are all replays of the computer hobbyist moment.

What does this have to do with biotech?

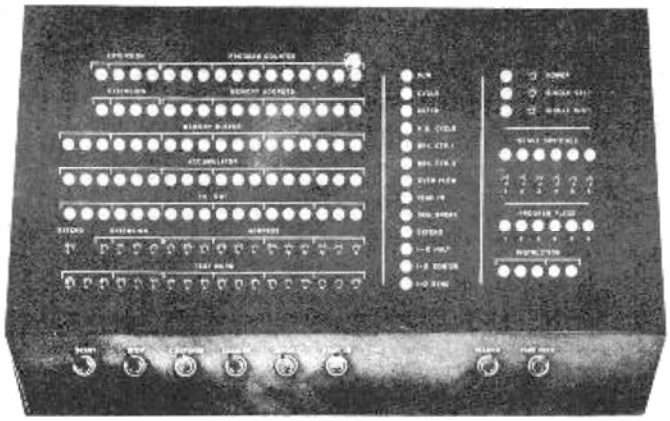
A third proposition: the Human Genome Project is the equivalent of the ENIAC. World War II, room-sized, mainframe computers, eventually gave way to IBM's business machines and miniaturization, which gave way to computer hobbyism and the idea of computer kits. Might biotech be on a similar trajectory? Already, sophisticated equipment like PCR thermal cyclers, DNA microarrays, and gene sequencing computers are standard fixtures in many biology labs. And older technologies, like gel electrophoresis, are routinely taught in high school biology courses. And we've seen a plethora of artists using glowing green genes for just about everything.

But the idea that biotechnology might become "personal biotechnology" is kind of ridiculous, if not science fictional. What would the average person want with PCR? So you learn how to sequence your own DNA – and then what?

What is interesting about the example of computer hobbyism is that the same question crops up again and again in relation to computers. When computer kits are marketed by Radio Shack or Tandy, all the emphasis is on the DIY approach. Nothing is ever said about application or usefulness. Furthermore, computer hobbyist zines give the constant impression of being not quite sure what to do with all these machines. There's almost an awkward quality to the writing, as if the increasing obsession over every detail of a computer will somehow compensate for the fact that no one has the slightest idea what to do once the computer is assembled and blinking away. Which indicates that computer hobbyism was largely about computing for computing's sake.

Except if we maintain – and this open to interpretation – that computer hobbyism was inseparable from the computer lib movement. If this is the case, then what is at stake is not just gadgets, switches, and blinking lights, but rather the social uses of specialized, technical knowledges. Anyone who has tried to learn a programming language or any other technical skill can testify to the frustrations, time-expenditure, and immersion of the process. Specialization is synonymous with technological development (presumably this is why we have college degrees...). But specialization is also a straightjacket. Cross-disciplinary art projects like EAT or Radical Software are the exceptions, not the rule.

Notes Towards a Sociology of Computer Hobbyism



So there are a few things to learn from computer hobbyism. First, it exists in this tension-filled space between technology and politics, between a geek culture of gadgets and socially-conscious communities. Second, the uses and application of hobbyism (let alone the idea of "personal computing") come after the fact. There is no practical application that then serves as the initiative for tinkering and noodling. It's the reverse.

But the context of biotech is radically different from that of computing. First, there is massive social and cultural anxiety surrounding biotech, not least because of the insecurities felt in the wake of terrorism and emerging infectious diseases. Around the time AIDS begins to make headlines in the U.S., you can catalog all the science fiction films which deal in some way with epidemics and body anxiety, from The Andromeda Strain (ok, that's pre-AIDS), to The Thing, to to Outbreak, to Lifeforce, to Virus, to 28 Days Later. And the genre of the "medical thriller" also emerges during this time (Robin Cook, Richard Preston). And then there are pop science books and PBS specials on "virus hunters" and heroic narratives involving the CDC. And so on. Biotech is on our minds. And it worries us.

All of which is to note that most people will feel frustrated about a computer virus, and mortally threatened about Mad Cow or SARS. Computers are not reality; but biotech is "life itself." Even though it's not hard to show how computers have very real, very material effects, and how bioterrorism is really about boosting national security policies, it is hard to deny the difference in context between computer hobbyism and a biotech hobbyism.

So what are we left with? We're left with a few lessons from the computer hobbyist movement as a technological analogy, but we're also left with a whole new set of problems in the era of genetic manipulation, bio-patenting, GM foods, bioinformatics, and so forth. Not only are there issues of hardware, software, and now, wetware, but there are issues pertaining to the privatization of biological data, as well as issues pertaining to the access to bio-knowledges.

This last aspect is especially crucial, because it reaches outside of both industry and research, and into the non-specialist, public understanding of biotechnology. In fact, we can outline a few pressing concerns for any "biotech hobbyist":

- Gaining a knowledge of techniques and technologies will need to happen right alongside gaining a knowledge of ethics and politics. This dual mindset will ensure that biotech hobbyism doesn't simply become technology-for-technology's sake.

- Ethics will have to be a foundational element of biotech hobbyism. This will mean several things: laboratory safety, continuous dialogue with scientists, a concern for the social aspects of biotech, and an interest in a critical approach to the instrumentalization of "life" that biotech represents.

- Interdisciplinarity between science, art, and politics, will have to be taken seriously. This should mean more than artists hiring out lab technicians to make DNA prints for the art gallery. Rather, it should mean an ongoing dialogue and debate between disciplines.

- Education, taken in the broadest sense, will provide a key context for biotech hobbyism. This is not limited to the classroom, but can take place in a range of contexts, including activism, performance, art installation, public art, and so forth.

- Paying attention to context will have to frame biotech hobbyism. What problems and benefits arise in the art context? What problems and benefits arise in corporate or institutional outreach programs? What problems and benefits arise in the contexts of academia, activism, or technology conventions?

- Finally, a biotech hobbyism can be fun, without being irresponsible. Put another way, a biotech hobbyism can be as playful as it can be serious, and both modes can be critical and political.

References

Ahl, David. "The Computer Threat to Society." *Creative Computing*, Vol. I (1976): 66-67.

Ceruzzi, Paul. *A History of Modern Computing*. Cambridge: MIT, 1998.

Creative Computing archives:
http://www.atarimuseum.org/bcc1/

Freiberger, Paul, and Michael Swaine. *Fire in the Valley: The Making of the Personal Computer*. New York: McGraw-Hill, 1999.

Hauben, Michael. "Participatory Democracy from the 1960s and SDS into the Future On-line."
Personal homepage: http://opencollector.org/history/homebrew/netdemocracy-60s.html

Lash, Bob. "Memoir of a Homebrew Computer Club Member." Personal homepage:
http://www.bambi.net/bob/homebrew.html

Levy, Steven. Hackers: *Heroes of the Computer Revolution*. New York: Penguin, 2001.

Nelson, Ted. *Computer Lib/Dream Machines*. Sausalito: Mindful Press. 1974.