

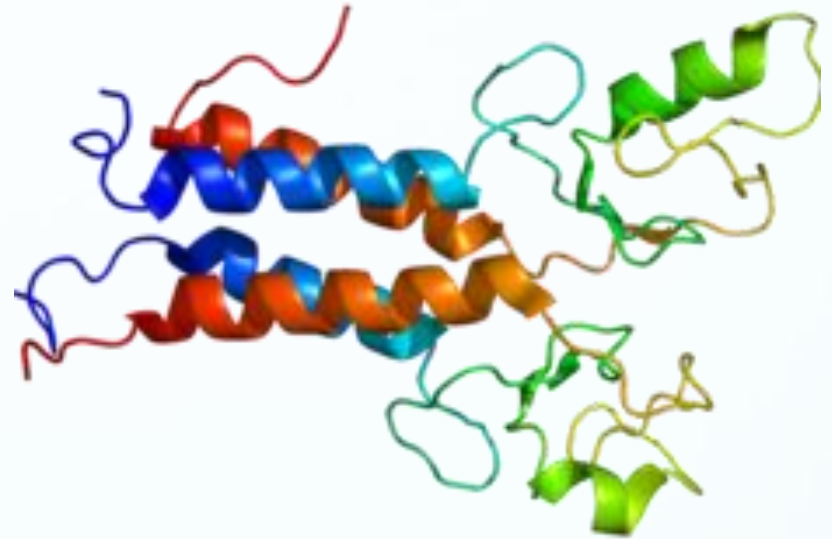
# Finding Where MoRFs Bind on Partner Proteins

The Reverse MoRF Predictor

Jonathan Patsenker

# Background

- Proteins
- Protein binding

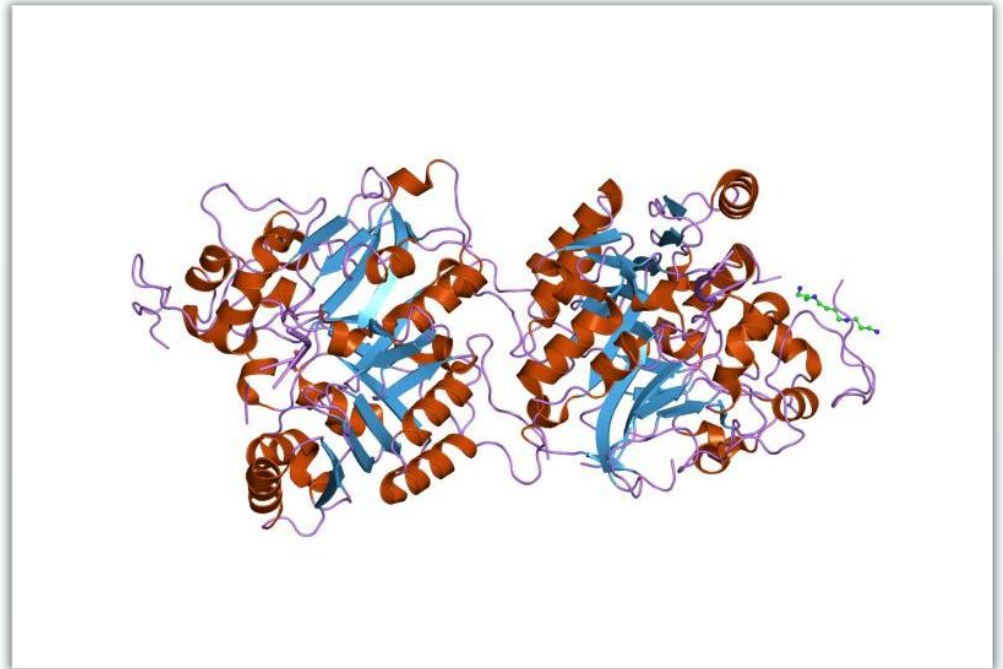


**BRCA1**

Source: <http://en.wikipedia.org/wiki/BRCA1>

# MoRFs

- Molecular recognition feature
  - Binding region
- Proteins
- Disordered regions
  - No definite structure
- Typically short

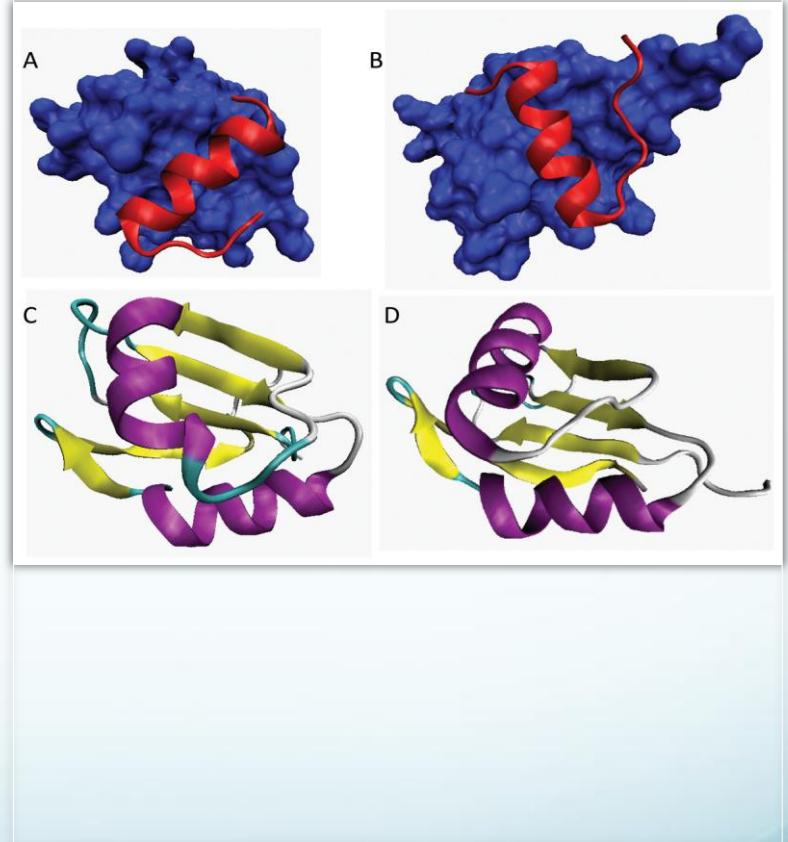


1RFI - Has a region that is a MoRF

Source: [http://commons.wikimedia.org/wiki/File:PDB\\_1rfi\\_EBI.jpg](http://commons.wikimedia.org/wiki/File:PDB_1rfi_EBI.jpg)

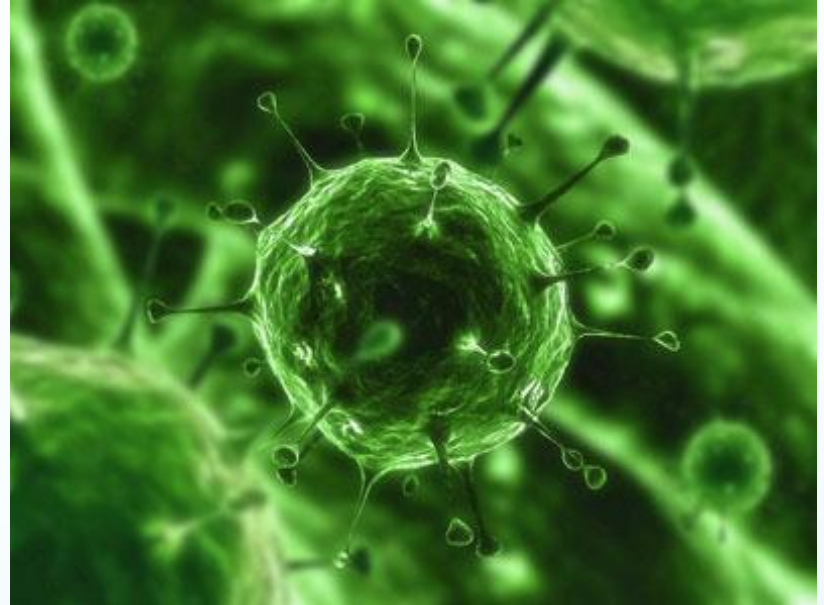
# Predicting MoRF-Protein Bind Sites

- Find regions on structured side where a MoRF is likely to bind
- Find main attributes that cause MoRF to bind on the location



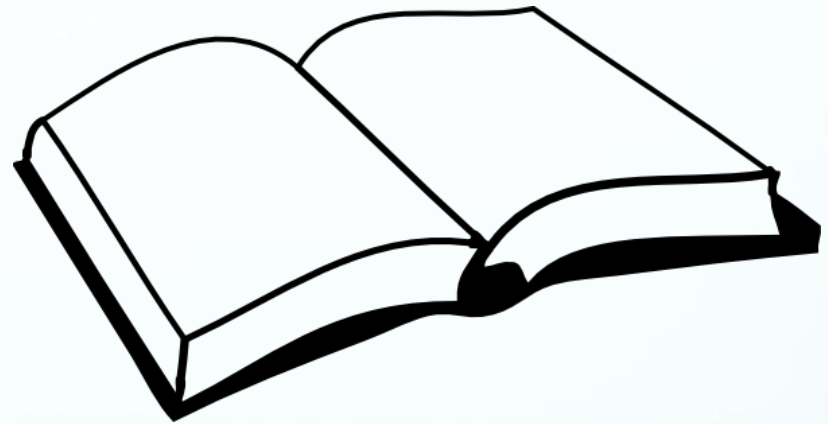
# Motivations

- Binding on partner has to do with how certain diseases function
- Useful for drug design



# Past Research

- MorfPred
  - Predicts where on MoRF protein is likely to bind
- MoRF Binary Predictor
  - Predicts whether or not MoRFs are likely to bind



# Data

- PDB (Protein Data Bank)
  - Data for attributes from here or calculated from information in here



# BRCA1

- Partner protein
- MoRFs bind to this protein
- 291 windows

ORIENTSCOR	SIDESCORE1	SIDESCORE2	GRAPHSHAP	SIZESCORE	AMINOACID:	CHARGETRAI	CHARGETRAI
1.434	4.88466667	3.61066667	3.61066667	3.61066667	5.8	0.26666667	5
1.68466667	5.126	3.68733333	3.68733333	3.68733333	5.18666667	0.26666667	7
1.454	5.09333333	3.846	3.846	3.846	5.28666667	0.2	6
1.40333333	4.976	3.76266667	3.76266667	3.76266667	5.70666667	0.2	3
1.63333333	5.076	3.896	3.896	3.896	5.62	0.2	4
1.358	4.932	4.206	4.206	4.206	4.51333333	0.06666667	9
1.43466667	5.244	4.50533333	4.50533333	4.50533333	4.68666667	0.13333333	10
1.97	5.188	4.10866667	4.10866667	4.10866667	5.52666667	0.13333333	9
2.18133333	5.00533333	3.89666667	3.89666667	3.89666667	5.50666667	0.2	9
1.92533333	5.072	3.974	3.974	3.974	5.33333333	0.13333333	7
1.99333333	5.41933333	4.02	4.02	4.02	5.58	0.2	7
1.882	5.41866667	4.026	4.026	4.026	5.88666667	0.13333333	6
1.512	5.20933333	3.95666667	3.95666667	3.95666667	5.92666667	0.13333333	4
1.406	5.018	3.88133333	3.88133333	3.88133333	5.88666667	0.2	4
1.538	4.608	3.718	3.718	3.718	5.82666667	0.2	5
1.65333333	4.38533333	3.91133333	3.91133333	3.91133333	5.86	0.33333333	6
1.99266667	4.78666667	4.18533333	4.18533333	4.18533333	5.6	0.4	9
2.112	5.12066667	4.22	4.22	4.22	5.6	0.4	8
1.83533333	4.99933333	3.834	3.834	3.834	5.38666667	0.33333333	5
1.71533333	5.07133333	3.908	3.908	3.908	5.44666667	0.26666667	6
1.572	5.38066667	4.30866667	4.30866667	4.30866667	5.49333333	0.13333333	8
1.948	5.53866667	4.47666667	4.47666667	4.47666667	5.65333333	0.26666667	8
1.92933333	5.17066667	3.876	3.876	3.876	5.84666667	0.26666667	4
1.80133333	4.79	3.63066667	3.63066667	3.63066667	6.02666667	0.2	3
1.616	4.52933333	3.32466667	3.32466667	3.32466667	6.34	0.2	3
1.57133333	4.97066667	3.87666667	3.87666667	3.87666667	6.24666667	0.13333333	5
2.03	5.40866667	4.00733333	4.00733333	4.00733333	5.56	0.06666667	7
2.194	5.51533333	4.202	4.202	4.202	5.20666667	0	8
2.00466667	4.71266667	3.562	3.562	3.562	5.9	0.2	4
1.60066667	4.458	3.45466667	3.45466667	3.45466667	6.08666667	0.4	2
1.79733333	4.48	3.506	3.506	3.506	5.97333333	0.26666667	4
1.738	4.51666667	3.524	3.524	3.524	5.96	0.26666667	5
1.85533333	4.75733333	3.63666667	3.63666667	3.63666667	5.47333333	0.2	5
1.972	5.12666667	3.65066667	3.65066667	3.65066667	5.58	0.26666667	4
1.95333333	4.754	3.50266667	3.50266667	3.50266667	6.18	0.26666667	3
1.77533333	4.712	3.52066667	3.52066667	3.52066667	6.29333333	0.26666667	4
1.76266667	4.562	3.37333333	3.37333333	3.37333333	6.38666667	0.2	4
1.666	4.88666667	3.84933333	3.84933333	3.84933333	4.94	0.33333333	6



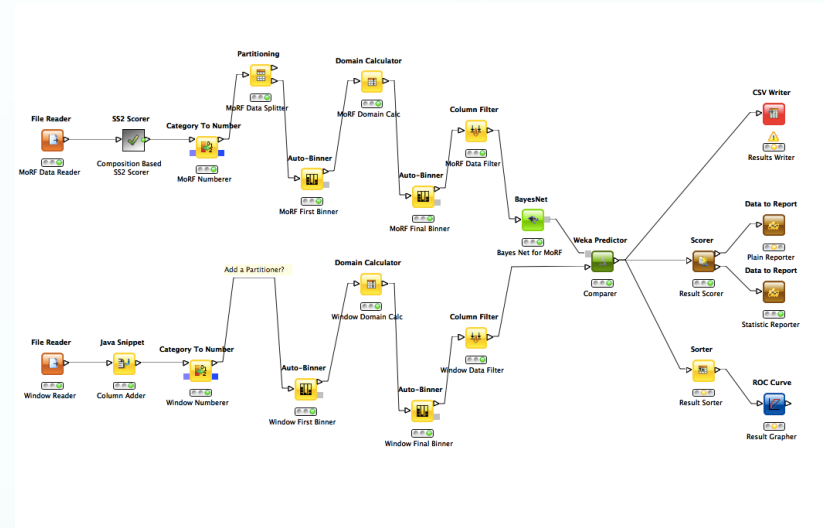
# MoRFs Used

- Using a list of MoRFs
- 812 MoRFs

PDB_ID	UNIPROT_ID	SS2	SP	MORF	VSL2B
1rfi_C	Q5UNT1	CCCCC	MDQYNKYIT	0000000000	11111
1d9k_P	P02789	CCSSCCCCC	MKLILCTVLSI	0000000000	10000
2bck_C	O14746	CCCCCCCCC	MPRAPRCRA	0000000000	11111
1j19_B	P35330	CCSSSSCCH	MSSFACWSL	0000000000	11110
1avp_B	P03274	CCSSSSSSSC	MEDINFASLA	1.1111E+10	1.11
1tg4_I	Q5QQ53	CCCCC	MEQSVSARV	0000000000	11111
1qr1_C	P04626	CCCCCCCCC	MELAALCRW	0000000000	11000
1br8_P	P01008	CSSSSSSSSS	MYSNVIGTV	0000000000	11111
1fch_C	COJAS2	CCCCC	WIMGHMVN	0000000000	11000
1osv_C	Q9WUI9	CCHHHHHHH	MSGMGENT	0000000000	11111
1q3p_C	P97838	CCSSSC	MRGYHGDR	0000000000	11111
1l7z_B	Q1RM09	CCCCCC	MGGKLSKKK	1.11E+195	1.11
3bze_P	Q95IT1	CCCCCCCCC	MVVMAPRTI	0011111111	11100
1yyp_B	Q65W77	CCCCCCHHH	MFFNPYLSG	0000000000	11111
3bxk_B	P97445	CHHHHHHHH	MARFGDEMI	0000000000	11111
3bhb_C	Q86UW6	CCCCCCCCC	MPRRRNKLG	0000000000	11111
2h pz_B	Q754R6	CCCCCCCCC	MDYWILLVL	1.11E+111	1.11
1s7r_C	Q14HX3	CCCCCCCCC	MQKNILVLG	0000000000	11110
2w0p_C	Q71FD7	CSSSSSSSC	MASKPEKRV	0000111111	11111
3d vu_C	Q5R878	CHHHHHHHH	MEGSKTSNN	0000000000	11111
1d4t_B	Q13291	CCSSSSCSC	MDPKLLSLT	0000000000	11111
1jmq_P	P97764	CCCCCCCCC	MARASSRNS	0000000000	11111
1ozs_B	Q8MKD5	CCCCCCCCC	MADESGDAF	1.11E+81	1.11
1l vb_C	P04517	CCSSSSCSC	MALIFGTVN	0000000000	00000
1t1w_C	B8F7P4	CCCCCCCCC	MNPKIIFLDA	0000000000	11110
1jd5_B	Q24570	CSSSSCC	MAIAYFIPDQ	1.11E+136	1.00
1vyt_E	Q13936	CCCHHHHHH	MVNENTRM	0000000000	11111
1t01_B	Q9Y490	CHHHHHHHH	MVALSLKISIC	0000000000	11110
1gbr_B	Q02384	CCCCCCCCC	MQQAPQPYI	0000000000	11111
1q1t_A	P05768	CCCCCCCCC	MEYVYAALIL	1.11E+45	1.10
1o9u_B	O70239	CHHHHHHHH	MNVQEQGF	0000000000	11111
1b0g_C	Q9NPA0	CCCCCCCCC	MAAALWGF	1.11E+238	1.00
1i4o_C	P98170	CCHHHHHHH	MTFNSFEGS	0000000000	11111
2gtk_B	Q4PJW2	CHHHHHHHH	MSGLGDSST	0000000000	11111
1zgl_C	P02688	CCSSCCCCC	MASQKRPSQ	1.11E+84	1.11
2q5d_D	O95149	CHHHHHHHH	MEELSQLAS	0000000000	11111
1pfb_B	Q5RCC9	CCSSSSSCC	MARTKQTAR	1.11E+115	1.11
1l0a_B	Q92844	CCSSCCCCC	MDKNIGEQL	0000000000	11111
1jbu_X	A1ALQ2	CCSSSSCC	MATDLLYGLI	0000000000	11110
1sjh_C	O89939	CSSSSSSC	MGARASVLT	0000000000	11111

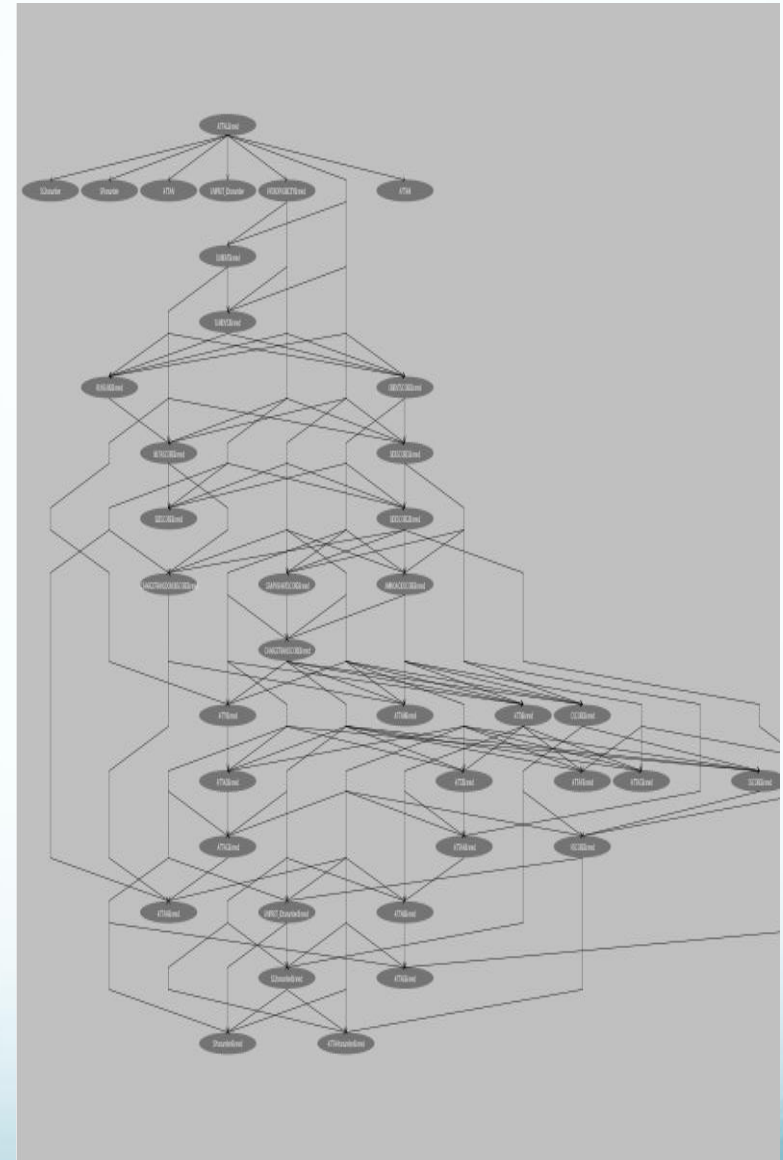
# Methods

- Bayes Nets
- Windowed Approach
- Molecular Geometry



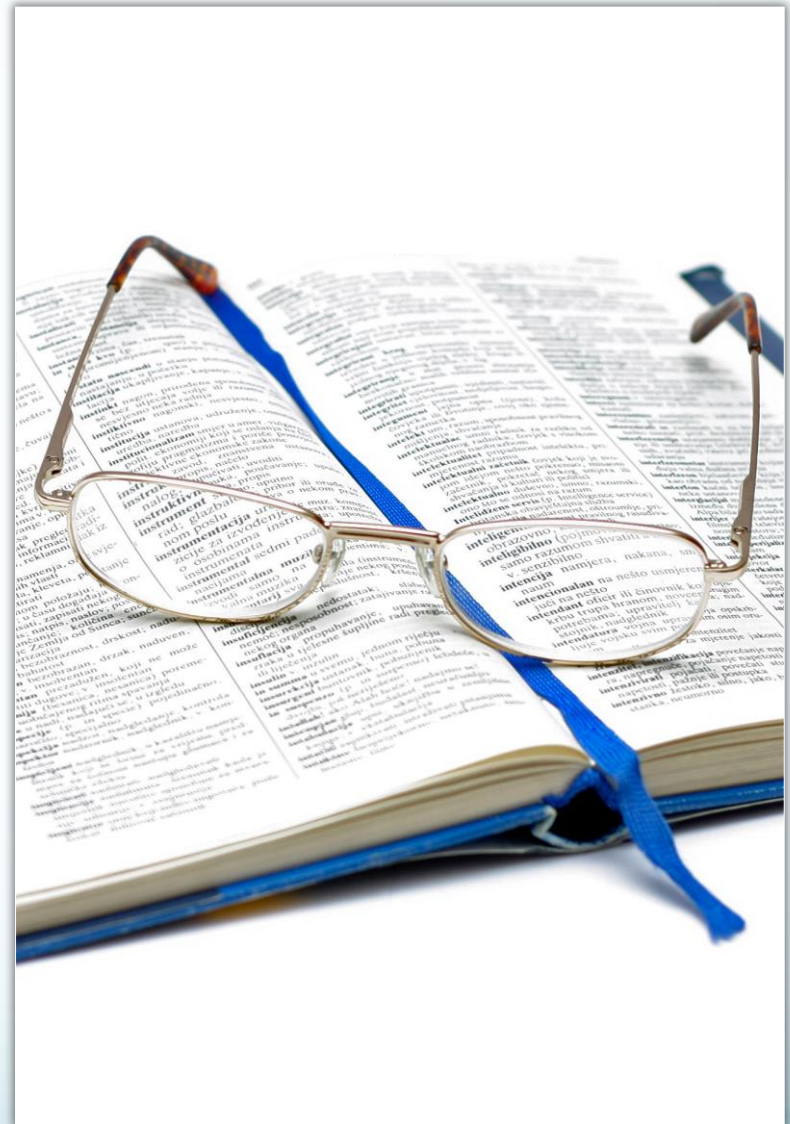
# Bayes Nets

- Attributes
  - On MoRF
  - On protein
- Implemented with Knime



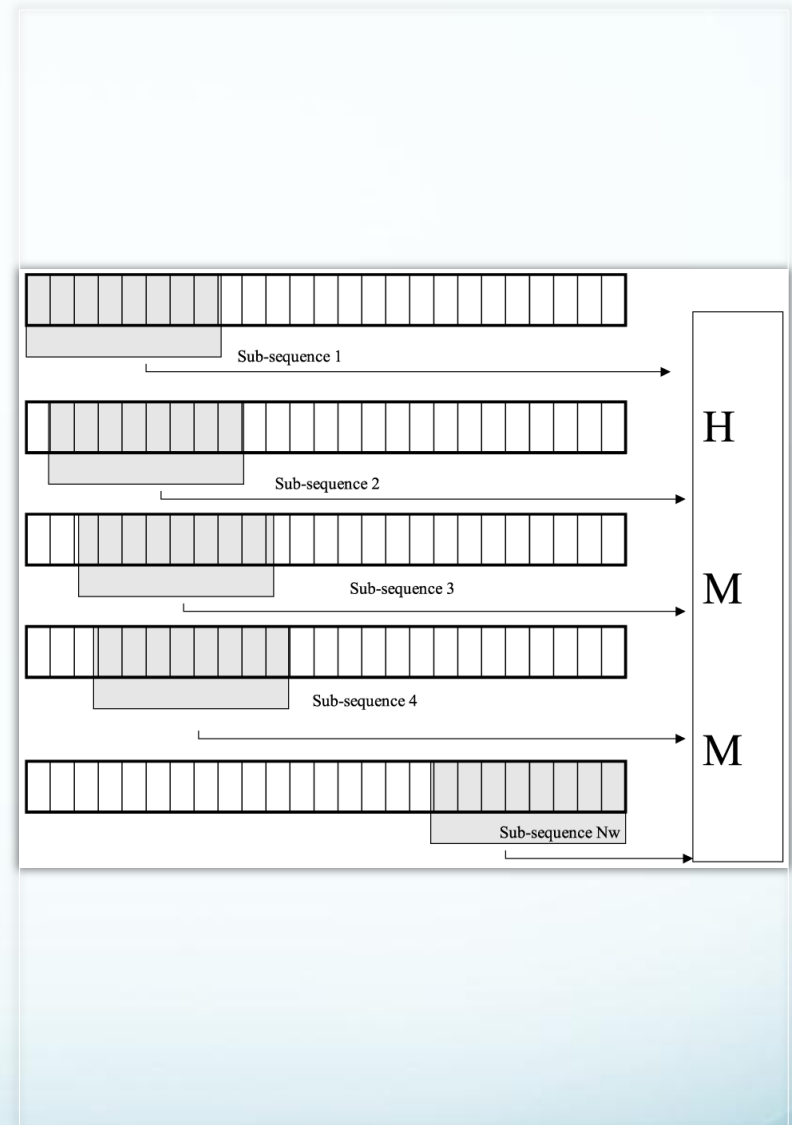
# Learning

- MoRF data for bound and unbound states
- Finds relationships between each of MoRF's attributes
- Creates a Bayes Net with all the relationships as they influence MoRF's state



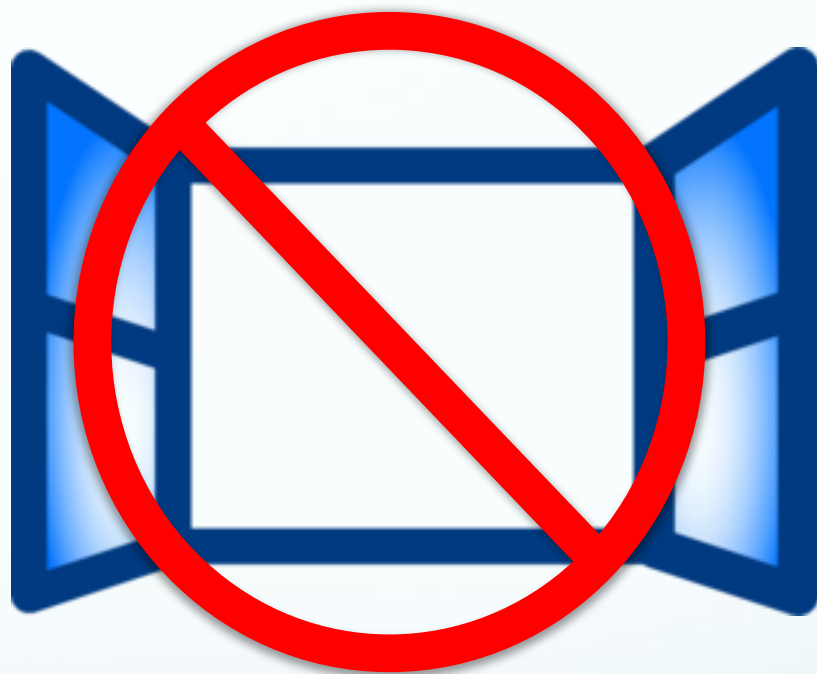
# Windowed Approach

- Windows on rigid protein
- Each window is assessed
- Window length is 15
- Predictor gives more favorable windows based on learned Bayes Net



# Deleting Windows

- Based on the molecular geometry
- Physically inaccessible areas



# Limitations

- False Attachment Points
- Extra Window Deletion

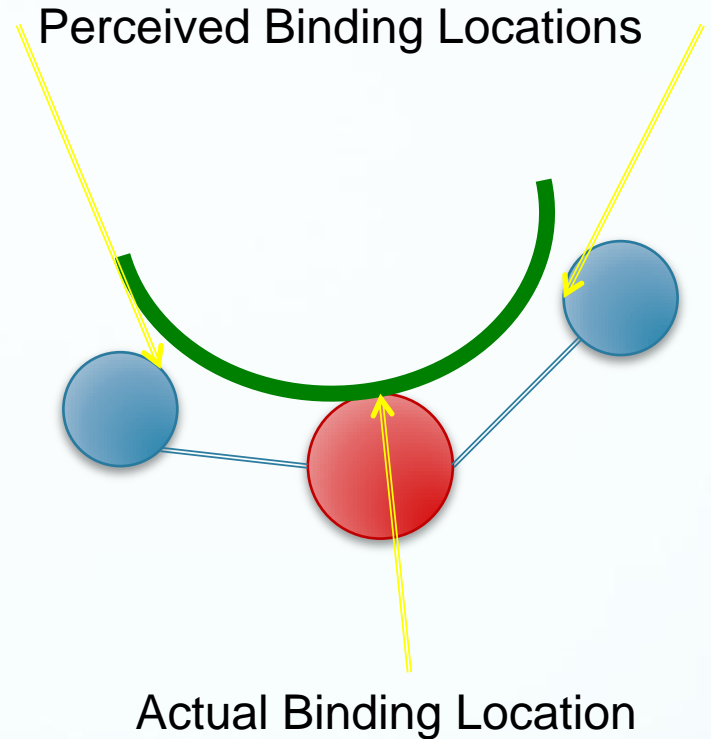
# Results

- Accuracy: ~40%
- $p < .0001$



# Future Research

- Improvement of Accuracy
- Fixing the deceptive attachments problem
- Implementing Window Deletion



# Thank You

- Dr. Alterovitz
- Jonah Kallenbach
- Wei-Lun Hsu
- Michael and Svetlana Patsenker