CASP 13

Predicting Contacts

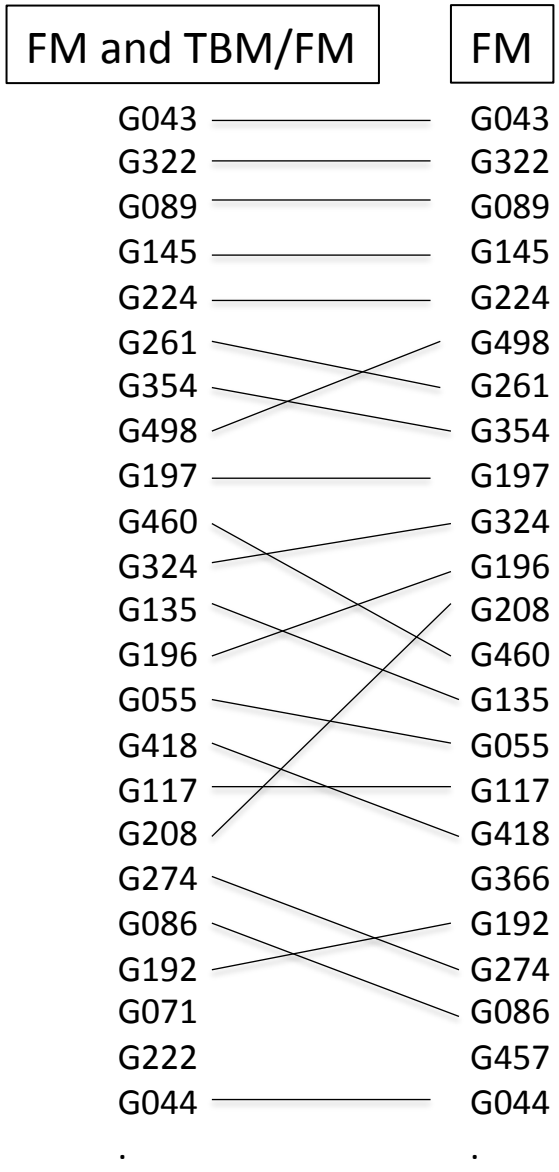
Assessor: András Fiser
Department of Systems and Computational Biology
Department of Biochemistry

Possible questions

- Does contact prediction accuracy correlate with that of structure modeling?
- How well you did among yourselves?
- How well you did compared to previous CASPs?

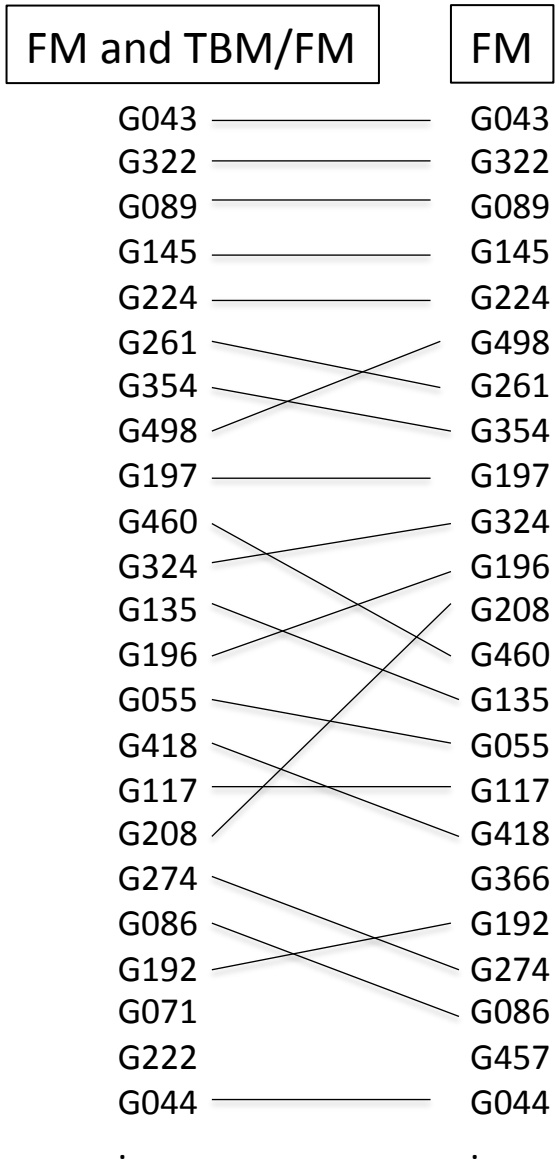
- Some insight analysis:
  - Are you capturing the same set of contacts?
  - Are there particular types of contacts that you are getting accurately?
  - How important is the quality of sequence information?

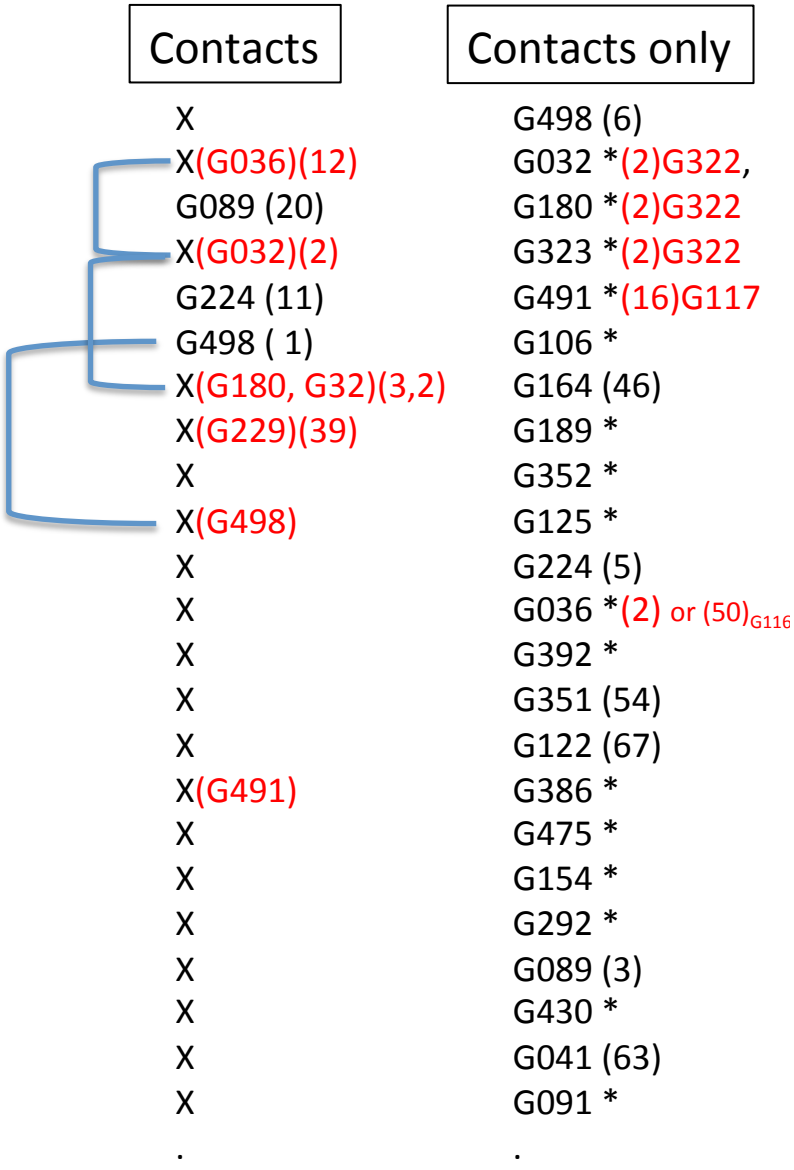
Best structure prediction (out of 98) vs. Best contact predictions (out of 46)



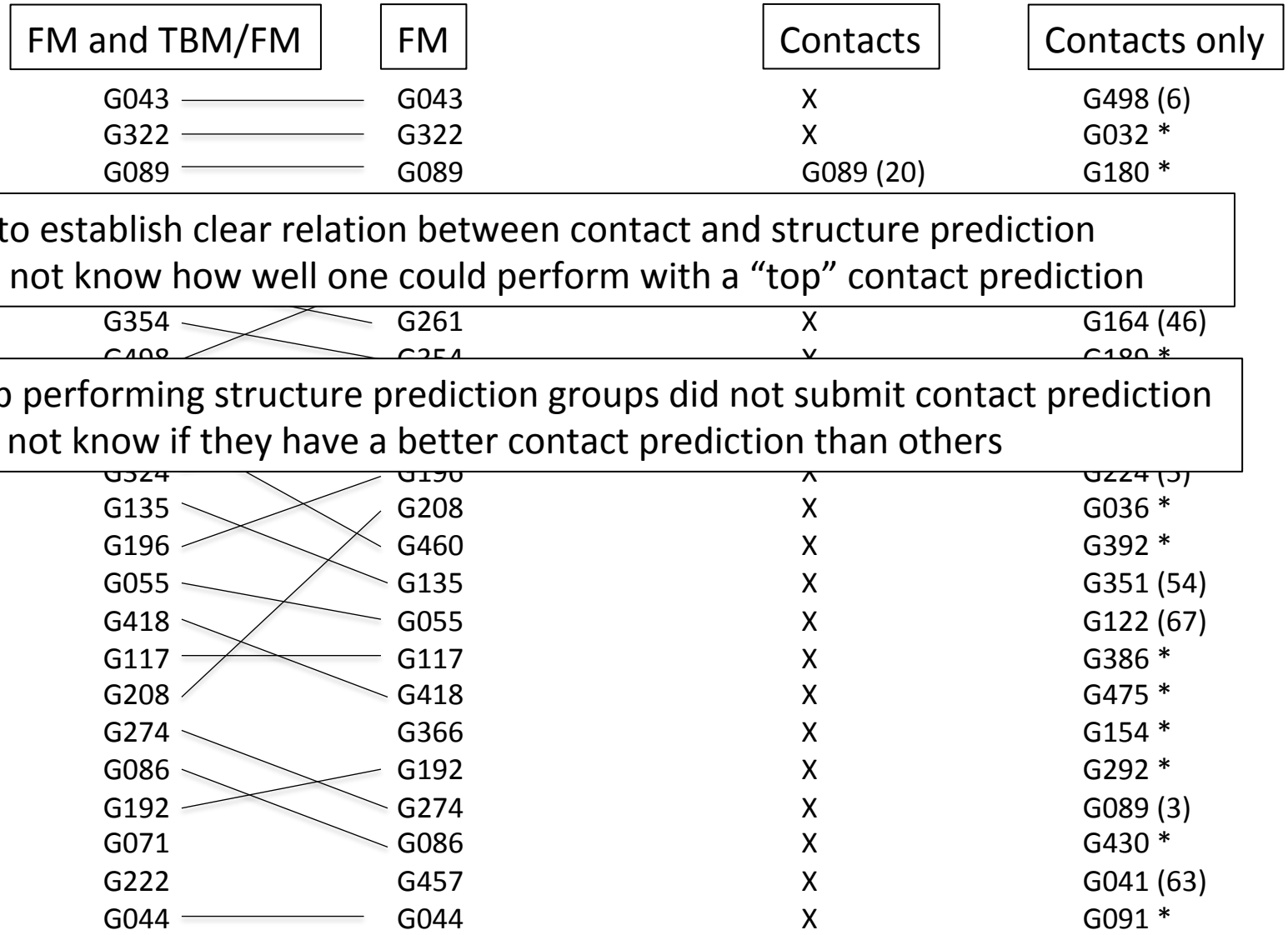
Contacts	Contacts only
X	G498 (6)
X	G032 *
G089 (20)	G180 *
X	G323 *
G224 (11)	G491 *
G498 (1)	G106 *
X	G164 (46)
X	G189 *
X	G352 *
X	G125 *
X	G224 (5)
X	G036 *
X	G392 *
X	G351 (54)
X	G122 (67)
X	G386 *
X	G475 *
X	G154 *
X	G292 *
X	G089 (3)
X	G430 *
X	G041 (63)
X	G091 *
.	.

Best structure prediction (out of 98) vs. Best contact predictions (out of 46)





Best structure prediction (out of 98) vs. Best contact predictions (out of 46)



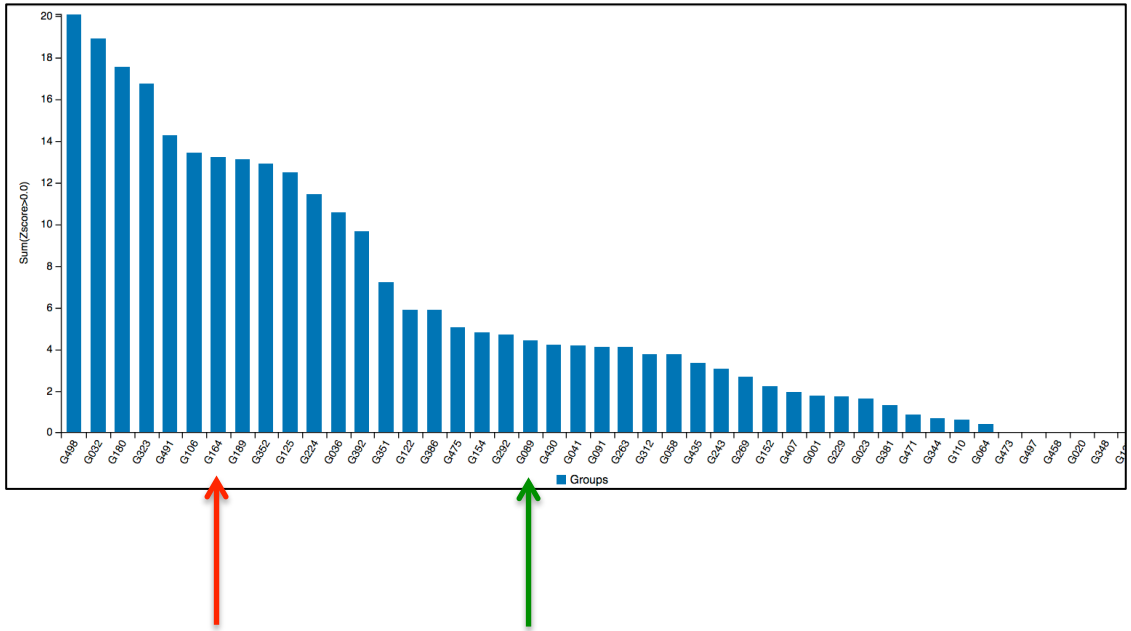
Difficult to establish clear relation between contact and structure prediction
-> we do not know how well one could perform with a “top” contact prediction

Some top performing structure prediction groups did not submit contact prediction
-> we do not know if they have a better contact prediction than others

Best structure prediction (out of 98) vs. Best contact predictions (out of 46)

Among groups that have submitted both structure and contact prediction: Surprising inconsistencies!!

It is important to know how to use contact information!
And/Or
Contact information is not as important as one thought

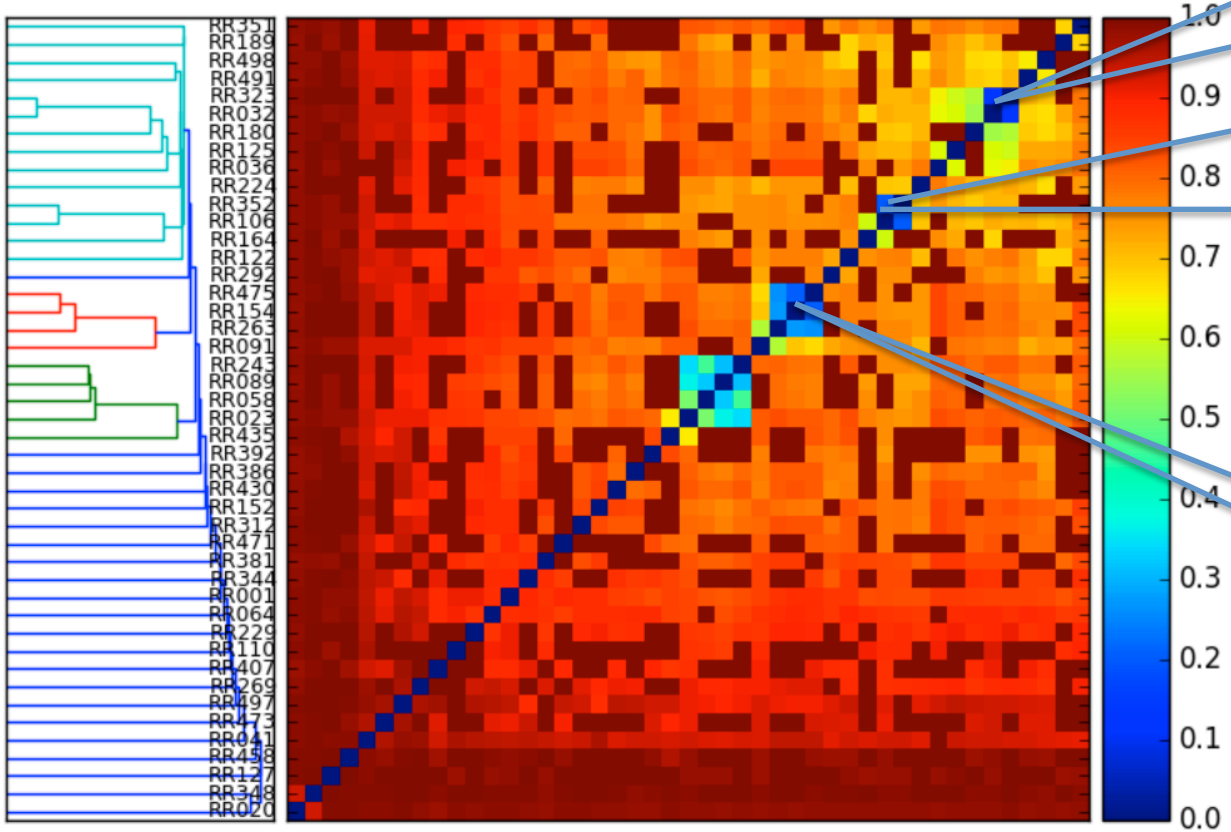


Contacts	Contacts only
X	498 (6)
X	032 *
089 (20)	180 *
X	323 *
224 (11)	491 *
498 (1)	106 *
X	164 (46)
X	189 *
X	352 *
X	125 *
X	224 (5)
X	036 *
X	392 *
X	351 (54)
X	122 (67)
X	386 *
X	475 *
X	154 *
X	292 *
X	089 (3) 89 submitted: 30/31 targets…)
X	430 *
X	041 (63)
X	091 *

Are we predicting different contacts? Jaccard distance ("1-Intersection over Union")

$$dj = \frac{|A \bigcup B| - |A \bigcap B|}{|A \bigcup B|}$$

0 (same) < d_j <1 (different)



Contacts only

498 (6)
032 *
180 *
323 *
491 *
106 *
164 (46)
189 *
352 *
125 *
224 (5)
036 *
392 *
351 (54)
122 (67)
386 *
475 *
154 *
292 *
089 (3)
430 *
041 (63)
091 *

Tripletres

Tripletres_AT

(RRMD)

(RRMD-plus)

Deltacontact

Gammacontact

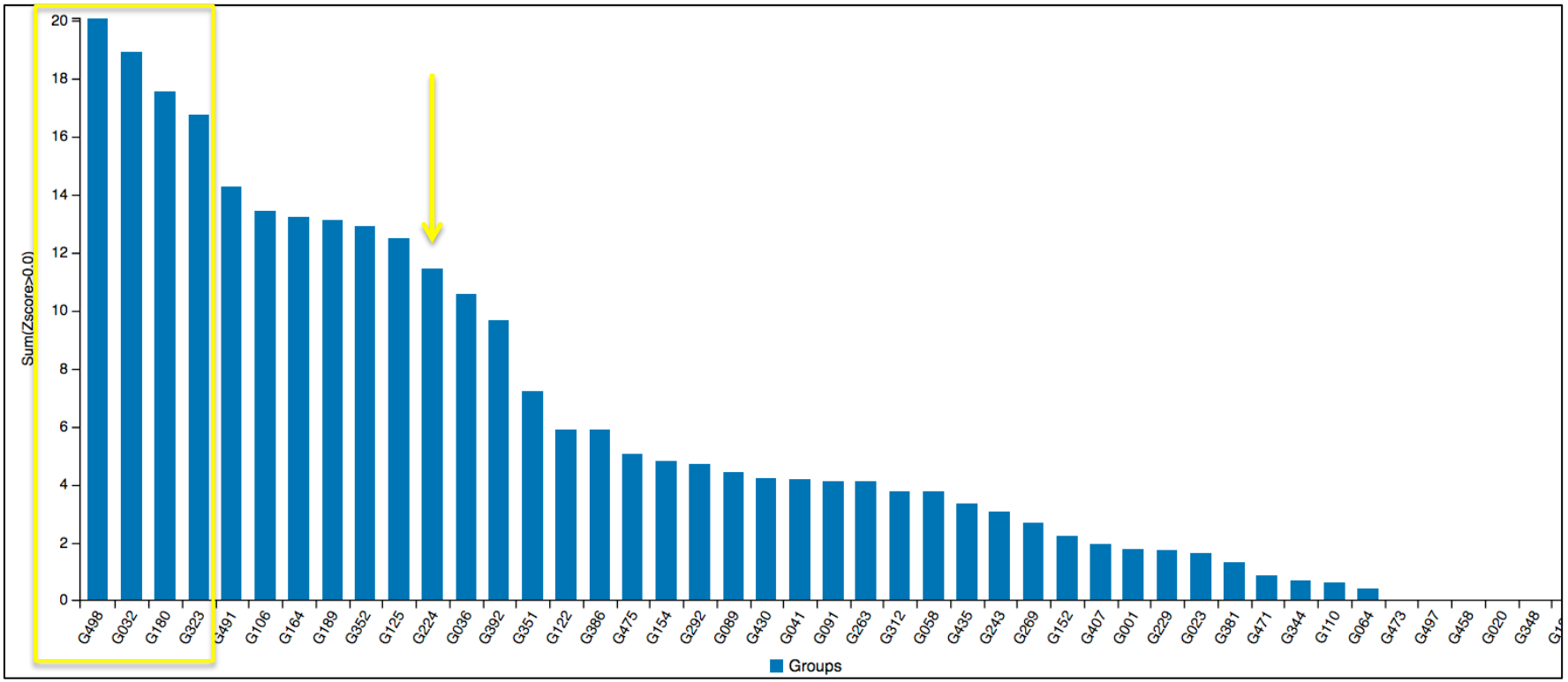
Top L/5 number of contacts, L is the length of sequence

Performance using different criteria

Models (FM or FM+TBM) **X** Contacts (top10 or L/5 or L/2 or L or FL) **X** probability (0 or 0.5) **X** contact definition (medium/long; long; extra long) => 60 combinations

evaluated by either: using F1; Precision/Recall; Z-score sum or Z-score average etc.

Long/medium contacts, FM only, Zscore >0



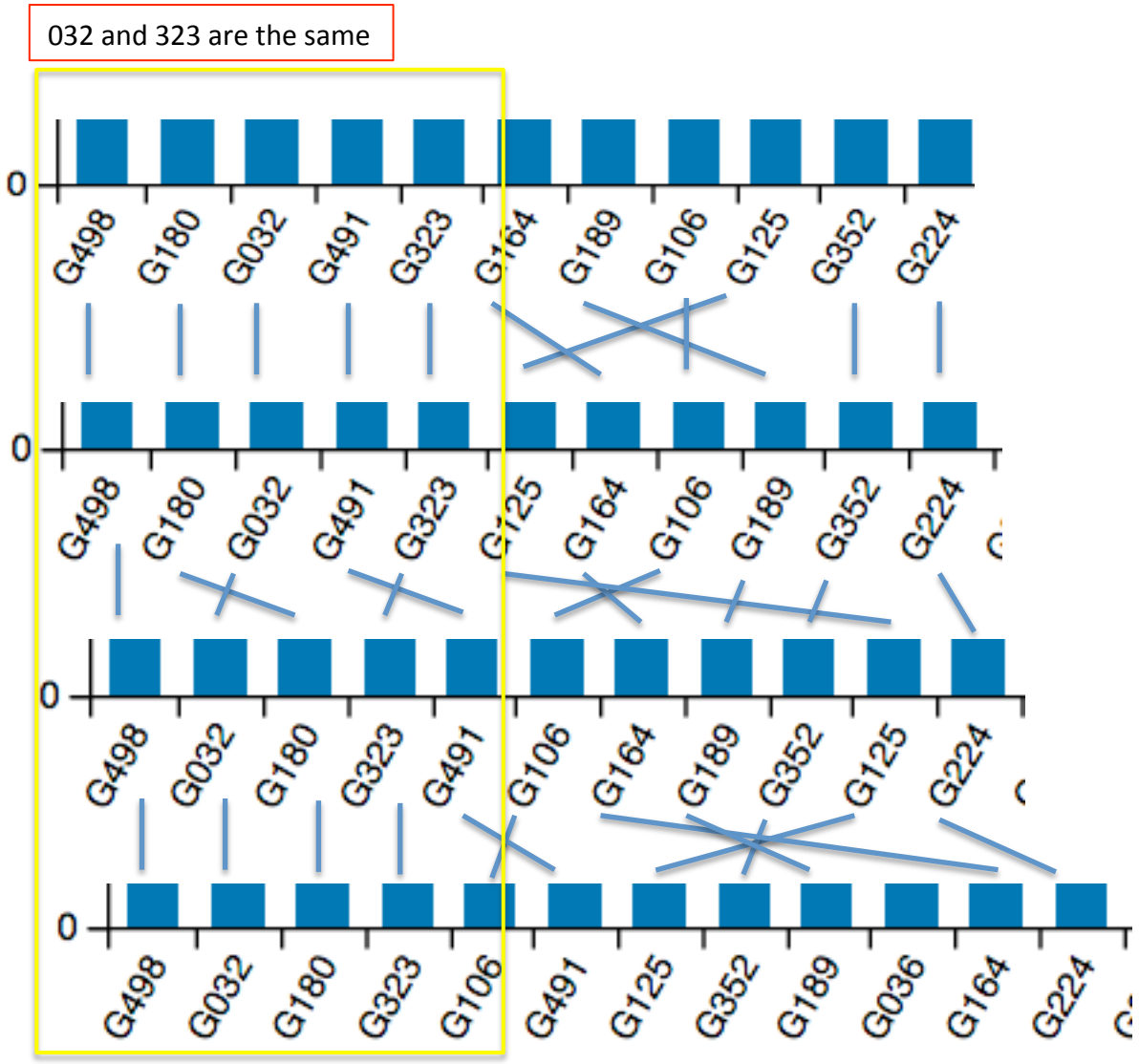
Long/medium contacts (FM only), sum Zscore (>0)

Long/medium: top10

Long/medium: L5

Long/medium: L2

Long/medium: L



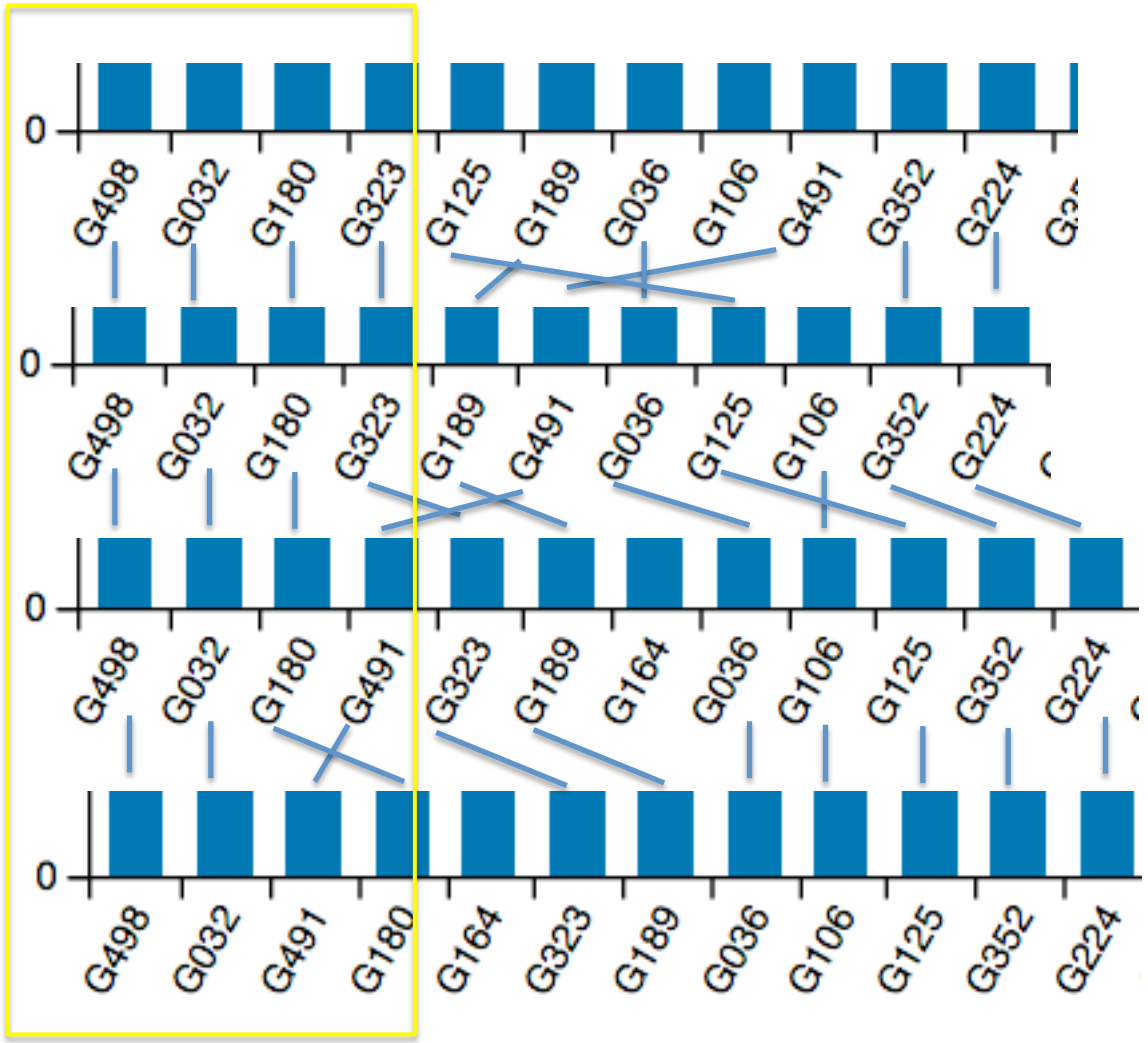
Long contacts, (FM only), sum Zscore (>0)

Long: L

Long: L/2

Long: L/5

Long: top10



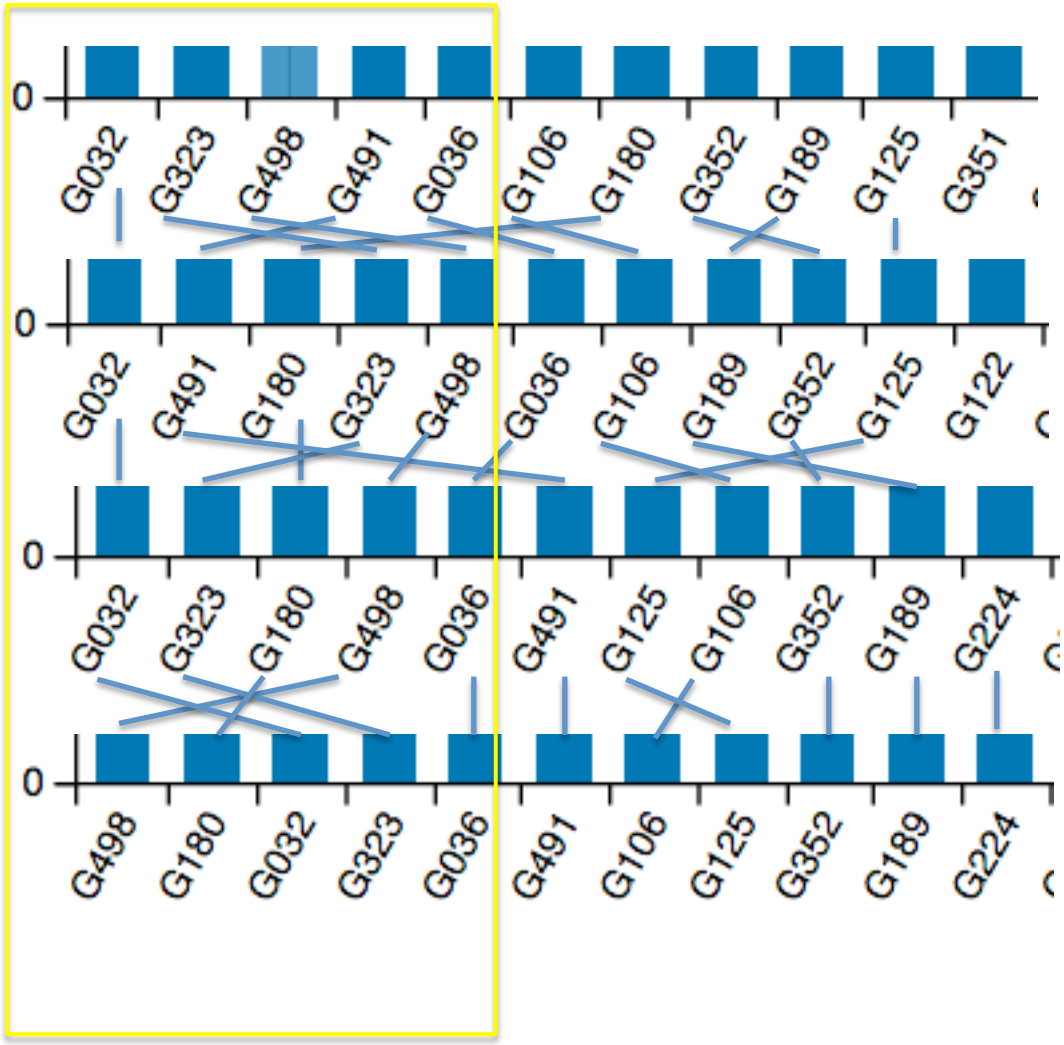
Extra long contacts only, (FM only), Zscore>0

Extra Long: top 10

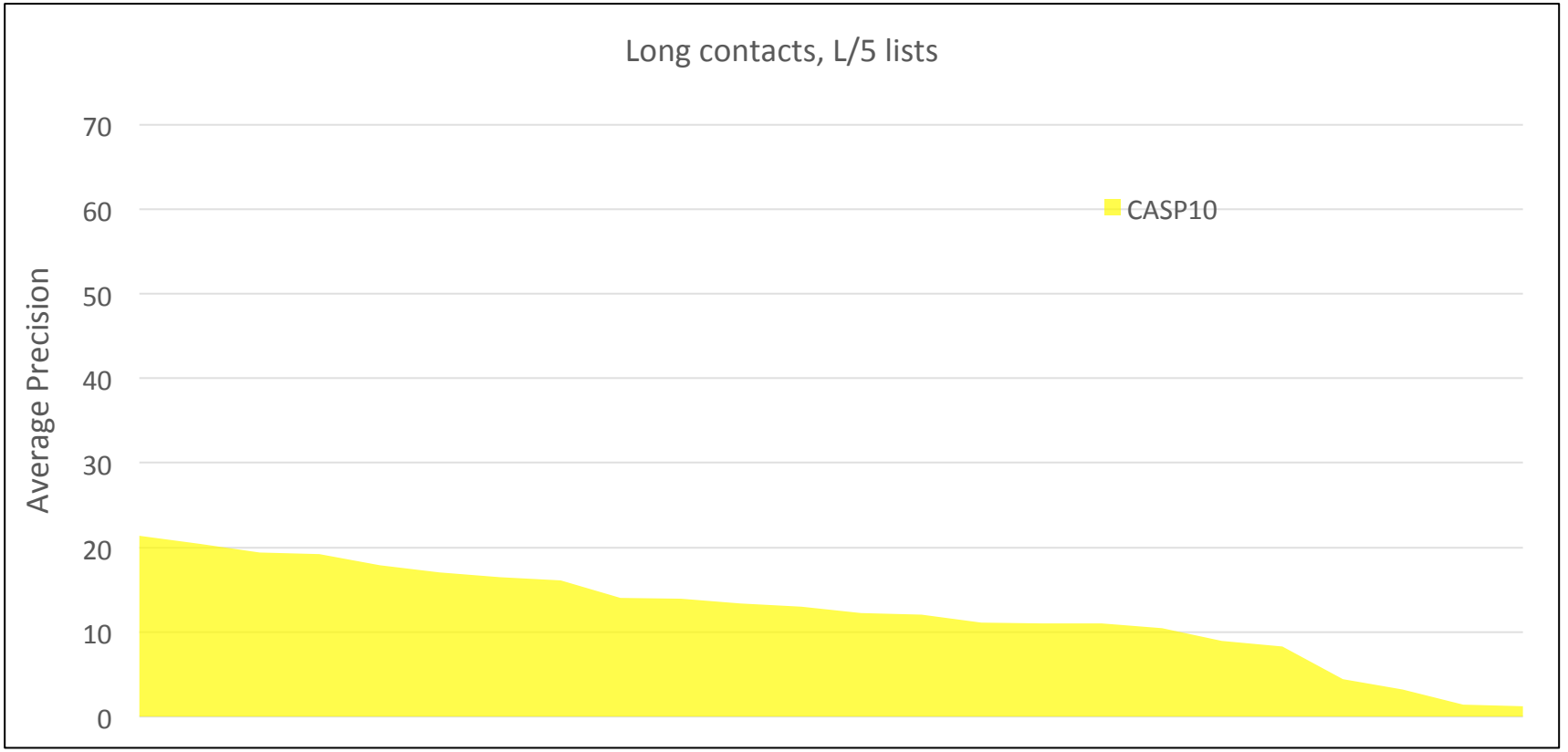
Extra Long: L/5

Extra Long: L/2

Extra Long: L



Improvement in contact prediction accuracy over CASP10-13 meetings

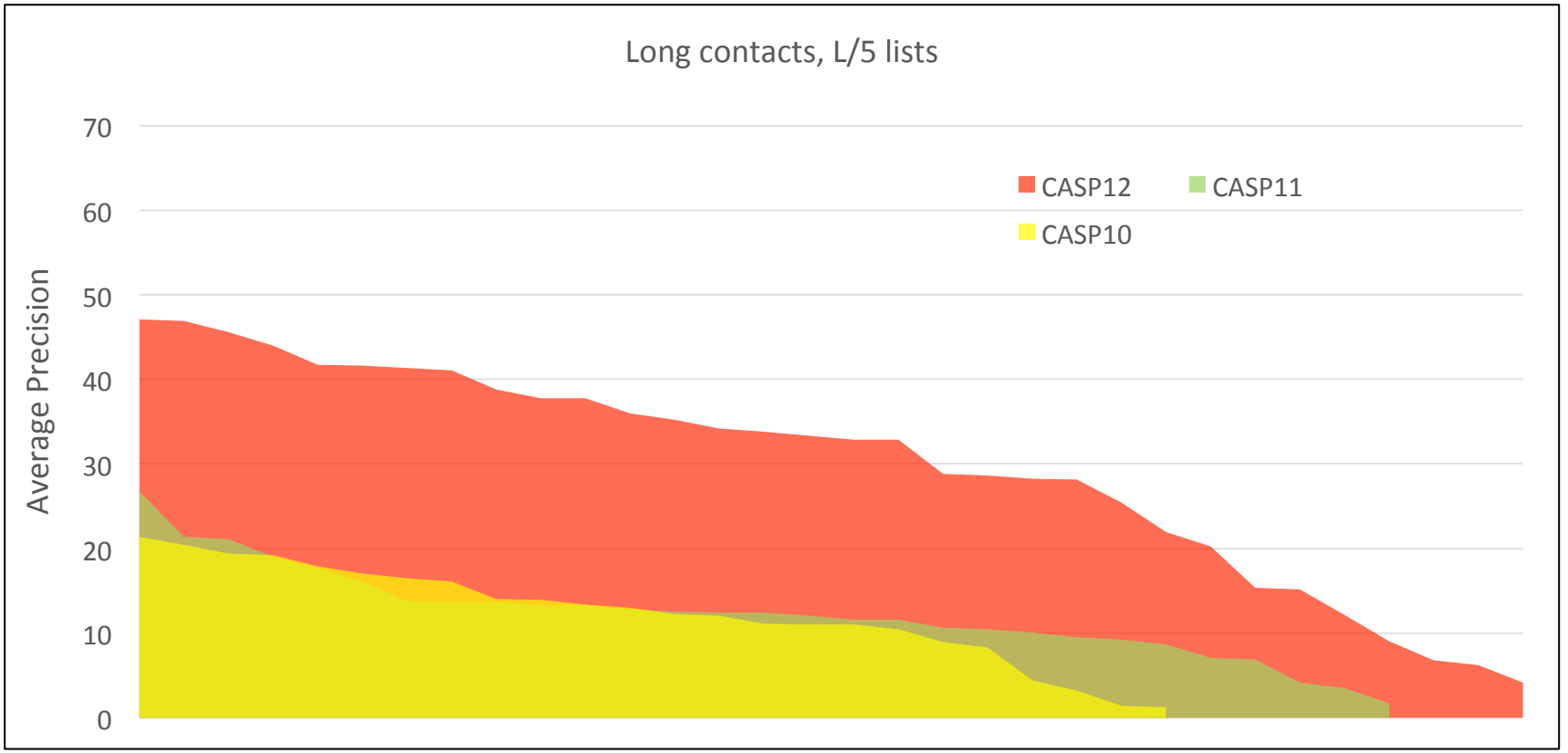


CASP10: 23 groups, 15 non-redundant

Improvement in contact prediction accuracy over CASP10-13 meetings

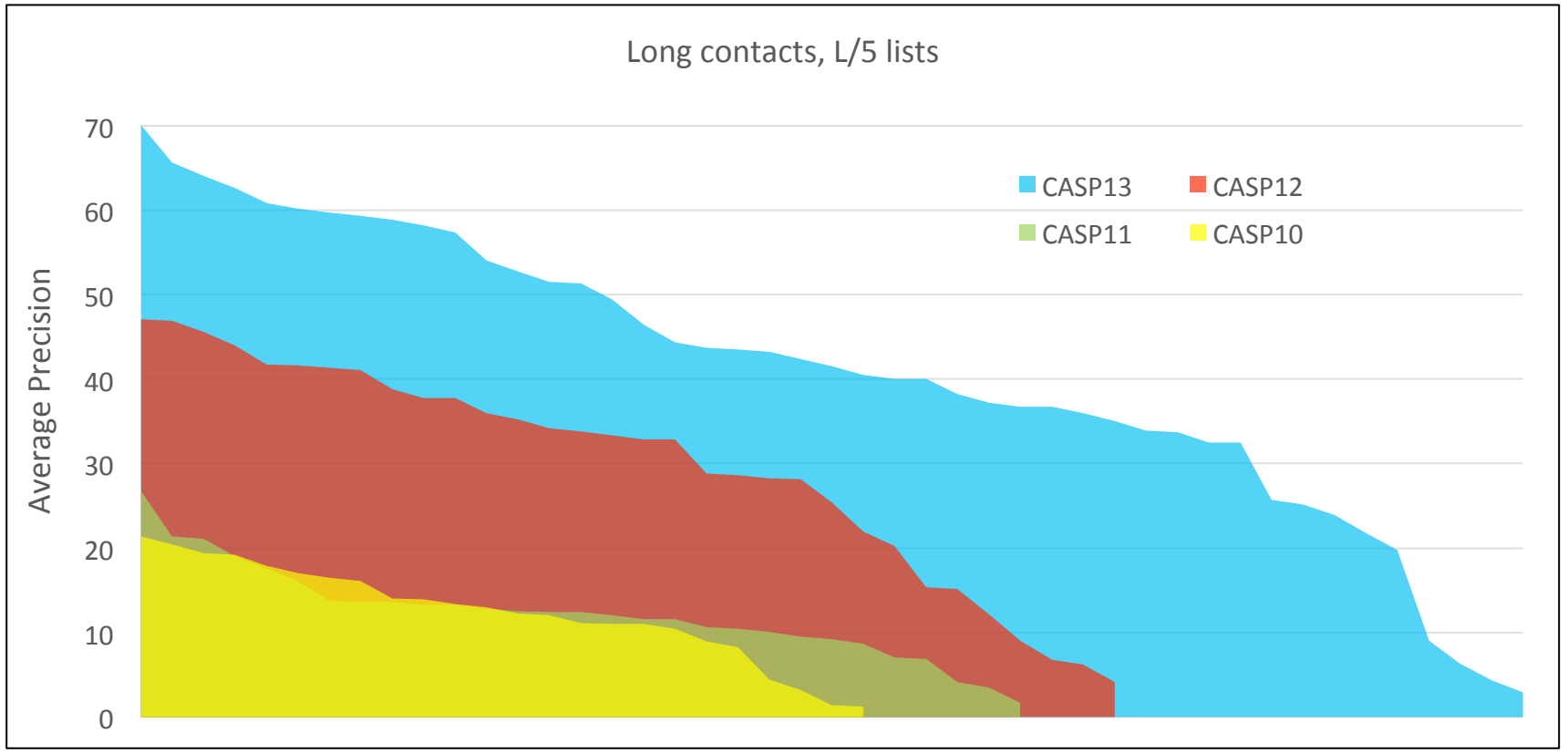
CASP10: 23 groups, 15 non-redundant
CASP11: 28 groups, 22 non-redundant

Improvement in contact prediction accuracy over CASP10-13 meetings



CASP10: 23 groups, 15 non-redundant
CASp11: 28 groups, 22 non-redundant
CASP12: 31 groups, 22 non-redundant

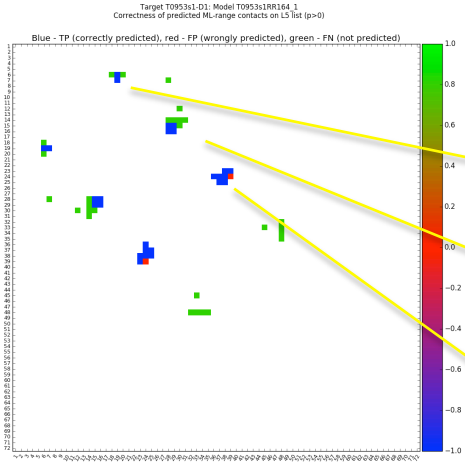
Improvement in contact prediction accuracy over CASP10-13 meetings



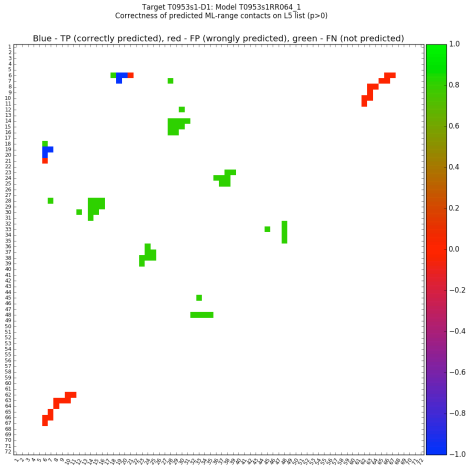
CASP10: 23 groups, 15 non-redundant
CASP11: 28 groups, 22 non-redundant
CASP12: 31 groups, 24 non-redundant
CASP13: 44 groups, 34 non-redundant

T0953s1d1

Good Fscore 63.4

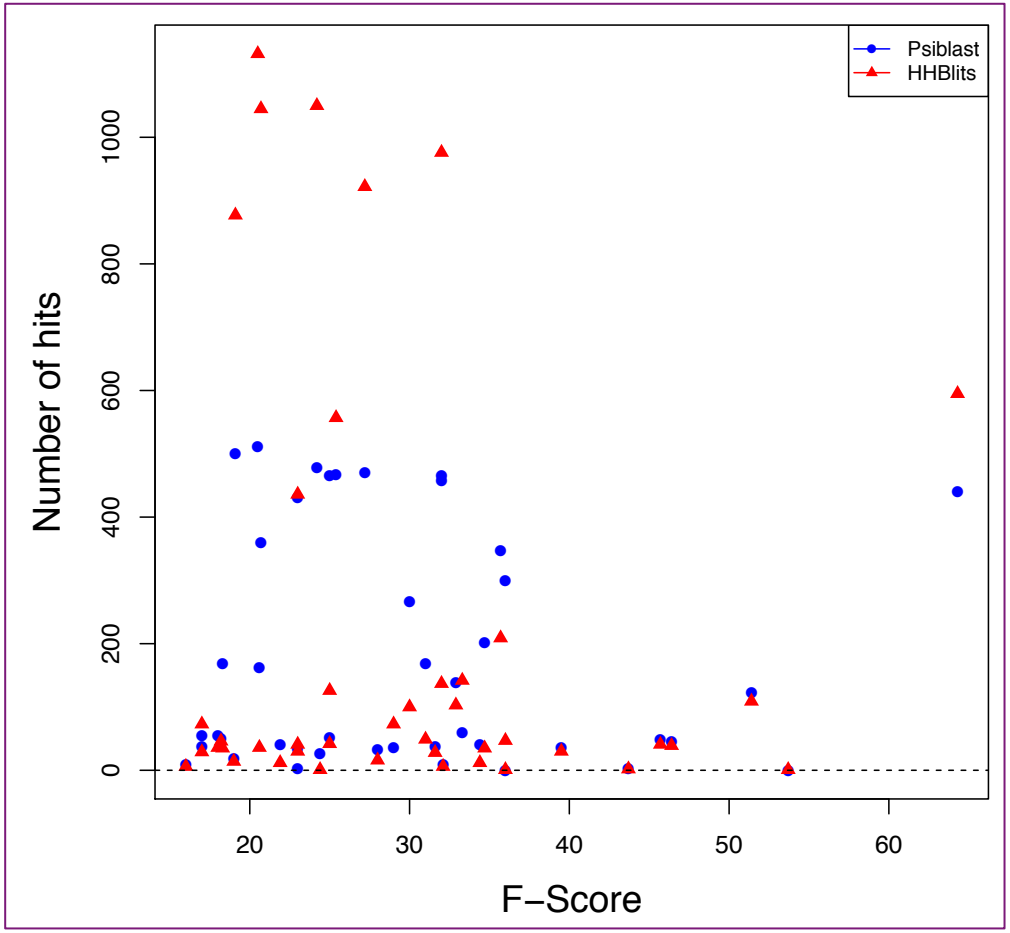


Poor Fscore: 14.64



Best TS model (G43), Cyan, GDT_TS 54.48
Contact model (G164), Green, GDT_TS 41.05

Relationship between *sequence profile depth* and success (F-score) of predicting contacts



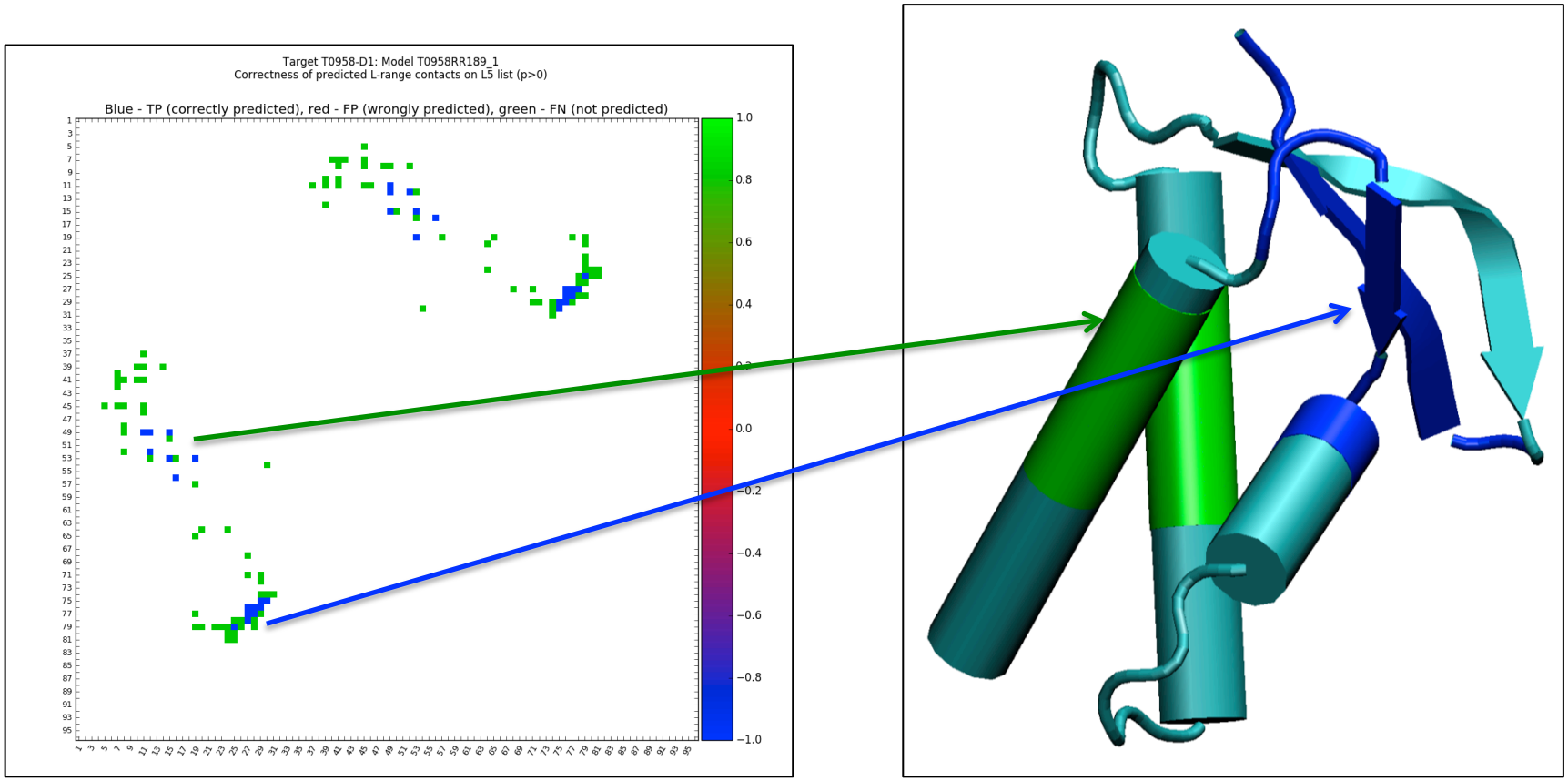
- Less reliant on sequence profiles.

Limited signal coming from sequence

	Blast		Blast+HHblits
Fscore	e-5	e-20	Neff
20.7	23	10	360
32.9	252	25	138
45.7	14	1	49
17	46		37
28	50	17	33
31.6	40	22	37
39.5	46	37	36
21.9	669	37	40
34.4	591		41
31	89	46	168
33.3	38	20	60
20.5	3021	1905	511
32	172	172	457
25.4	609	183	467
25	6130	129	465
34.7	132	31	201
34.7	91	111	201
16	30	6	9
17	30	6	55
35.7	278	17	347
23	19	9	35
20.6	38	23	162
18.3	38	23	169
36	194	1	300
24.4	194	1	27
46.4	58	31	45

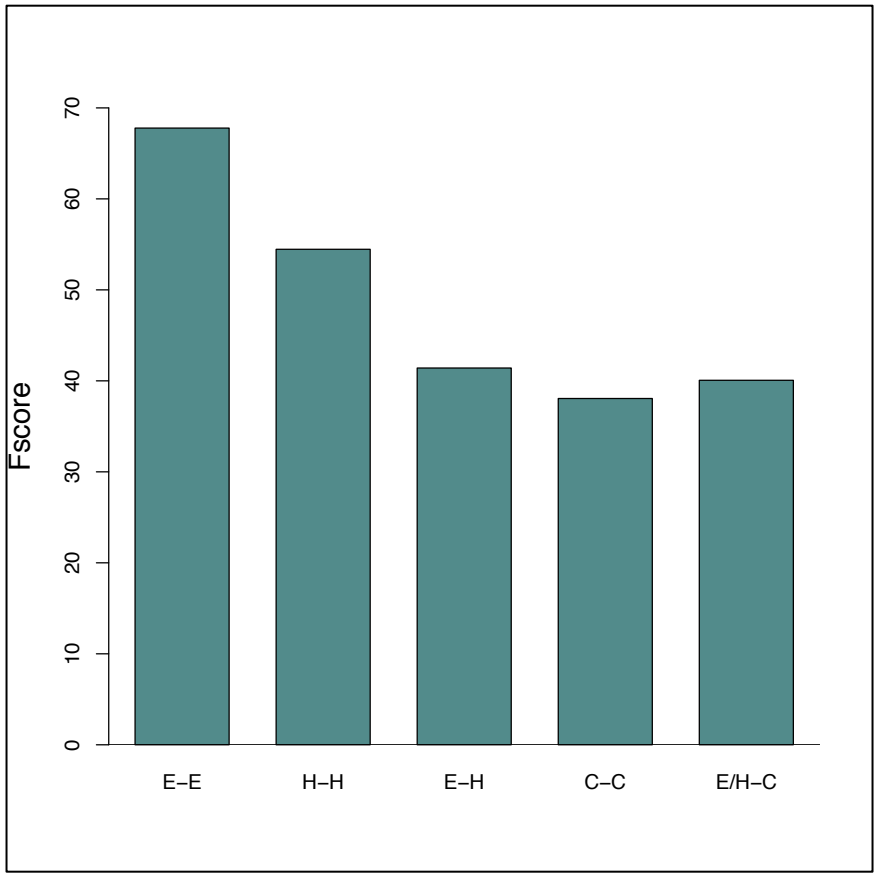
	Blast		Blast+HHblits
Fscore	e-5	e-20	Neff
18	58	31	54
25	58	31	51
53.7	1	1	0
29	14	14	36
24.2	1266	1028	478
43.7	1	1	2
19.1	3752		500
32.1	4	3	9
23	584	110	430
36	1	1	0
19	7	4	18
51.4	21	6	123
51.4	77	13	123
30	231	126	267
64.3	302	85	441
64.3	545	343	441
27.2	1730	68	470
18.2	629	1163	50
18.2	3755	38	50
32	1380	53	465

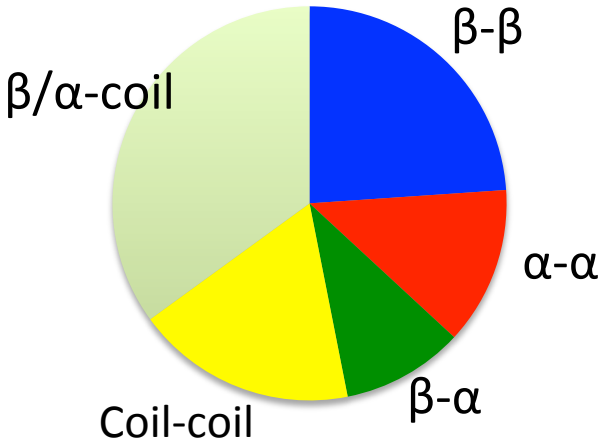
What is what?

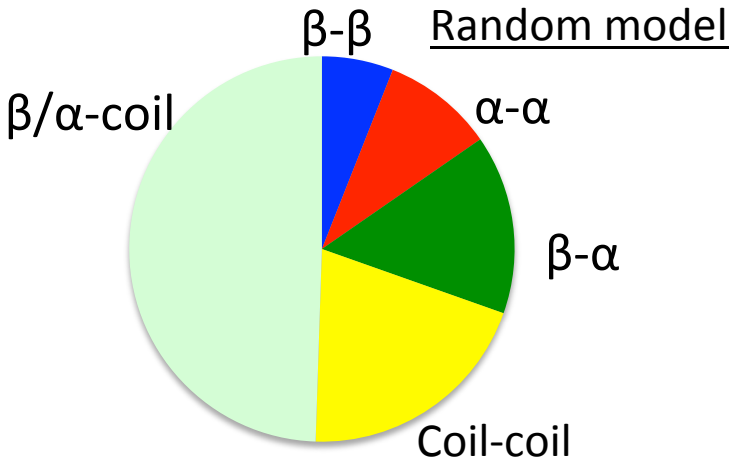


Green: parallel (parallel with diagonal) + diffuse (helical)
Blue: Anti-parallel (orthogonal with diagonal) + compact (strand)

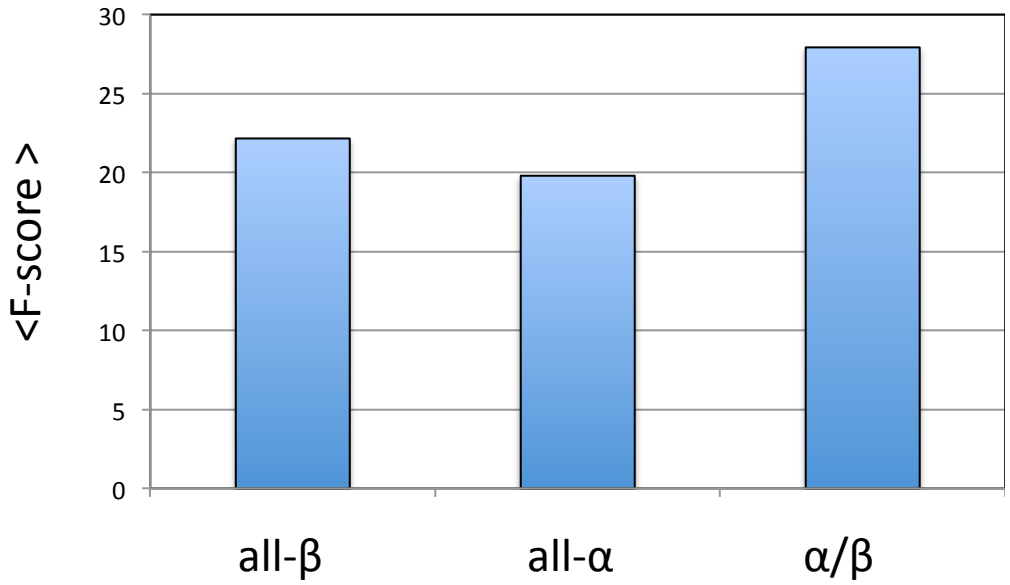
Performance vs. secondary structure interactions



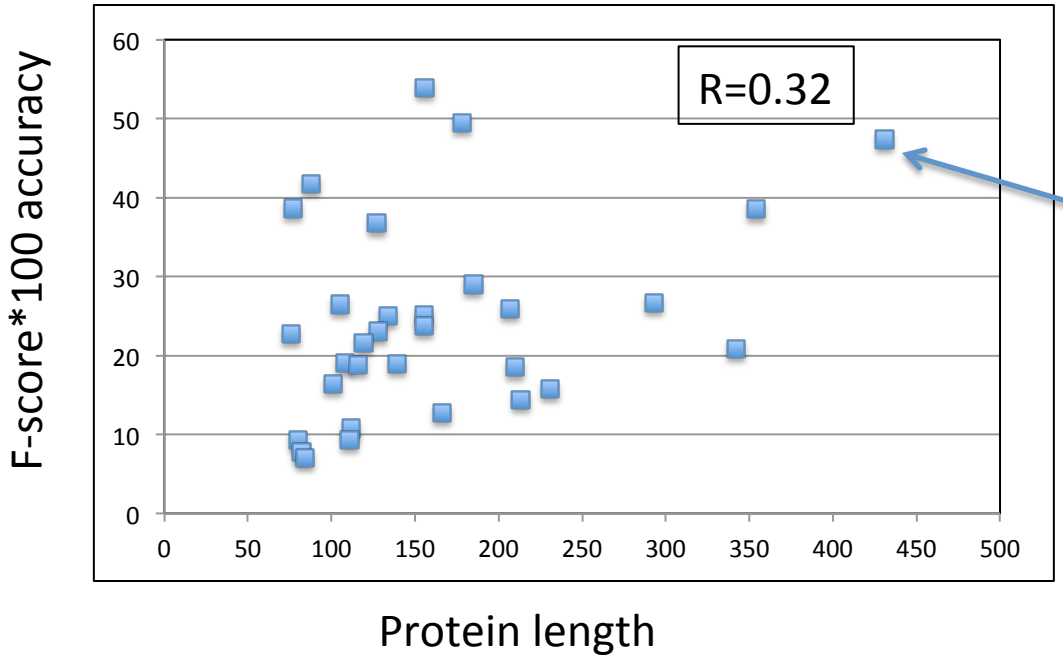




Topology dependence of success rates, Class level



Correlation with size



Without this single point:
R=0.19

Conclusions

- Contact prediction methods made a major advance for the last two years
- A lot of different subsets of correct contacts can be made and used successfully in 3D modeling
- It is difficult to directly correlate predicted contacts with 3D predictions because of ambiquity and lack of overlap between categories but :
  - Best 3D predictors have either even superior contact predictions or better ways to use contact information
  - From the few examples when both contacts and 3D structures were predicted we see strong inconsistencies: it is important to know how to use contact information
- Often very few homologous sequences were available, but very good contact predictions were made
  - Less emphasis on co-variance based methods (supported by the abstract of invited groups)

Acknowledgement

CASP and Predictioncenter at UC Davies, Davies, USA:

Andriy Kryshtafovych
Bohdan Monastyrskyy
Krzysztof Fidelis

CASP organizers

Albert Einstein College of Medicine, New York, USA:

Rojan Shrestha
Eduardo Fajardo
Nelson Gil