The Significance of Disordered Residues in: 1) Bacterial Drug Resistance and 2) SNP Interactions in Relation to Disease Associations

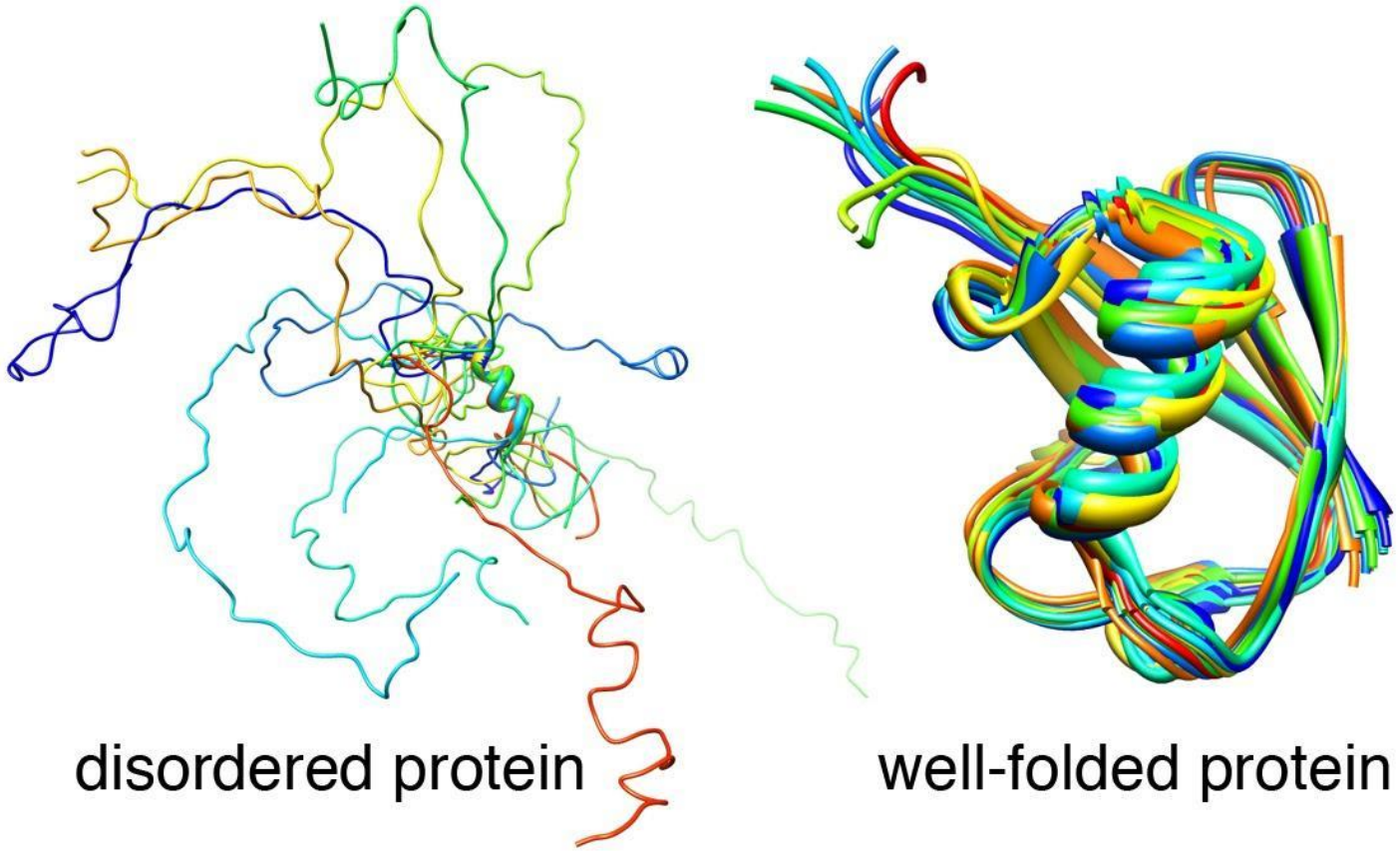
By Arul Prasad
PRIMES Conference
Mentor: Dr. Gil Alterovitz
May 17, 2015

Outline

1. **Introduction**
2. Bacteria Methods
3. Bacteria Results
4. Disease Association Methods
5. Disease Association Results
6. Conclusion

Introduction

- What is a disordered protein?



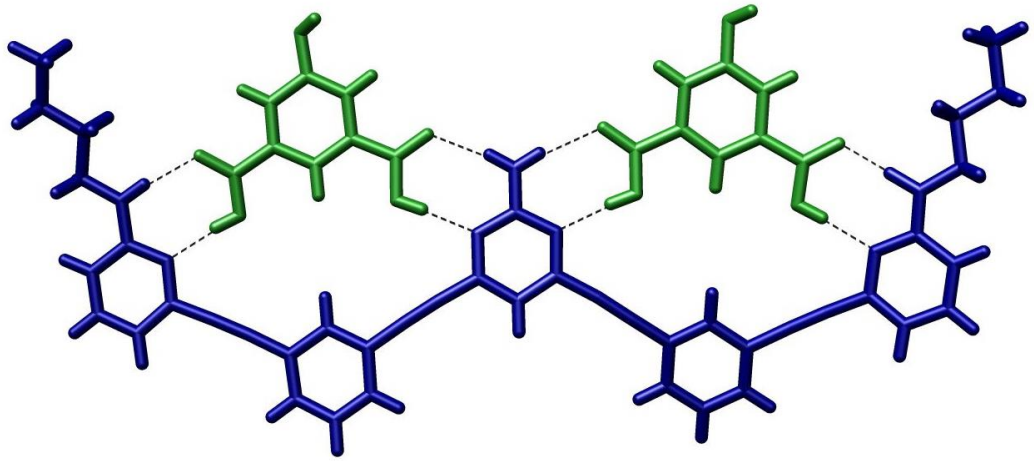
Introduction (Cont.)

- Previous research on disordered proteins in relation to drug resistance

	Bacteria Proteome (1145411)			Drug Resistance Proteins (38)				
Threshold	Disordered	Ordered	Disordered p	Disordered	Ordered	Disordered p	Percent Chai	fisher p-value
1	371085	774326	0.3241	17	21	0.4474	12.33%	**0.0369023**
5	371085	774326	0.3241	17	21	0.4474	12.33%	**0.0369023**
10	278085	867326	0.2428	11	27	0.2895	4.67%	0.1139565
15	179389	966022	0.1566	5	33	0.1316	-2.50%	0.1712697
20	134122	1011289	0.1171	5	33	0.1316	1.45%	0.181349
25	100226	1045185	0.0875	2	36	0.0526	-3.49%	0.1992268
30	73164	1072247	0.0639	0	38	0	-6.39%	0.0814126
35	57601	1087810	0.0503	0	38	0	-5.03%	0.1407665
40	43343	1102068	0.0371	0	38	0	-3.71%	0.2308866

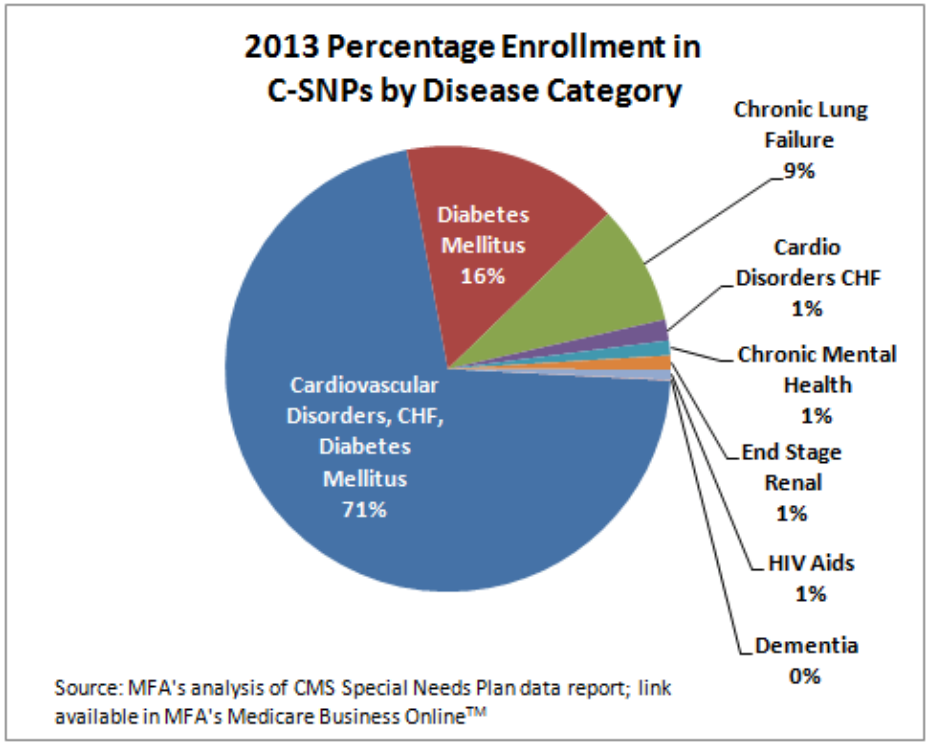
Introduction (Cont.)

- Significant threshold range potentially equivalent to MORF length
- What is a MORF?



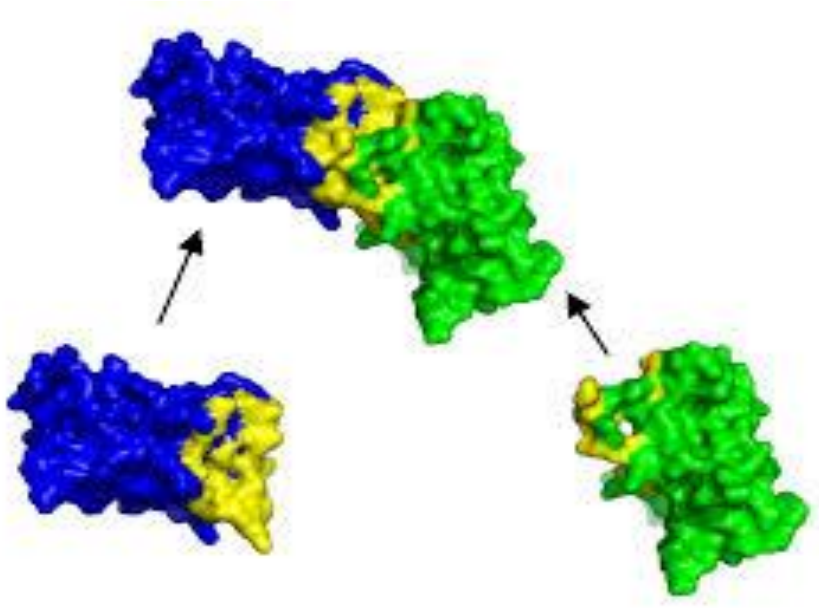
Introduction (Cont.)

- What is a SNP?
- SNPs have been found to be disease associated



Introduction (Cont.)

- Protein interaction related
  - Domain-domain
- Objective is to find SNPs in which type of interaction cause disease

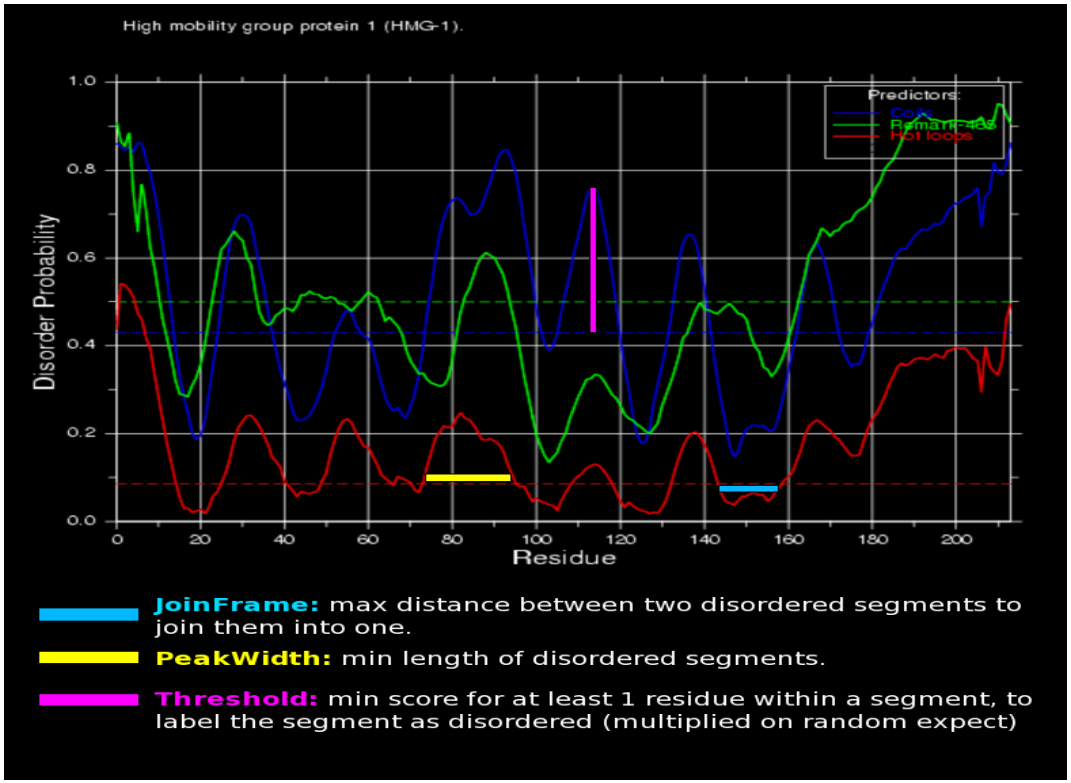


Outline

1. Introduction
2. **Bacteria Methods**
3. Bacteria Results
4. Disease Association Methods
5. Disease Association Results
6. Conclusion

Bacteria Methods

- Collect bacteria PDBs and amino acid sequences
- Run through DisEMBL pipeline



Outline

1. Introduction
2. Bacteria Methods
3. **Bacteria Results**
4. Disease Association Methods
5. Disease Association Results
6. Conclusion

Bacteria Results

1 to 39

Threshold	Disordered	Ordered	Dresistance	Oresistance	P-value
1	15678	16713	695	618	0.00012651
2	15678	16713	695	618	0.00012651
3	15678	16713	695	618	0.00012651
4	15678	16713	695	618	0.00012651
5	15678	16713	695	618	0.00012651
6	15678	16713	695	618	0.00012651
7	15678	16713	695	618	0.00012651
8	15678	16713	695	618	0.00012651
9	12719	19672	564	749	0.000644801
10	10561	21830	476	837	0.000549522
11	8589	23802	396	917	0.000375914
12	7118	25273	343	970	0.00005967
13	5994	26397	300	1013	1.5588E-005
14	4896	27495	252	1061	1.3996E-005
15	4144	28247	216	1097	2.7605E-005
16	3493	28898	174	1139	0.000785313
17	3067	29324	156	1157	0.000648922
18	2628	29763	131	1182	0.002407791
19	2350	30041	107	1206	0.019970993
20	2049	30342	82	1231	0.046033593
21	1797	30594	67	1246	0.0398393
22	1604	30787	50	1263	0.008680559
23	1421	30970	40	1273	0.003140312
24	1285	31106	35	1278	0.002892637
25	1129	31262	29	1284	0.00223885
26	990	31401	28	1285	0.010059851
27	908	31483	28	1285	0.024741835
28	853	31538	25	1288	0.019077118
29	791	31600	20	1293	0.006804668
30	721	31670	16	1297	0.002944621
31	672	31719	14	1299	0.002267782
32	624	31767	12	1301	0.001614064
33	589	31802	12	1301	0.00331164
34	536	31855	6	1307	6.1602E-005
35	480	31911	4	1309	2.4836E-005
36	446	31945	3	1310	1.6401E-005
37	412	31979	3	1310	5.0717E-005
38	378	32013	2	1311	3.0715E-005
39	344	32047	2	1311	9.9576E-005

40 to 79

40	332	32059	2	1311	0.000150093
41	306	32085	2	1311	0.000361605
42	299	32092	2	1311	0.000457051
43	285	32106	2	1311	0.000727631
44	276	32115	2	1311	0.000978578
45	256	32135	2	1311	0.001875175
46	241	32150	2	1311	0.003029251
47	237	32154	2	1311	0.003438072
48	235	32156	2	1311	0.003661934
49	228	32163	2	1311	0.004561231
50	221	32170	2	1311	0.005670472
51	217	32174	2	1311	0.00641573
52	204	32187	2	1311	0.009536888
53	204	32187	2	1311	0.009536888
54	204	32187	2	1311	0.009536888
55	198	32193	2	1311	0.011420572
56	194	32197	2	1311	0.012865629
57	193	32198	2	1311	0.013252849
58	192	32199	2	1311	0.013650985
59	180	32211	2	1311	0.019386451
60	168	32223	2	1311	0.027286725
61	140	32251	2	1311	0.058056698
62	139	32252	2	1311	0.059565685
63	135	32256	2	1311	0.06593556
64	128	32263	2	1311	0.078431892
65	123	32268	2	1311	0.088465744
66	118	32273	2	1311	0.099458593
67	114	32277	2	1311	0.108952453
68	113	32278	2	1311	0.111422868
69	109	32282	2	1311	0.121686915
70	103	32288	2	1311	0.138187886
71	93	32298	2	1311	0.16823719
72	80	32311	2	1311	0.209875231
73	73	32318	2	1311	0.231660846
74	73	32318	2	1311	0.231660846
75	73	32318	2	1311	0.231660846
76	71	32320	2	1311	0.237548556
77	71	32320	2	1311	0.237548556
78	71	32320	2	1311	0.237548556
79	69	32322	2	1311	0.24321408

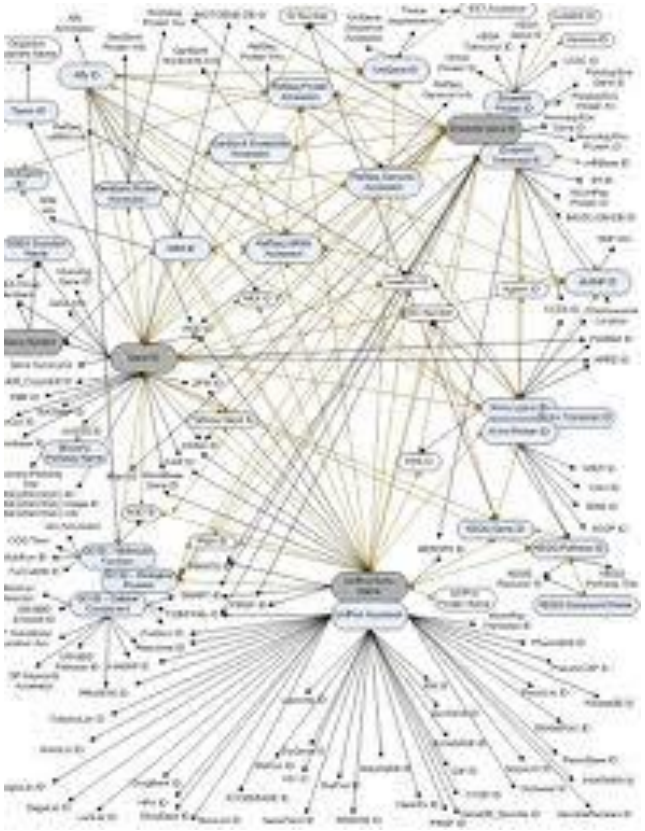
80	64	32327	2	1311	0.25610536
81	64	32327	2	1311	0.25610536
82	64	32327	2	1311	0.25610536
83	64	32327	2	1311	0.25610536
84	64	32327	2	1311	0.25610536
85	64	32327	2	1311	0.25610536
86	62	32329	2	1311	0.260621268
87	62	32329	2	1311	0.260621268
88	62	32329	2	1311	0.260621268
89	62	32329	2	1311	0.260621268
90	62	32329	2	1311	0.260621268
91	62	32329	2	1311	0.260621268
92	62	32329	2	1311	0.260621268
93	62	32329	2	1311	0.260621268
94	62	32329	2	1311	0.260621268
95	59	32332	2	1311	0.266537563
96	57	32334	2	1311	0.269823681
97	55	32336	2	1311	0.272510706
98	53	32338	2	1311	0.274533704
99	46	32345	2	1311	0.275408097
100	46	32345	2	1311	0.275408097

Outline

1. Introduction
2. Bacteria Methods
3. Bacteria Results
4. **Disease Association Methods**
5. Disease Association Results
6. Conclusion

Disease Association Methods

- Take in protein interaction data and retrieve SNPs present in those proteins
- Map SNPs to associated diseases
- Statistical analysis



Outline

1. Introduction
2. Bacteria Methods
3. Bacteria Results
4. Disease Association Methods
5. **Disease Association Results**
6. Conclusion

Disease Association Results

Disease associated	Non-disease associated:	Percentages
Domain : 3878	Domain : 39122	Domain : 0.0901860465116
_rare__binding__SP_ : 0	_rare__binding__SP_ : 6700	_rare__binding__SP_ : 0.0
_rare__binding__MORF_ : 3	_rare__binding__MORF_ : 906	_rare__binding__MORF_ : 0.003300330033
_rare__binding__ : 244	_rare__binding__ : 19365	_rare__binding__ : 0.0124432658473
_rare__nonbinding__SP_ : 95	_rare__nonbinding__SP_ : 6205	_rare__nonbinding__SP_ : 0.0150793650794
_rare__nonbinding__MORF_ : 0	_rare__nonbinding__MORF_ : 230	_rare__nonbinding__MORF_ : 0.0
_rare__nonbinding__ : 216	_rare__nonbinding__ : 17314	_rare__nonbinding__ : 0.01232173417
_rare___SP_ : 391	_rare___SP_ : 21609	_rare___SP_ : 0.0177727272727
_rare___MORF_ : 404	_rare___MORF_ : 10525	_rare___MORF_ : 0.0369658706195
_rare___ : 1157	_rare___ : 54772	_rare___ : 0.0206869423734
_com__binding__SP_ : 1382	_com__binding__SP_ : 5618	_com__binding__SP_ : 0.197428571429
_com__binding__MORF_ : 113	_com__binding__MORF_ : 788	_com__binding__MORF_ : 0.125416204218
_com__binding__ : 3675	_com__binding__ : 24226	_com__binding__ : 0.131715709114
_com__nonbinding__SP_ : 1231	_com__nonbinding__SP_ : 6769	_com__nonbinding__SP_ : 0.153875
_com__nonbinding__MORF_ : 27	_com__nonbinding__MORF_ : 200	_com__nonbinding__MORF_ : 0.118942731278
_com__nonbinding__ : 2956	_com__nonbinding__ : 17771	_com__nonbinding__ : 0.142615911613
_com___SP_ : 4592	_com___SP_ : 21158	_com___SP_ : 0.178330097087
_com___MORF_ : 2991	_com___MORF_ : 8047	_com___MORF_ : 0.270973002355
_com___ : 11461	_com___ : 57827	_com___ : 0.165411037986
__binding__SP_ : 1382	__binding__SP_ : 12018	__binding__SP_ : 0.103134328358
__binding__MORF_ : 113	__binding__MORF_ : 1657	__binding__MORF_ : 0.0638418079096
__binding__ : 3675	__binding__ : 34495	__binding__ : 0.0962798008908
__nonbinding__SP_ : 1231	__nonbinding__SP_ : 10169	__nonbinding__SP_ : 0.10798245614
__nonbinding__MORF_ : 27	__nonbinding__MORF_ : 423	__nonbinding__MORF_ : 0.06
__nonbinding__ : 2956	__nonbinding__ : 28894	__nonbinding__ : 0.0928100470958
___SP_ : 4592	___SP_ : 36708	___SP_ : 0.111186440678
___MORF_ : 2991	___MORF_ : 14939	___MORF_ : 0.166815393196
___ : 11463	___ : 90769	___ : 0.112127318257

Outline

1. Introduction
2. Bacteria Methods
3. Bacteria Results
4. Disease Association Methods
5. Disease Association Results
6. **Conclusion**

Conclusion

1. Disordered residues approximately the size of MoRFs are significant in causing drug resistance in bacteria.
2. Protein interactions with
   a. common
   b. MoRF
   c. Non-domain

   SNPs are significant in disease associa

Future Work

1. Expanding the DisEMBL results
2. Target SNPs in protein interactions that were found to be disease associated

Acknowledgements

- MIT PRIMES program
- Dr. Gil Alterovitz
- Mentor group
- Parents