IPRO 318

Searching for Novel Drug Targets

Final Report

Instructor: Yuzhu Zhang

Team: Kathan Amin
Evan Himchak
Xiao-Xuan Hu
Suyao Huang
Kyle Laster
Hyun Lee
Alfred Liu
Joshua Nedrud*
Megha Pansara
Snehalata Topagi
Amanda Wicker
Hannah Zwibelman
*Team leader

November 29, 2007
Fall 2007

Introduction Over the course of the term, IPRO 318 made progress in the "[Search] for Novel Drug Targets." This IPRO is part of an ongoing effort to try to find potential drug targets on pathogenic proteins. During past terms, IPRO 318 has compiled information on proteins that are known to be involved in diseases in a disease gene database. By comparing this to a database of hypothetical genes, possible high value proteins were found. This term IPRO 318 continued the lab work on the high value proteins found in previous semesters, found new ones and began lab work on those proteins, continued work on the databases, and started a new allergen database.

Background The majority of the gene functions in the human genome (see ref.1 for the first draft) are still unknown. Many of those genes may be a valuable drug target for numerous human health issues. The discovery of novel molecular drug targets is becoming more efficient with the use of powerful high throughput technologies, gene expression microarrays, and proteomics methods (2,3). However, presently these targets comprise of only a small fraction of the human genome and determining the potential pharmaceutical applications of genes and their encoded proteins is still the rate-limiting step in deriving value from this gene sequence information. Thus, it is important to identify new drug targets in the post genomic era.

The final goal of this IPRO is to find novel drug targets. In this process the first objective that must be met is the creation of three databases. The first database, which contained hypothetical gene products predicted from the human genome, was almost completed during the Spring of 2007. The second database contains proteins that are known to be involved in diseases. The final database contains potential drug targets based on sequence alignment between the first two databases. Our team also started work on a forth database, an allergen database. Our next objective this term was to continue research on one of the proteins found last term and to find its interaction with CryAB in yeast. Our final objective was to find more high value proteins and begin work on determining their roles in the cells. Hopefully future IPRO 318s will continue to work on the databases, do research on the high value proteins found, and find more high value proteins.

Purpose The purpose of this IPRO project is to identify pathogen proteins from human genomic sequences with unknown functions that may be targeted for future drug development. With the knowledge that certain genes induce the production of proteins related to specific diseases, the genes themselves become potential targets for drugs that treat the disease. With genomic databases and other bioinformatic tools, potential drug targets can be based off of sequence alignment of the human genome and diseases protein databanks. Any potential drug target gene then undergoes further experimentation to test the validity of the target. With such a vast array of protein sequences within the human genome as well as the plethora of disease related proteins, this IPRO is a continuous project which can take many years before a valid drug target can be found. However, as of the Fall 2007 semester, the IPRO team has set out to expand the protein databases, update the website, experimentation with new potential genes, as well as expansion of research on the potential gene discovered to have interactions with a specific protein related to a disease.

Research Methodology (accompanying figure: 1)

- **Computer Analysis and Web Design Team**

  Materials:
  - UNIX based POSIX compliant operating system
  - Perl interpreter
  - Blast (Basic Local Alignment and Search Tool)
  - NCBI GenBank (Public genome database)

The computer analysis and web design team's main objectives for the semester included blasting the current databases produced by previous semesters IPROs. The results from this blast were then screened for well qualified genes. Well qualified in this case indicated that they were globular, and had a good mix of hydrophobic and hydrophilic regions. This was important as globular proteins are more likely to carry drug targets in them. Other than the website, the other

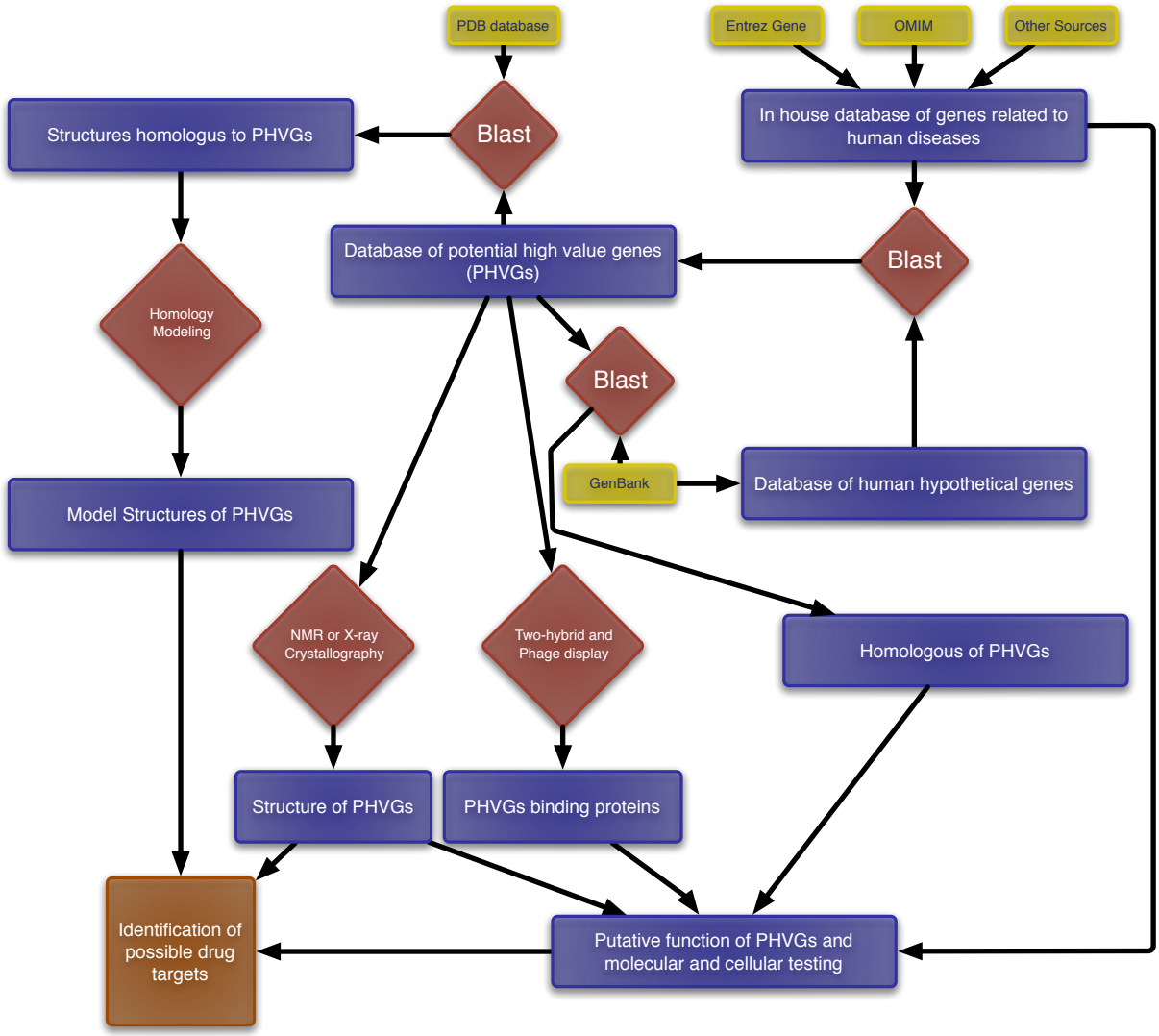


Figure 1:
Diagram of the research organization of our project for the current semester

main task of this group was to build up a new allergen database for future IPROs. This process involved querying established allergen databases and then linking these results with the public genome databases to create a custom in-house database which will be blasted with the hypothetical gene database in semesters to come.

- **New Gene Group**

  Materials:
  - Genes of interest
  - Growth mediums
  - Laboratory glassware
  - Centrifuge
  - RNA primer sequence
  - DNA polymerase
  - Forward and reverse DNA primer sequences
  - Gel electrophoresis kit
  - DNA marker ladder
  - Micro pipettes
  - Two hybrid system kit

Using an existing database compiled from the previous Spring IPRO318, hypothetical gene sequences of various illnesses were analyzed. From a list of thousands, several sequences were subjected to the BLAST algorithm to determine their homology with known illnesses. From our list of candidates, 2 sequences were chosen due to their high homology in addition to their globular

characteristics. Based on the findings, the team ordered the sequences from a commercial source in order to find potential novel drug targets through experimentation. Molecular biological principles, namely Polymerase Chain Reaction (PCR), were exploited in order to amplify the sequence. Once amplified, the sequence was converted to a plasmid for insertion into a yeast cell, which was induced to produce the protein coded by the insert. This protein can then be purified and concentrated. This can be used to study the structure-function relationship. We can also use protein expression to study the protein-protein interaction, as shown in Figure 2. The team members will extract from vectors using affinity columns and various electrophoresis separation techniques. Finally, the extracted gene will be spliced into custom plasmids, and it will be run by two-hybrid system.

- **Continuation Group**

  Background to Methods:
  - Last semester H1 (a hypothetical human gene related to a disease causing gene) was used as bait and found to interact with CRY-AB (α,β subunit of a major eye protein, acquired from a gene bank, also found in other tissues such as the heart)
  - CRY-AB mutations that cause disease:
    - P205 – eye
    - D140N – heart

  Materials:
  - Genes of interest
  - Growth mediums
    - Plates used in H1 interaction with CRY-AB

  1. Gal/UHWL- If colonies survive on this plate, the two proteins interacted (necessary in order for the gene needed to synthesize L to be transcribed)
  2. Gal/UHW-/X-Gal- If colonies that turn blue on this plate interaction has occurred (necessary in order to transcribe β-galactosidase)
  3. Glu/UHW-/X-Gal- Colonies should grow but remain white on this plate; it is used as a test for false positives on the previous plate
  4. Glu/UHW- All colonies should grow; this plate is to keep the yeast around for future use

  - Laboratory glassware

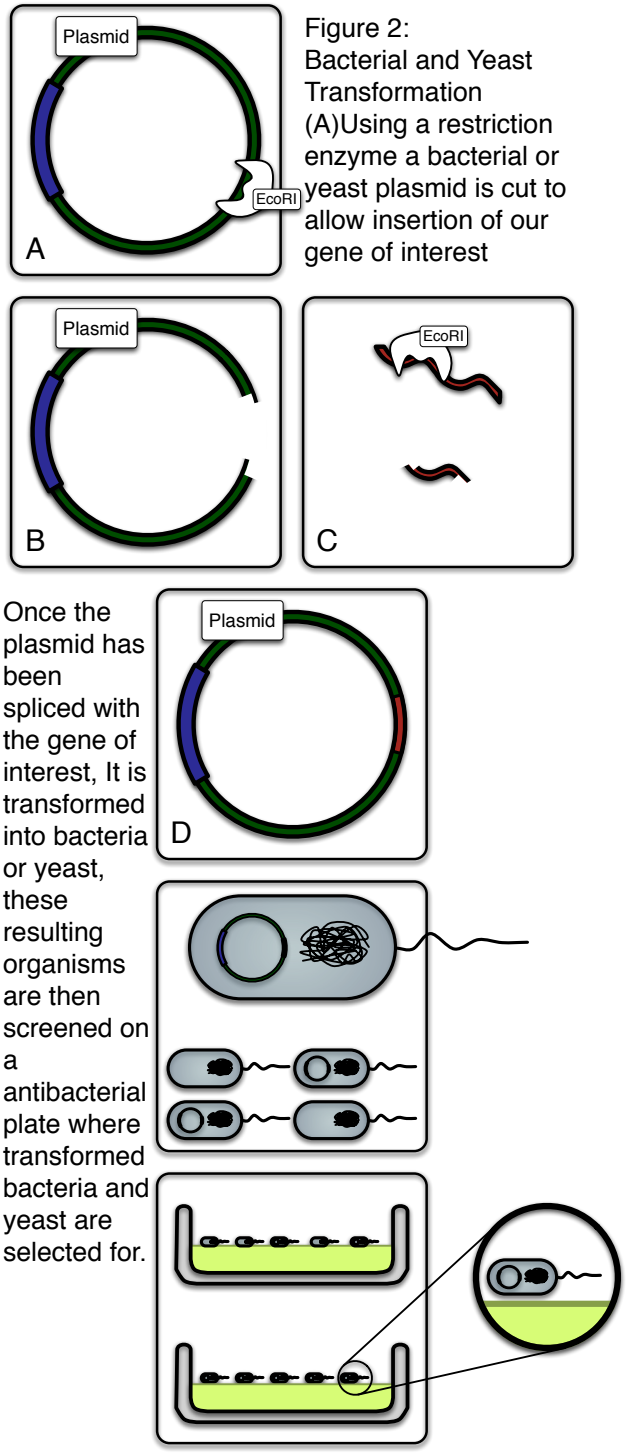


Figure 2:
Bacterial and Yeast Transformation
(A)Using a restriction enzyme a bacterial or yeast plasmid is cut to allow insertion of our gene of interest

Once the plasmid has been spliced with the gene of interest, It is transformed into bacteria or yeast, these resulting organisms are then screened on a antibacterial plate where transformed bacteria and yeast are selected for.

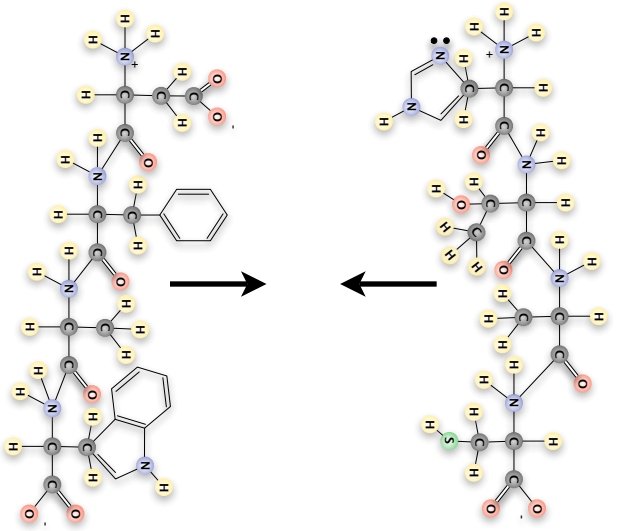
Transcription factor preparations:
- Has two domains normally encoded by one gene, therefore one protein
- We have them split into two genes on different vectors

    - The DNA binding domain will be transcribed, and therefore physically linked on the protein level, to H1
    - The Activator is linked to CRY-AB (our unknown of choice)
- The only way to physically bring these two together, and therefore transcribe the genes under their control on the reporter plasmid, is if H1 and CRY-AB interact
- Genes under control of this transcription factor:
    - β-galactosidase
    - The gene that makes the production of Leucine (L) possible

Procedural List:

- Cloned CRY-AB via PCR
- Put CRY-AB into T2 vector
- Put gene into 202 or 4-5 vector
- Created CRY-AB mutations via PCR
    - One mutation has already been sequenced and proven successful
    - The second is most likely done as well, but has yet to be sequenced
- Put the 1st mutation into plasmid and 202 vector
- Transform this mutation and the H1 vector into yeast

The goal of this IPRO is to study the interaction of the gene crystalline AB and hypothetical gene H1. Cry-AB mutations have been indicated in several key human diseases such as heart disease and blindness. To test this hypothesis, our group has created two mutations of Cry-AB to observe whether these mutations effectively prohibit Cry-AB interaction with the H1 gene in a yeast two-hybrid system using a reporter plasmid. The wild type (normal gene) Cry-AB as well as both mutant strains were put into the plamid vectors, 4-5. In order to test the Cry-AB/H1 interaction the H1 gene was put into the 202 vector. These plasmids were then transformed into yeast to test the observe the interaction, or possible lack there of, between the wild type, mutant strains, and H1. If the transcribed and translated proteins of Cry-AB and H1 are interacting, they will form a DNA binding-activator complex that will in turn transcribe and translate an indicative (usually colored) protein. Figure 3. By testing the wild type of Cry-AB we continue to confirm the interaction between the non-mutant strain of Cry-AB and hypothetical gene H1. We will then be able to test the interaction between the mutant strains of Cry-



Protein A (depicting* CryAB) Protein B (depicting* H1) Binding Activator Complex

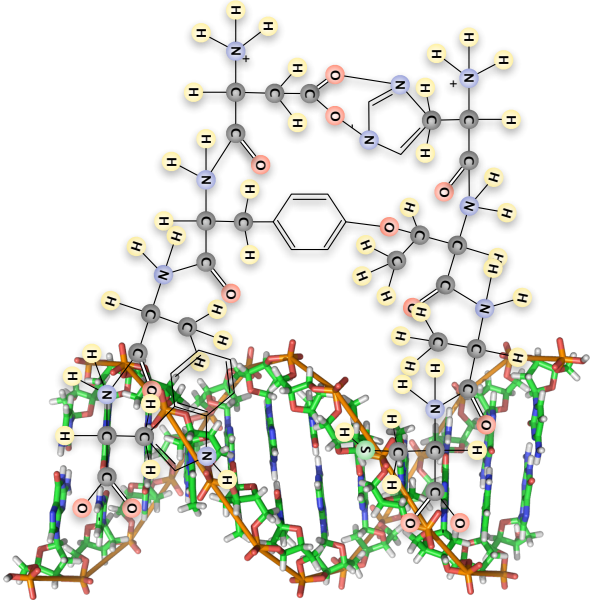


Figure 3:
Two Hybrid Assay
(A)The two proteins are expressed in the cell
(B)If they interact this will produce a third protein encoded at the binding site

AB 590 and 600 and H1. The empty 202 vector, meaning H1 was not transformed into it, is used as a control to assure that it is indeed H1 Cry-AB is interacting with and not part of the carrier vector. Figure 4.

*depiction is of a example protein and is not intended to resemble either H1 or CryAB, but is for demonstration value only

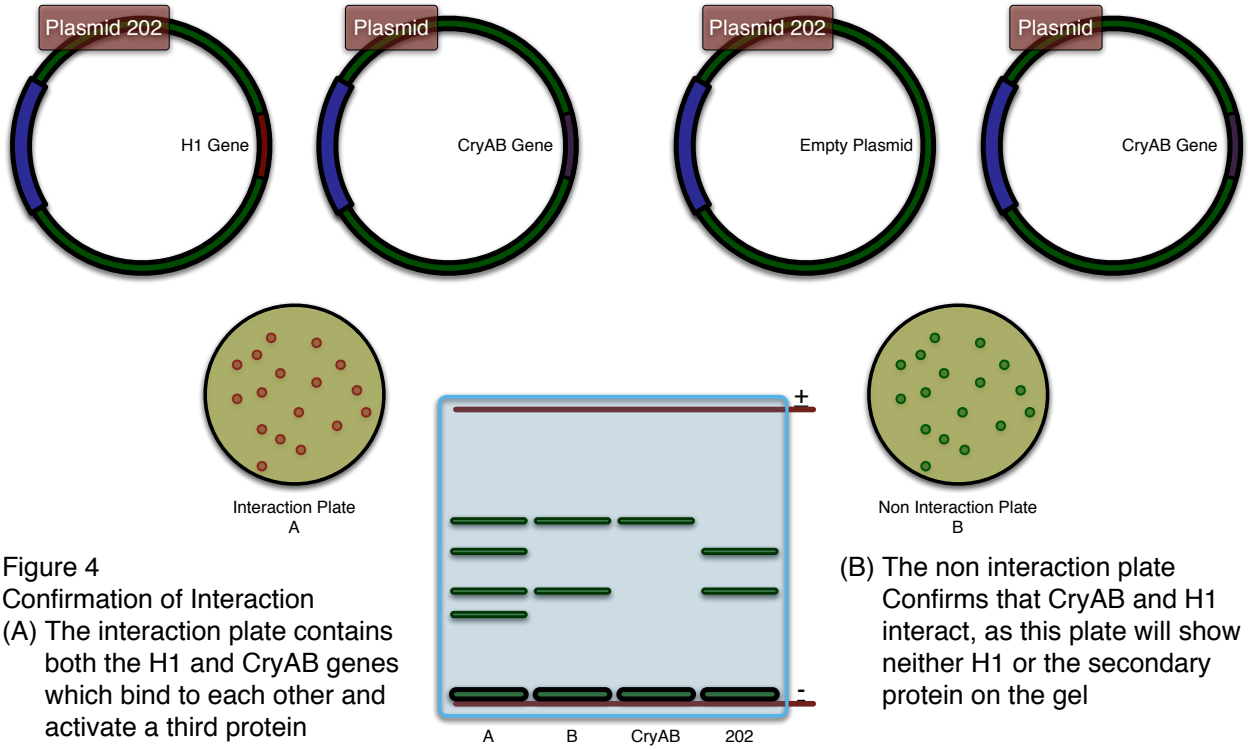


Figure 4
Confirmation of Interaction
(A) The interaction plate contains both the H1 and CryAB genes which bind to each other and activate a third protein

(B) The non interaction plate Confirms that CryAB and H1 interact, as this plate will show neither H1 or the secondary protein on the gel

Assignments

Team Assignments

Team	Members
Computer Analysis and Web Design Team	Joshua Nedrud Suyao Huang
New Gene Group	Kathan Amin Xiao-Xuan Hu Kyle Laster Hyun Lee Snehalata Topgi
Continuation Group	Evan Himchak Alfred Liu Megha Pansara Amanda Wicker Hannah Zwibelman

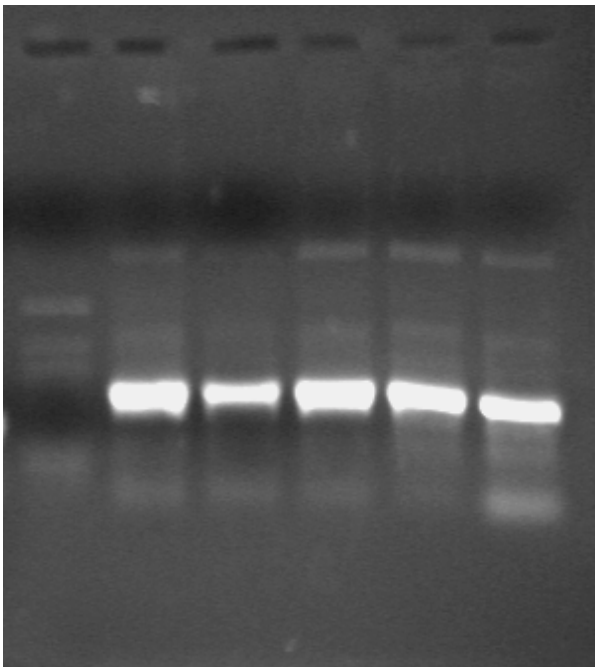
Individual Assignments

Name	Initiatives
Amin, Kathan	Isolation of gene
Himchak, Evan	Protein interaction, cloning vector
Hu, Xiao-Xuan	PCR based amplification of gene
Huang, Suyao	Web design and implementation, creation of computer models
Laster, Kyle	placement of gene into vector
Lee, Hyun	two hybrid system
Liu, Alfred	Experimental team 2 leader, protein interaction
Nedrud, Joshua	Team leader, database programmer, creation of computer models
Pansara, Megha	Protein interaction, transforming yeast
Topgi, Snehalata	extract gene
Wicker, Amanda	Protein interaction, PCR mutagenesis
Zwibelman, Hannah	Protein interaction, two hybrid system

Obstacles

- **Computer Analysis and Web Design Team**

The computer group faced the first obstacle of correctly using the right parameters to blast the protein databases. We also faced the obstacle of finding the right website host to be compatible with our website. There were some obstacles with finding the right allergen database for us to use.



stuff here

- **New Gene Group**

Since this project is continued from the last semester, the new gene group basically recapitulates the procedures the previous semester's IPRO 318 team took in order to familiarize our team with the whole design of the project. In our group, we first worked on selecting proteins to start with. We have done extensive researches to find out the most appropriate proteins which are involved in various diseases. We had some difficulties selecting the proteins since it was hard to search for globular proteins from the large pool of proteins. This issue was helped by the computer team by narrowing our selections from huge numbers of genes available from the databases. Also, getting the proteins took longer time than expected. Furthermore, after getting the proteins, our members needed to be at the lab almost every day to continue with research procedures. We had overcome this by going into the lab two members at a time so that we could divide workload and also help one another in the lab with the supervision of Dr. Zhang and T.As in the lab, who were very helpful.

- **Continuation Group**

Through the semester lab work was done in order to further research with the old genes. There were not too many obstacles. The one major obstacle that was encountered was contamination. Some portions of our experiment had to be repeated due to the contamination. The source of contamination was unknown. Overall, the experiments ran smoothly with few obstacles.

Results

- **Computer Analysis and Web Design Team**

A website has been finished. The website currently displays most of our research information, our goal and our progress. The website also contains a gallery of homolog protein 3D structures of our hypothetical genes. Aside from that, the website has a guest book service as well as a shout box for people to communicate short bursts of information in real time. There are molecular biology protocols on the website that we currently use or will use in the future. The allergen database has been mostly done. The protein blast database has been done too and have been used for the search of high potential value genes.

- **New Gene Group**

The selection of viable proteins was completed early in the course. There were several proteins that did not match our criteria of being viable (globular, small, similar to human protein, etc). We narrowed down the selection to three genes that would be viable targets for research. The new gene group successfully made it to the lab and began to amplify the gene through PCR. Two of the three genes selected have been replicated. The third gene had trouble showing up on the gels and therefore was unable to be advanced further.

Several PCR runs have been completed on two of the genes. They have been placed through gel electrophoresis and one of them has already been inserted into a plasmid. That plasmid was then transfected into bacteria and the colonies were isolated. These colonies were also grown in nutrient media and checked for the presence of the plasmid. The second gene has been placed into a plasmid and is currently in the process of being transfected into bacteria. The transfected bacteria will then be used to run several tests and to find out if the protein can be replicated and mutated. The step after would be trying to create a two-hybrid system and transfecting yeast, a eukaryotic specimen. Team members were all

exposed to techniques in labs to amplify and create plasmids. They were also exposed to methods of transfection and gel electrophoresis.

• **Continuation Group**
At the end of the previous team's term, there was interaction discovered between the hypothetical H1 gene product and the proteins related to heart disease and cataracts known as Crystallin AB or Cry-AB. Through literature, two separate mutations of the Cry-AB gene were known to exist. The mutations were then reproduced in the lab in order to test if Cry-AB mutations would disrupt its interaction with the H1 gene. Through PCR recombination, the mutation P205 was induced and placed into T2 bacterial vectors. Once replicated, the bacterial vectors were transformed into yeast cells in which interaction could be tested. Therefore, with the mutations induced and vectors for yeast prepared the next logical step would be to proceed with experimentation to test for interactions.

Recommendations Throughout the semester, the team faced numerous problems. We would like the next team to be aware that many experiments in the lab may not work out as expected. Procedures may have to be repeated several times in order to get significant results. In the future, we would like to see the group further experiment with the new genes and create eukaryotic vectors. Also, the beginnings of an allergen database were initiated this semester; we would like to see further advancement with this database.

References

1. genome Isaaoth, Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. Initial sequencing and analysis of

the human genome. Nature 2001;409(6822): 860-921.
2. Workman P. Drug discovery strategies: technologies to accelerate translation from target to drug. J Chemother 2004;16 Suppl 4:13-15.
3. Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 2003;13(10): 2291-2305.

Acknowledgments

We would like to thank Dr. Zhang, Tengchuan Jin, and the rest of the group members in the lab for helping us with all of our procedures.