Protein Interactions Extrapolated from Feline Protein Complexes

S. Martin*, Z. Mao†, L. S. Chan†, S. Rasheed†

10/21/2006

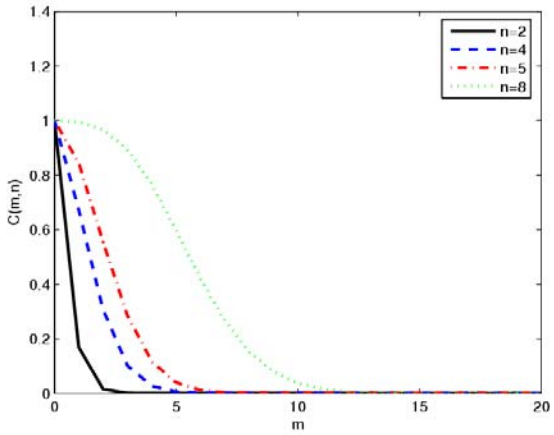
*Sandia National Laboratories
Albuquerque, NM

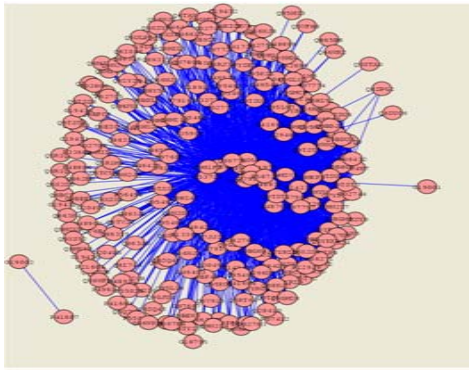
†University of Southern California
Los Angeles, CA

Sandia National Laboratories

Outline of Talk

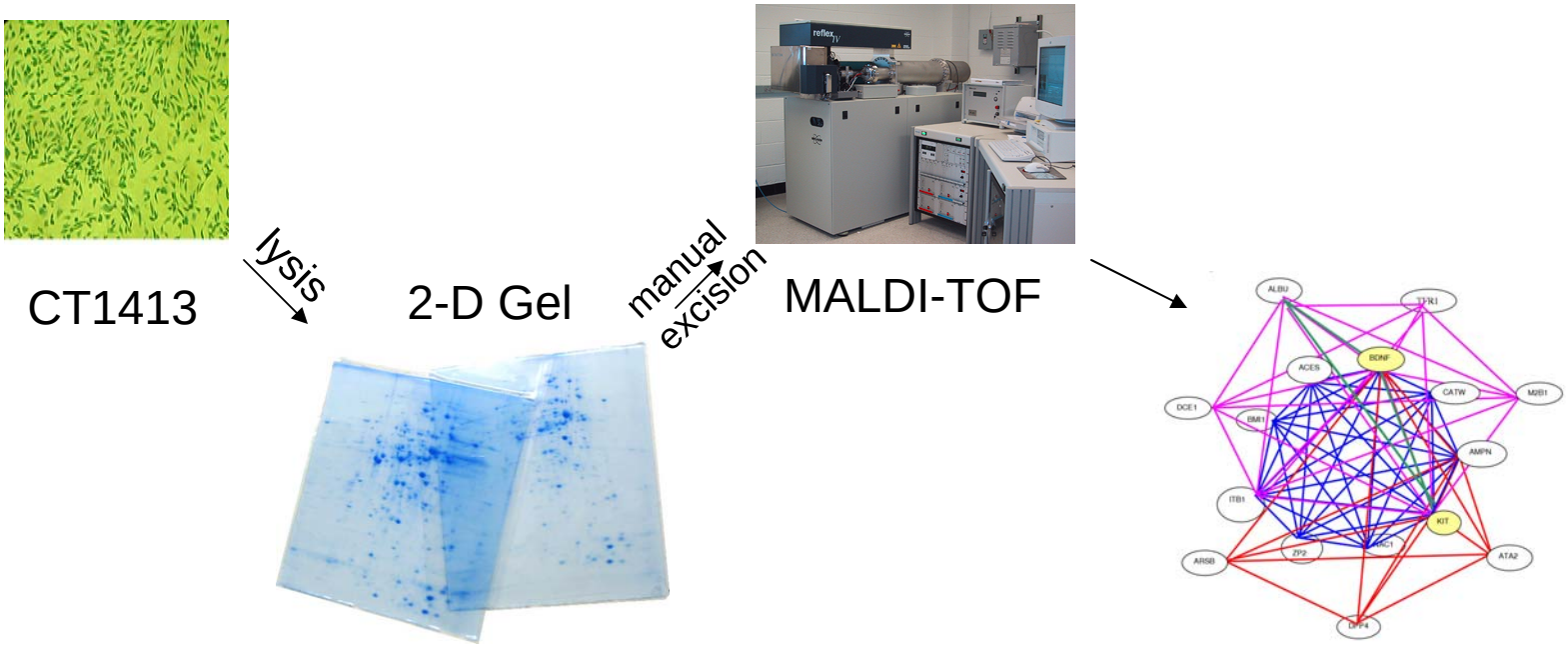
- Feline Protein Complexes
  - Experimental Data
  - Inferring Protein Interactions
- Extrapolating Protein Interactions
  - Relating Protein Interactions to Amino Acid Sequence
  - Application to Experimental Data
- Conclusion





Protein Complexes

- Feline protein complexes were identified using cat melanoma cell line CT1413, 2-D gel electrophoresis, and MALDI-TOF mass spectrometry.



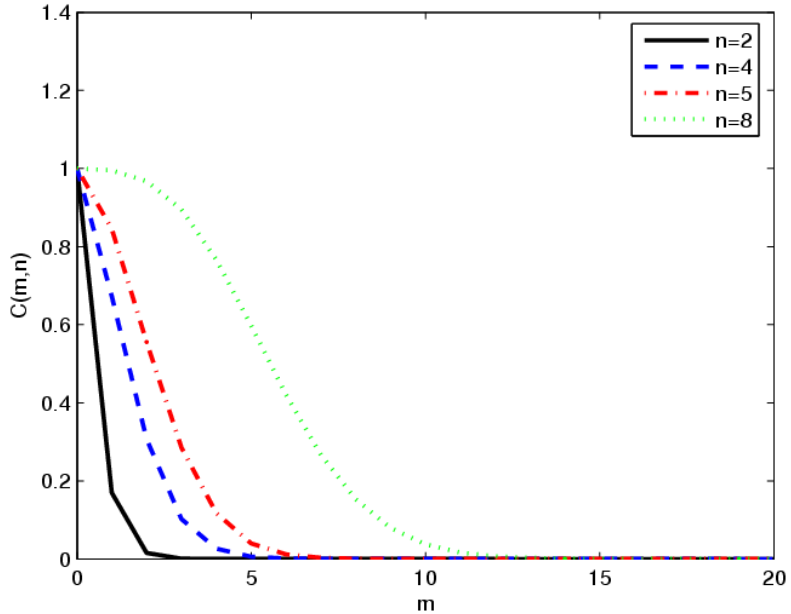
- The end result: groups of proteins assumed to be related by function/interaction potential.

Inferring Interactions from Complexes

- Initial assumption: if two proteins occur in the same complex they may interact.
- Natural extension: if a protein pair occurs in multiple complexes it is more likely to interact.
  - We computed the probability C(m,n) that two proteins occur in *m* or more complexes of size *n*.

$$C(m,n) = \sum_{i \geq m} \binom{c}{i} P_n^i (1 - P_n)^{c-i}$$

$$P_n = \frac{n(n-1)}{a(a-1)}$$



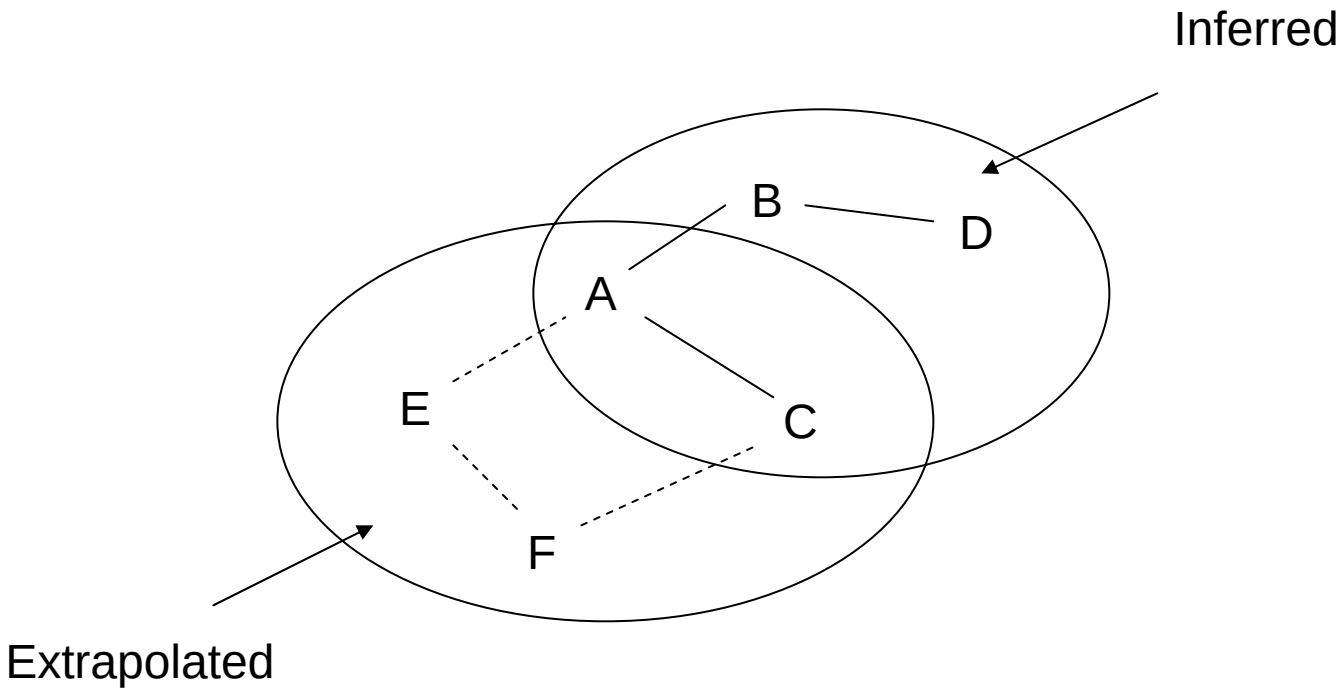
  - We compared the actual counts to C(m,n) to find protein pairs that were likely to interact.

Interactions Inferred from Complexes

Pair	Prot. 1 Name	Abbr.	Acc. #	Prot. 2 Name	Abbr.	Acc. #	p-value
1	Integrin beta-1 precursor	ITGB1	P53713	Mast/stem cell growth factor recept	KIT	Q28889	<.001
2	Integrin beta-1 precursor	ITGB1	P53713	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	<.001
3	Integrin beta-1 precursor	ITGB1	P53713	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
4	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	Mast/stem cell growth factor recept	KIT	Q28889	<.001
5	Zona pellucida sperm-binding protei	ZPB	P48834	Integrin beta-1 precursor	ITGB1	P53713	<.001
6	Serum albumin precursor	ALB	P49064	Integrin beta-1 precursor	ITGB1	P53713	<.001
7	Proto-oncogene tyrosine-protein kin	FES	P14238	Integrin beta-1 precursor	ITGB1	P53713	<.001
8	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
9	Serum albumin precursor	ALB	P49064	Mast/stem cell growth factor recept	KIT	Q28889	<.001
10	Mast/stem cell growth factor recept	KIT	Q28889	Transferrin receptor protein 1	TFRC	Q9MYZ3	<.001
11	Pyruvate kinase, M1 isozyme	PKM2	P11979	Integrin beta-1 precursor	ITGB1	P53713	<.001
12	Integrin beta-1 precursor	ITGB1	P53713	Aminopeptidase N	APN	P79171	<.001
13	Serum albumin precursor	ALB	P49064	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	<.001
14	Proto-oncogene tyrosine-protein kin	FES	P14238	Mast/stem cell growth factor recept	KIT	Q28889	<.001
15	Sodium/calcium exchanger 1 precurso	NCX1	P48767	Integrin beta-1 precursor	ITGB1	P53713	<.001
16	Pyruvate kinase, M1 isozyme	PKM2	P11979	Mast/stem cell growth factor recept	KIT	Q28889	0.002
17	Sodium/calcium exchanger 1 precurso	NCX1	P48767	Mast/stem cell growth factor recept	KIT	Q28889	0.002
18	Proto-oncogene tyrosine-protein kin	FES	P14238	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	0.007
19	Zona pellucida sperm-binding protei	ZPB	P48834	Sarcoplasmic/endoplasmic reticulum	SERCA2	Q00779	0.007
20	Zona pellucida sperm-binding protei	ZPB	P48834	Mast/stem cell growth factor recept	KIT	Q28889	0.007
21	Integrin beta-1 precursor	ITGB1	P53713	Alkaline phosphatase, tissue-nonspe	ALPL	Q29486	0.007
22	Zona pellucida sperm-binding protei	ZPB	P48834	Serum albumin precursor	ALB	P49064	0.016
23	Beta-glucuronidase precursor	GUSB	O97524	Integrin beta-1 precursor	ITGB1	P53713	0.016
24	Glutamate decarboxylase, 67 kDa iso	GAD67	P14748	Integrin beta-1 precursor	ITGB1	P53713	0.016
25	Aminopeptidase N	APN	P79171	Mast/stem cell growth factor recept	KIT	Q28889	0.016
26	Zona pellucida sperm-binding protei	ZPB	P48834	Transferrin receptor protein 1	TFRC	Q9MYZ3	0.016
27	Mast/stem cell growth factor recept	KIT	Q28889	Interleukin-1 beta convertase precu	CASP1	Q9MZV6	0.016
28	Integrin beta-1 precursor	ITGB1	P53713	Cathepsin W precursor	CTSW	Q9TST1	0.016
29	Lysosomal alpha-mannosidase precurs	MANB	O46432	Integrin beta-1 precursor	ITGB1	P53713	0.038
30	Integrin beta-1 precursor	ITGB1	P53713	Toll-like receptor 4 precursor	TLR4	P58727	0.038
31	Serum albumin precursor	ALB	P49064	Transferrin receptor protein 1	TFRC	Q9MYZ3	0.038

Extrapolating Interactions to Cat Proteome

- Inferred Interactions involved 46 different proteins. We want to extrapolate to the 569 known proteins in the feline proteome.



- This can be done using the protein amino acid sequences.

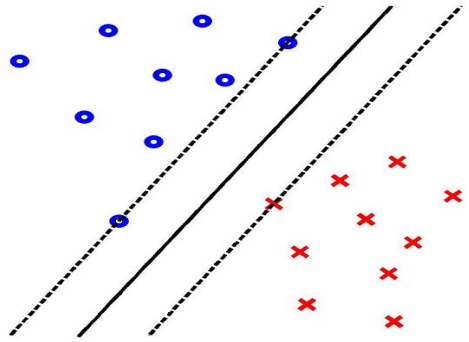
Amino Acid Tri-mer Kernel Support Vector Machine

Φ

Concept

LTRPKYRNP

$$\begin{pmatrix} \vdots \\ 1 \quad K(PY) \\ 0 \quad R(LP) \\ \vdots \end{pmatrix}$$



Implementation

$$\text{maximize} \quad W(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subject to} \quad 0 \le \alpha_i \le C, \quad i = 1, \ldots, \ell, \text{ and } \sum_{i=1}^{\ell} \alpha_i y_i = 0.$$

Replace inner product with *tri-mer kernel*

$$k(seq_1, seq_2) = (\Phi(seq_1), \Phi(seq_2))$$

and solve QP problem.

Tri-mer Product

Tri-mers can be used to describe a single peptide, but in order to describe protein-protein *pairs* we use *tri-mer products*.

$$\Phi \otimes \Phi : X \times X \to F \otimes F \cong \Re^{N^2},$$

where $\otimes$ is the tensor product, or all possible pairs.

Example

$$\Phi(\text{LTRPKYRNP}) \otimes \Phi(\text{EYVEALYDFEAQQ}\ldots\text{NKIGIFPANYVKPA}) =$$

$$\begin{pmatrix} \vdots \\ 1 \;\; \text{K(PY)A(EL)} \\ 0 \;\; \text{N(PL)A(EL)} \\ \vdots \end{pmatrix}$$

Tri-mer Product

The tri-mer product results in an explosion of tri-mer pairs.

To avoid producing and analyzing the large number of pairs we use the following relation

$$\left(\Phi \otimes \Phi\left(A,B\right),\Phi \otimes \Phi\left(C,D\right)\right) = k\left(A,C\right)k\left(B,D\right)$$

This relation allows us to use tri-mer products while avoiding actual enumeration of products.

In fact we use a symmetric tri-mer product kernel

$$k\left(A,C\right)k\left(B,D\right)+k\left(A,D\right)k\left(B,C\right).$$

Accuracy Assessment of Tri-mer SVMs

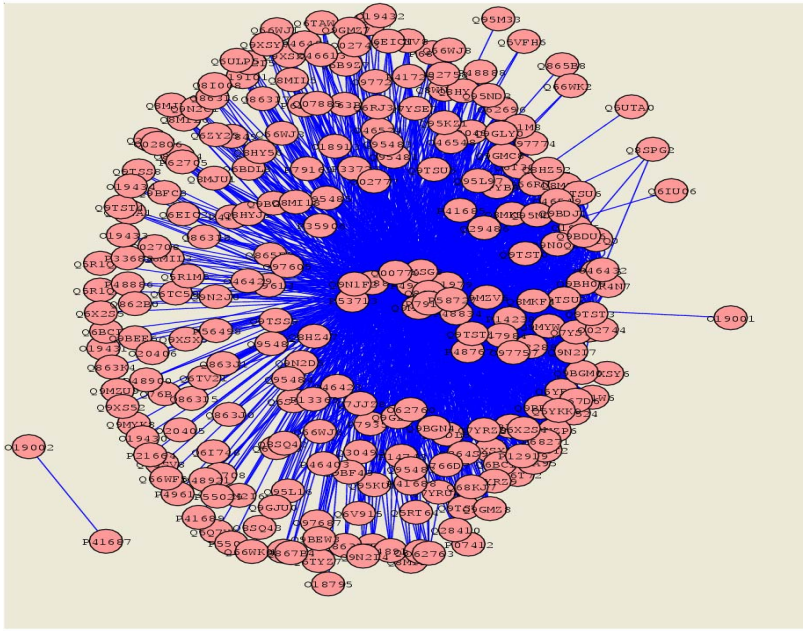
- To assess the accuracy of the extrapolation method we used 10-fold cross validation on the inferred protein interactions.

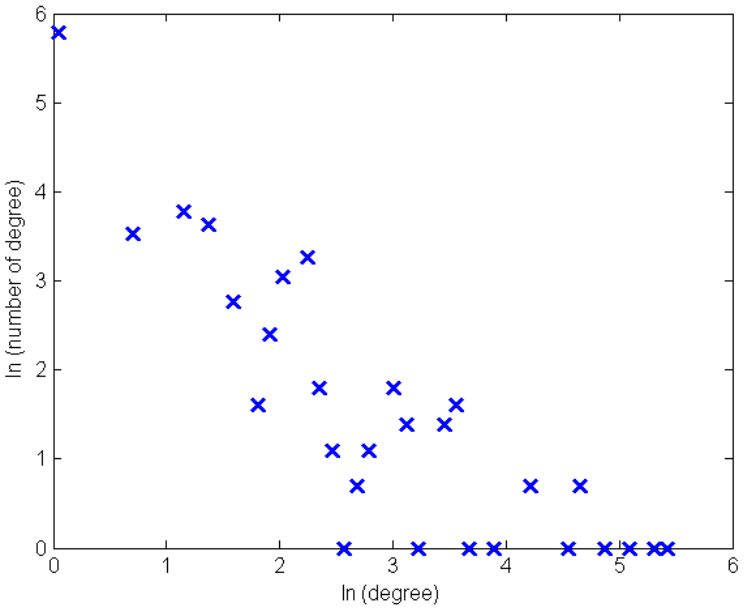
Num. Pairs	Num. Comps.	Comp. Size	Acc.	Spec.	Sens.
300	1		83.5	84.7	81.6
142	3	2	89.9	92.2	89.4
98	4	3	92.8	91.8	92.8
77	5	4	94.1	92.4	96.0
69	6	5	95.7	95.6	96.3
48	8	6	96.8	95.5	98.3
40	9	7	96.3	95.0	96.7
31	11	8	96.7	97.5	97.5

- The confidence levels were varied to obtain more or less inferred protein pairs, and the negatives (non-interactions) were taken at random.

Extrapolated Network

- Using the SVM trained on 69 protein pairs (inferred interactions from at least 6 complexes at .05) we extrapolated to the entire cat proteome.





- As a check we plotted the degree distribution of the network to see that it was scale-free.

Conclusions

- We performed a computational analysis of an experimental technique which produced protein complexes.
  - Suggested a new way to infer protein interactions from complexes
  - Added evidence to validity of experimental technique.
  - Extrapolated from limited set of proteins to entire proteome.
  - Produced hypothetical feline protein network.

Questions?