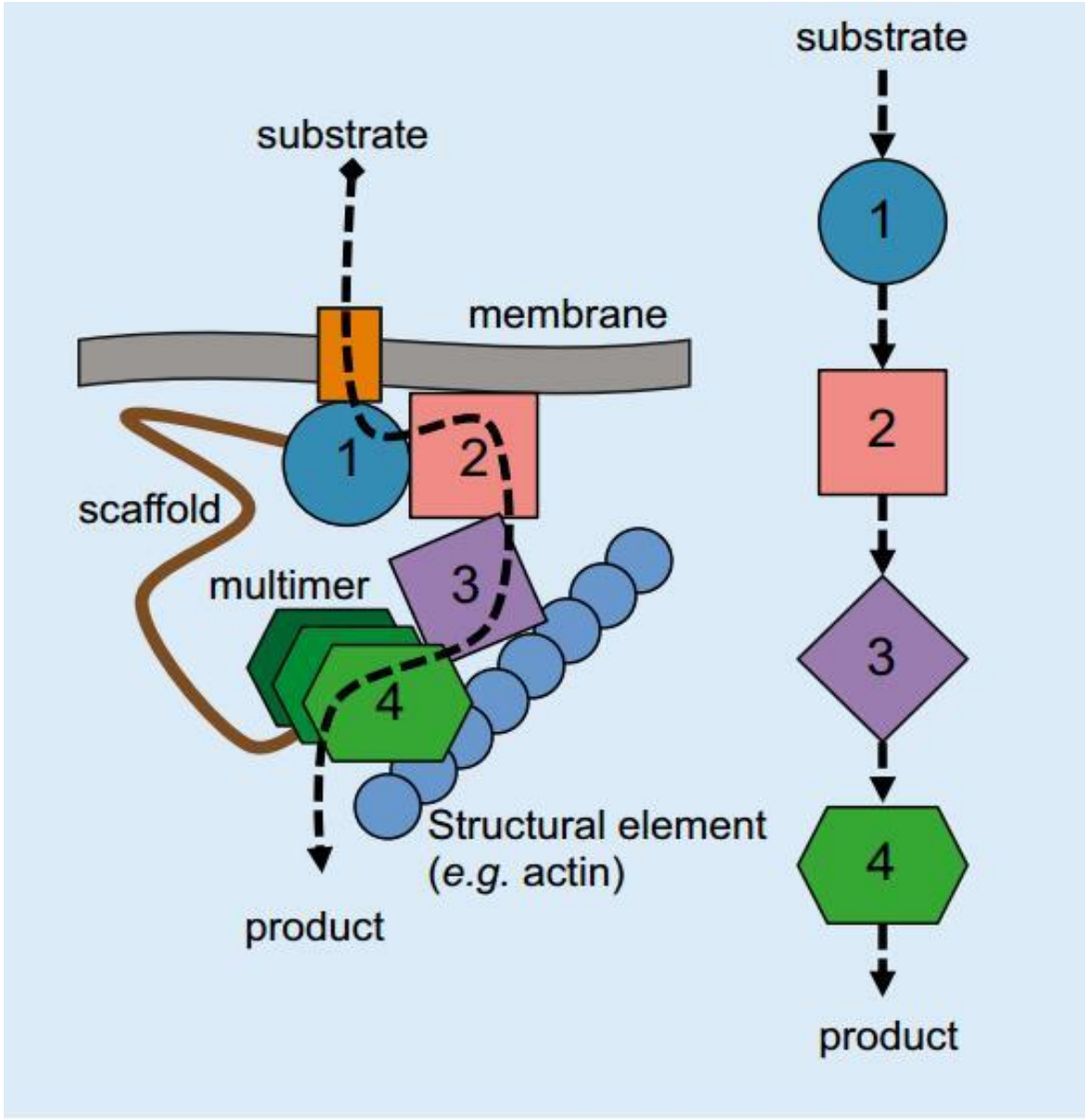
Capturing protein communities by structural proteomics in a thermophilic eukaryote

利用结构蛋白组学捕获嗜热真核生物中的蛋白质群落

汇报人：陈刚

蛋白质群落：

Protein communities are higher-order, often dynamically associated,assemblies of multiple macromolecular complexes that benefit from their close proximity to each other in the cell.



研究对象：Chaetomium thermophilum（嗜热毛壳菌）

1. protein interactions in thermophiles have higher stability
2. protein communities may be more robust than those from other model systems

研究方法：quantitative mass spectrometry ,electron microscopy(EM) , computational modeling

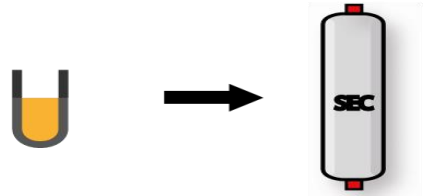
Method

1.we obtained simple and crude cellular fractions (simplified cell lysates) from the thermophilic fungus C.thermophilum by singlestep analytical size exclusion chromatography(SEC，体积排阻色谱)

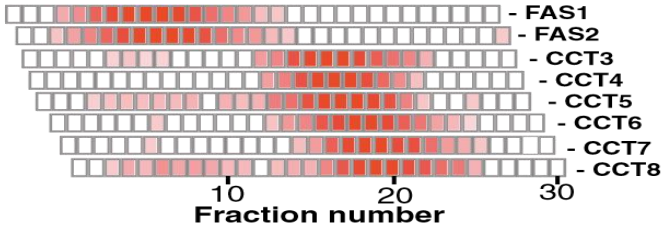
2.analyzed these fractions in biological triplicate by label-free quantitative liquid chromatography–mass spectrometry (LC-MS/MS) to characterize co-eluting proteins, complexes, and communities.

Result Separation of native higher-order assemblies from a eukaryotic thermophile

Chaetomium thermophilum

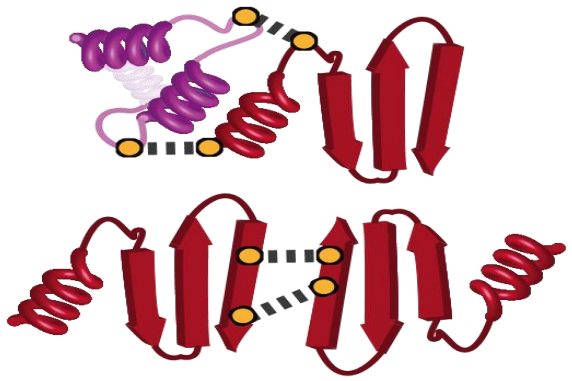


lyse cells, enrich for large complexes, and seperate by SEC



Interface profiling

Cross-linking MS and integrative docking

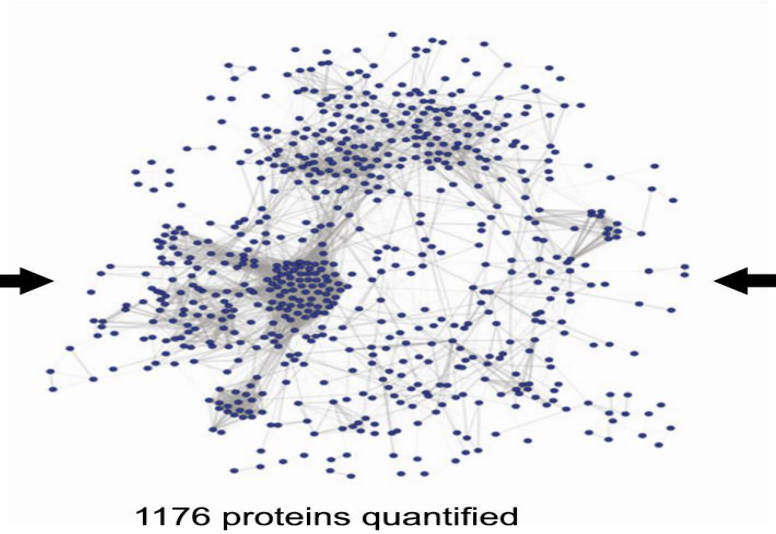


3139 high-quality crosslinks
239 interfaces
69 novel homomultimers
66 novel heteromultimers

Validation

Molecular Profiling

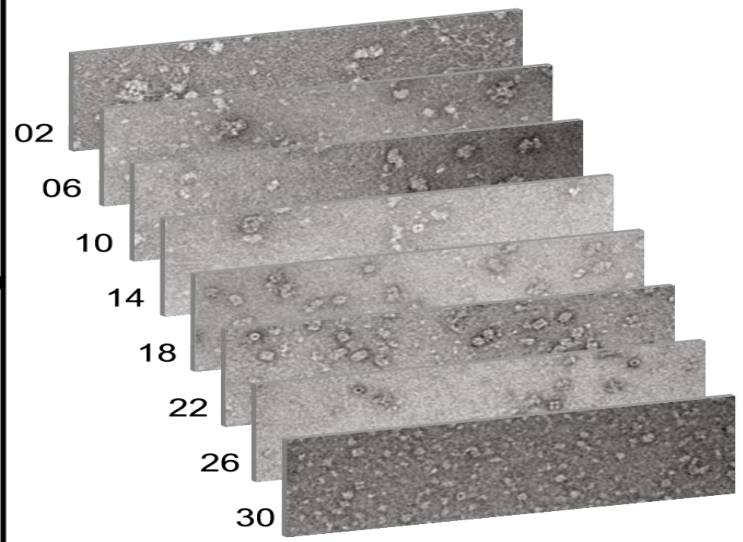
LC-MS/MS and protein interaction prediction



1176 proteins quantified
6313 interactions
27 communities
108 interconnected complexes

Structural profiling

Electron Microscopy and image processing of each SEC fraction



- Structural signatures per fraction
- Shape detection of protein complexes in cell extracts

Validation

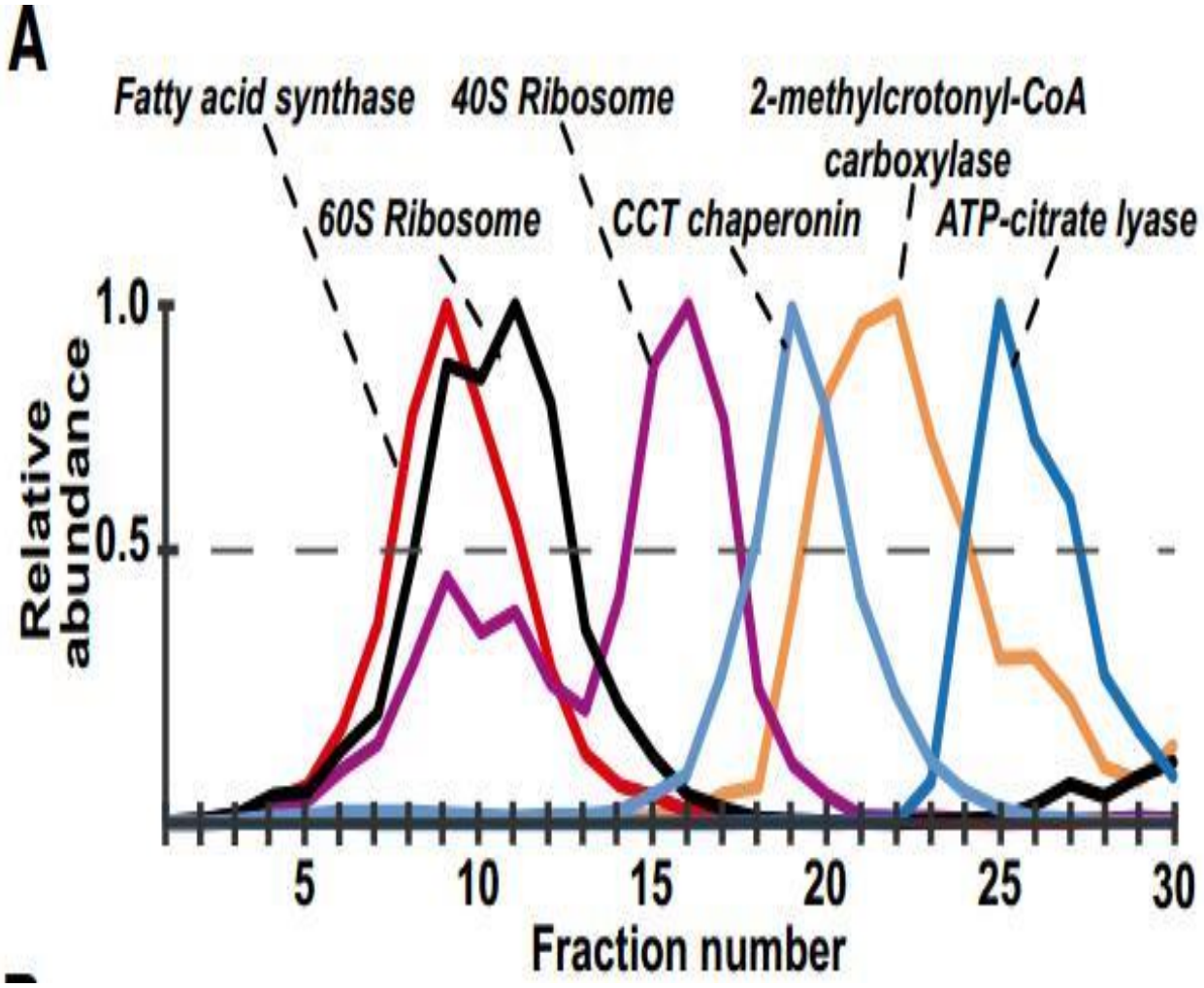
Structural insights into native protein communities by integrative molecular biology and cryoEM

Method

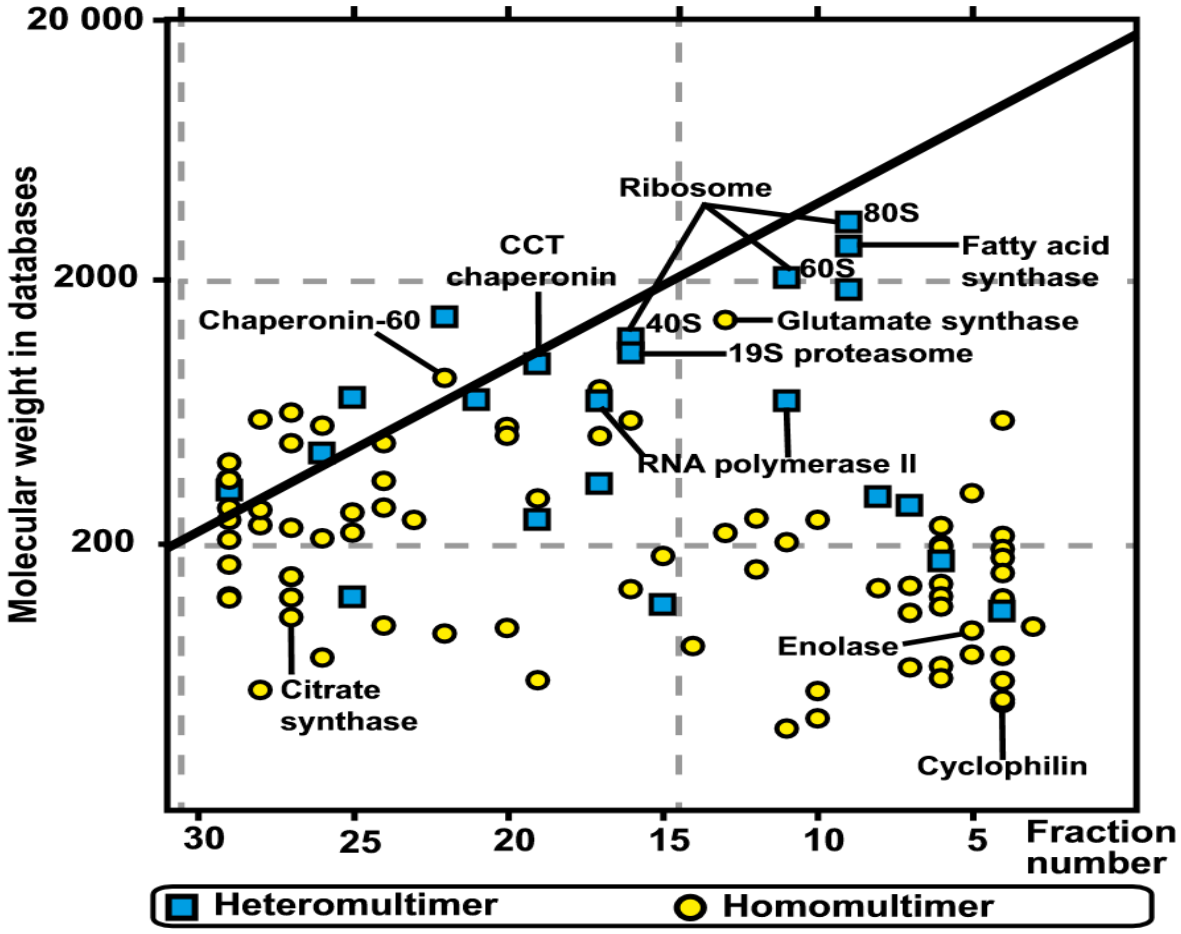
1.通过基于iBAQ评分量化蛋白质丰度来确定每种蛋白质的实验洗脱曲线

2.为了进一步评估生化分离的质量和有效性，我们确定观察到的洗脱图谱是否与蛋白质数据库中所包含的良好表征和保守的蛋白质复合物的组成，分子量和化学计量匹配.

Result



Experimentally measured molecular weight

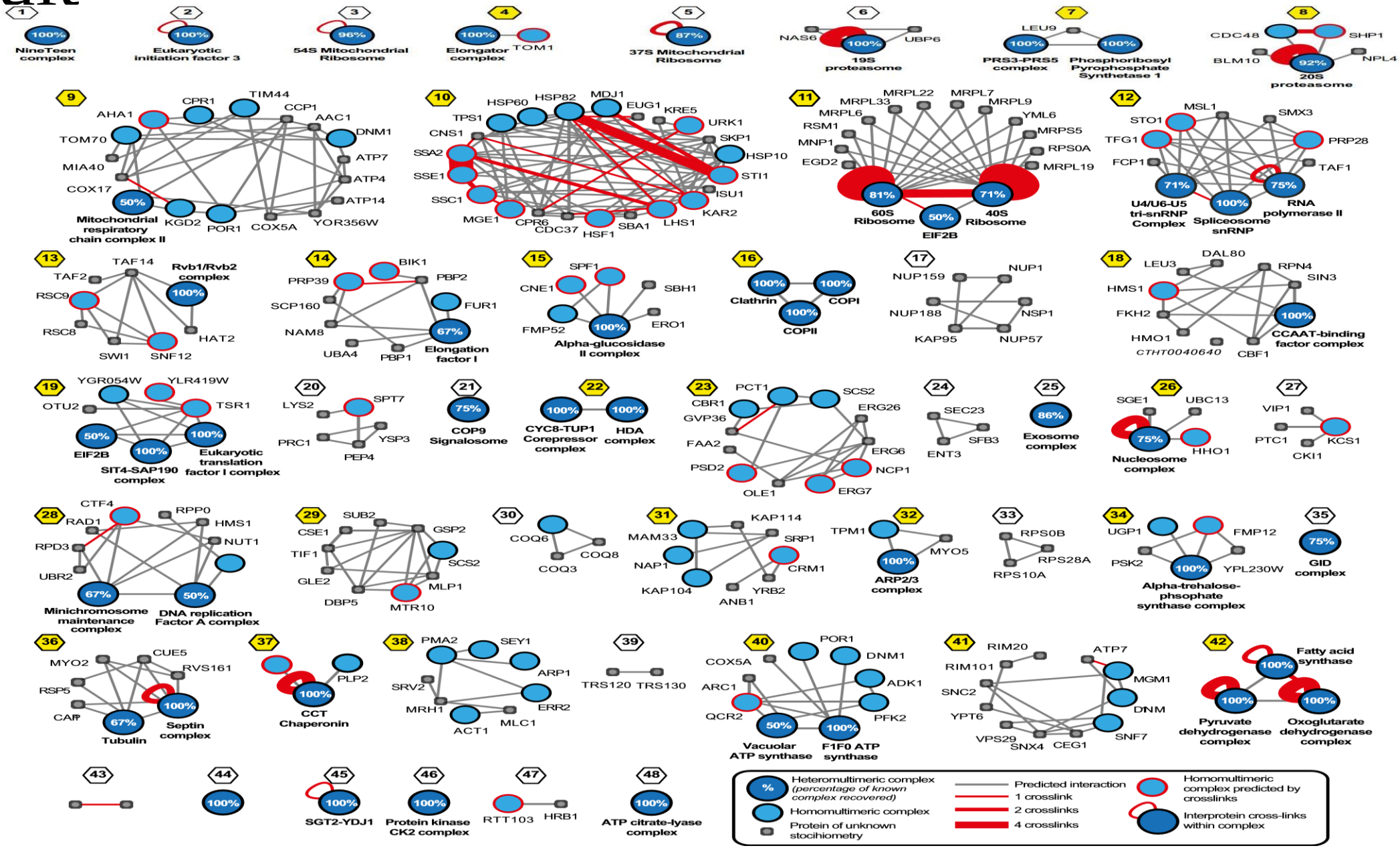


Method

1.接下来将蛋白质洗脱曲线与已知的功能性关联，一起用于系统地定义蛋白质群落。 洗脱曲线之间的相关性可以指示在时空背景下执行功能的相同复合物或蛋白质群落的成员。 对于数据集中所有可能的蛋白质对，我们计算了Pearson相关系数（互相关共洗脱（CCC）评分），以测量其洗脱曲线的相似性。

2.采取了0.85的最小交互概率构造一个蛋白质 - 蛋白质相互作用网络，从这个网络中，我们使用了一种聚类方法，有效地发现密集连接的重叠区域（代表蛋白质复合物和群落）。

Result


1
100%
NineTeen complex
2
100%
Eukaryotic initiation factor 3
3
96%
54S Mitochondrial Ribosome
4
100%
Elongator complex
TOM1
5
87%
37S Mitochondrial Ribosome
6
NAS6
100%
UBP6
19S proteasome
7
LEU9
100%
100%
PRS3-PRS5 complex
Phosphoribosyl Pyrophosphate Synthetase 1
8
CDC48
SHP1
BLM10
92%
NPL4
20S proteasome
9
AHA1
CPR1
TIM44
CCP1
AAC1
DNM1
TOM70
ATP7
MIA40
ATP4
COX17
ATP14
50%
KGD2
POR1
COX5A
YOR356W
Mitochondrial respiratory chain complex II
10
HSP60
HSP82
MDJ1
EUG1
KRE5
TPS1
URK1
CNS1
SKP1
SSA2
HSP10
SSE1
STI1
SSC1
ISU1
KAR2
MGE1
CPR6
CDC37
HSF1
SBA1
LHS1
11
MRPL22
MRPL7
MRPL33
MRPL9
MRPL6
YML6
RSM1
MRPS5
MNP1
RPS0A
EGD2
MRPL19
81%
71%
60S Ribosome
50%
40S Ribosome
EIF2B
12
MSL1
STO1
SMX3
TFG1
PRP28
FCP1
TAF1
71%
75%
U4/U6-U5 tri-snRNP Complex
100%
Spliceosome snRNP
RNA polymerase II
13
TAF14
TAF2
Rvb1/Rvb2 complex
RSC9
100%
RSC8
HAT2
SWI1
SNF12
14
BIK1
PRP39
PBP2
SCP160
FUR1
NAM8
67%
UBA4
PBP1
Elongation factor I
15
SPF1
CNE1
SBH1
FMP52
100%
ERO1
Alpha-glucosidase II complex
16
100%
100%
Clathrin
COPI
100%
COPII
17
NUP159
NUP1
NUP188
NSP1
KAP95
NUP57
18
DAL80
LEU3
RPN4
HMS1
SIN3
FKH2
100%
HMO1
CTHT0040640
CBF1
CCAAT-binding factor complex
19
YGR054W
YLR419W
OTU2
TSR1
50%
100%
100%
EIF2B
SIT4-SAP190 complex
Eukaryotic translation factor I complex
20
LYS2
SPT7
PRC1
YSP3
PEP4
21
75%
COP9 Signalosome
22
100%
100%
CYC8-TUP1 Corepressor complex
HDA complex
23
PCT1
CBR1
SCS2
GVP36
ERG26
FAA2
ERG6
PSD2
NCP1
OLE1
ERG7
24
SEC23
ENT3
SFB3
25
86%
Exosome complex
26
SGE1
UBC13
75%
HHO1
Nucleosome complex
27
VIP1
PTC1
KCS1
CKI1
28
CTF4
RAD1
RPP0
HMS1
NUT1
RPD3
UBR2
67%
50%
Minichromosome maintenance complex
DNA replication Factor A complex
29
SUB2
CSE1
GSP2
TIF1
SCS2
GLE2
MLP1
DBP5
MTR10
30
COQ6
COQ8
COQ3
31
KAP114
MAM33
SRP1
NAP1
CRM1
KAP104
YRB2
ANB1
32
TPM1
MYO5
100%
ARP2/3 complex
33
RPS0B
RPS28A
RPS10A
34
UGP1
FMP12
PSK2
100%
YPL230W
Alpha-trehalose-phosphate synthase complex
35
75%
GID complex
36
MYO2
CUE5
RVS161
RSP5
CAP
100%
67%
Tubulin
Septin complex
37
PLP2
100%
CCT Chaperonin
38
PMA2
SEY1
ARP1
SRV2
ERR2
MRH1
MLC1
ACT1
39
TRS120
TRS130
40
COX5A
POR1
DNM1
ARC1
ADK1
QCR2
PFK2
50%
100%
Vacuolar ATP synthase
F1F0 ATP synthase
41
RIM20
ATP7
RIM101
MGM1
SNC2
DNM
YPT6
VPS29
SNX4
CEG1
SNF7
42
Fatty acid synthase
100%
100%
100%
Pyruvate dehydrogenase complex
Oxoglutarate dehydrogenase complex
43
44
100%
45
100%
SGT2-YDJ1
46
100%
Protein kinase CK2 complex
47
RTT103
HRB1
48
100%
ATP citrate-lyase complex
%
Heteromultimeric complex (percentage of known complex recovered)
Homomultimeric complex
Protein of unknown stoichiometry
Predicted interaction
1 crosslink
2 crosslinks
4 crosslinks
Homomultimeric complex predicted by crosslinks
Interprotein cross-links within complex

Method

1.由共现推断的物理相互作用也可以是间接的，所以接下来我们通过将的全蛋白组XL-MS(交联质谱）应用来表征预测的蛋白质群落中的成员之间发生的相互作用界面。为了捕获大部分相互作用组，我们整合了三个独立的XL-MS(交联质谱)数据集，这些数据集使用不同的互补方案获得，例如使用不同的化学交联剂，以及基于序列和基于结构不同的错误发现率。

Table 1. Cross-linking statistics at a false discovery rate of 10%.

FDR 10%	Cross-links	Structurally mapped	Total interfaces covered	Novel interfaces
Total cross-links	3,139	931	239	135
Cross-links on monomers	2,732	851	–	–
Cross-links on homomultimers	230	36[a]	121	69
Cross-links on heteromultimers	177	44	118	66

[a]These cross-links show decrease in intra-residue distance when measured on known homomultimers by 26.3 ± 13.4 Å.

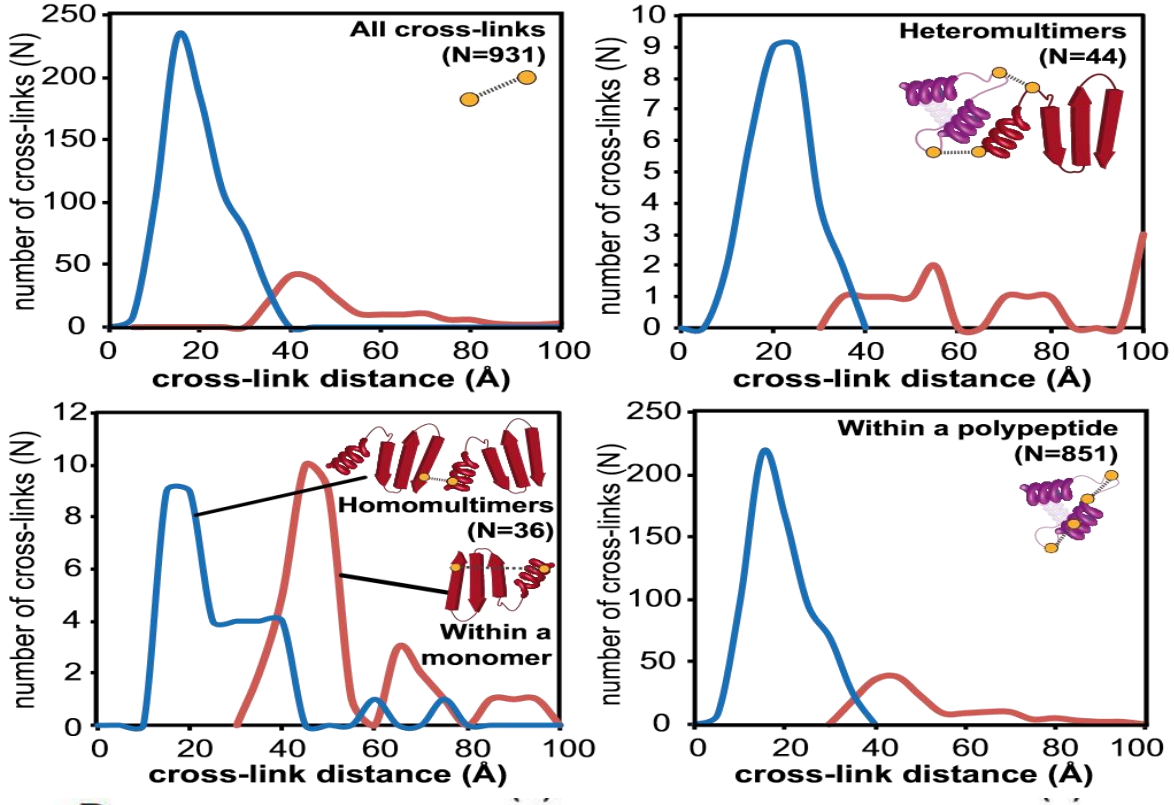
Method

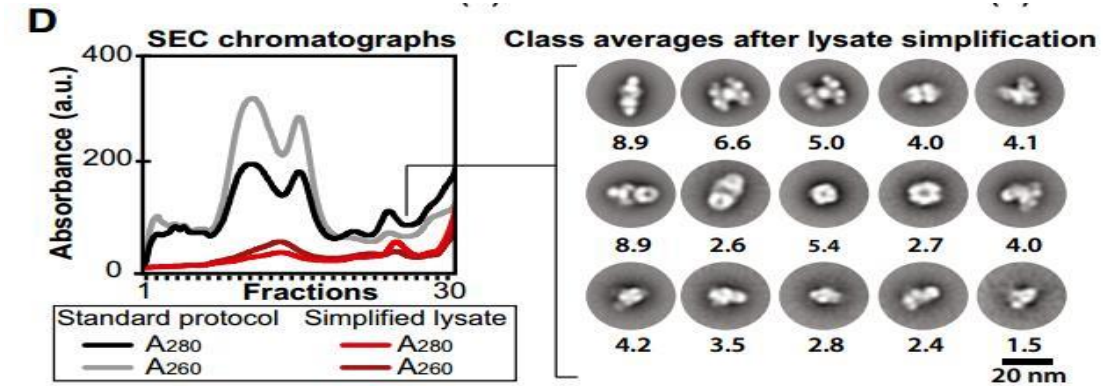
1.为了验证数据，检查了那些交联的肽识别在结构水平上是合格的，即对应于交联的赖氨酸残基的C原子之间的距离小于33埃(十分之一纳米);与所有结构已知的复合物的比较显示73％的分子间交联和84％的分子内交联合格。

2.使用EM检查了不同部分的结构特征,我们获得了大量负染色的所有级分的电子显微照片，确定了单个颗粒，并对其进行二维分类。

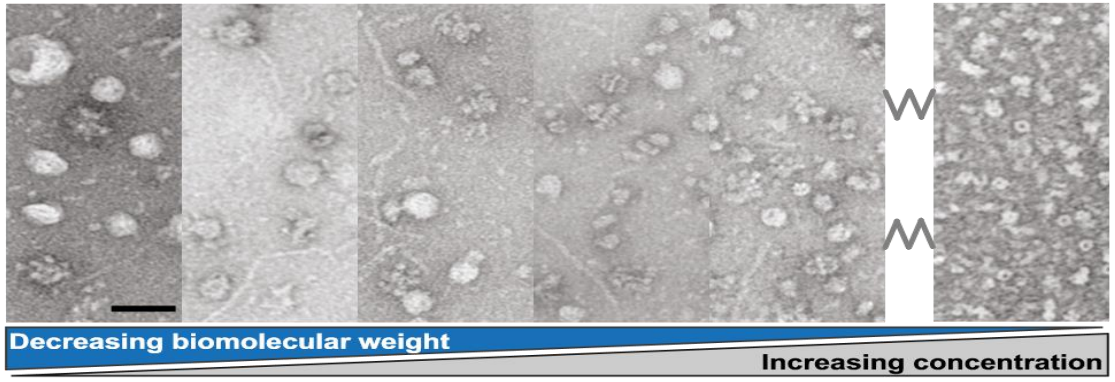
Result

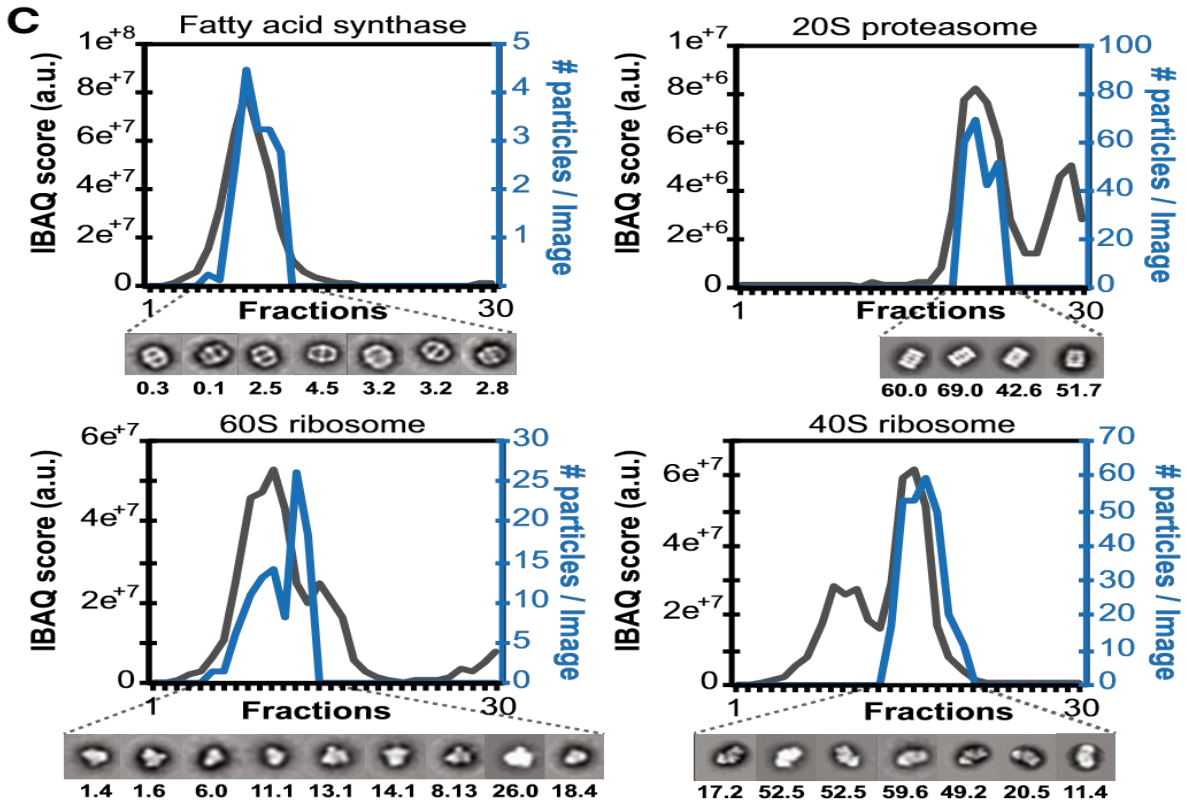
Structural mapping of cross-links





Fraction 3 Fraction 4 Fraction 5 Fraction 6 Fraction 7 Fraction 27



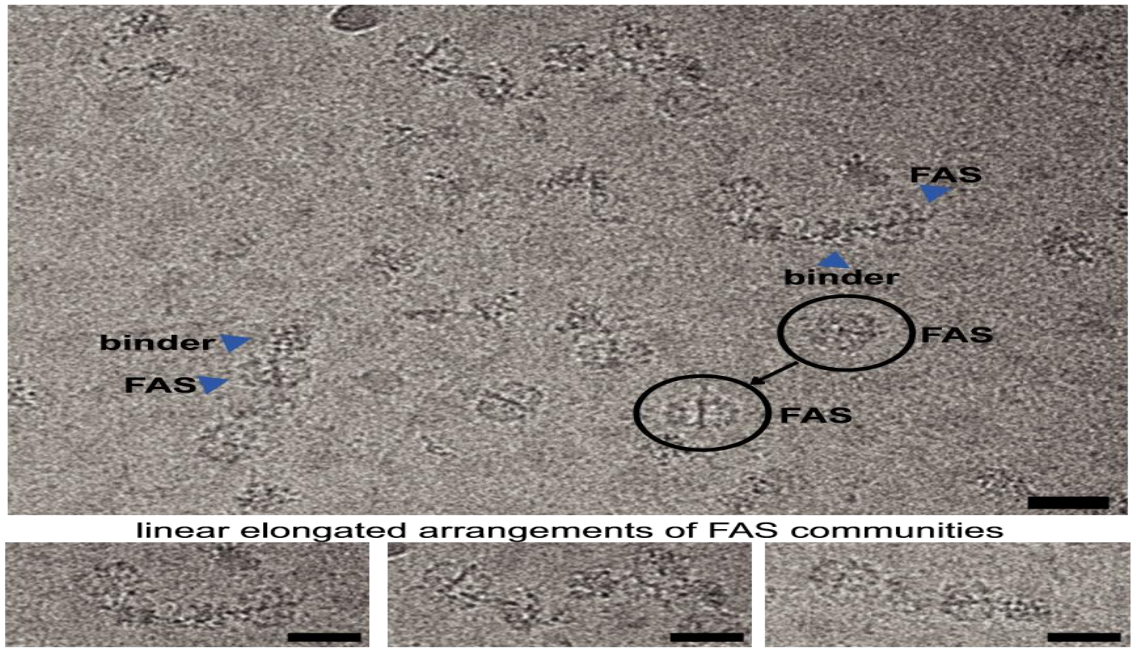


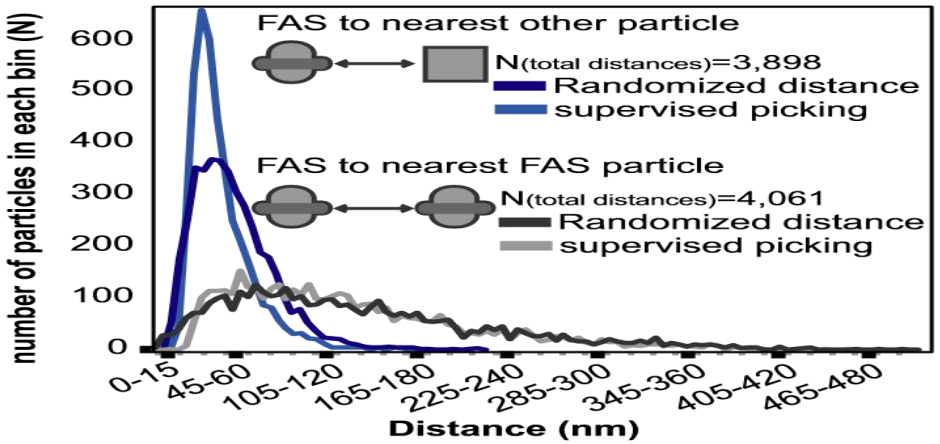
Method

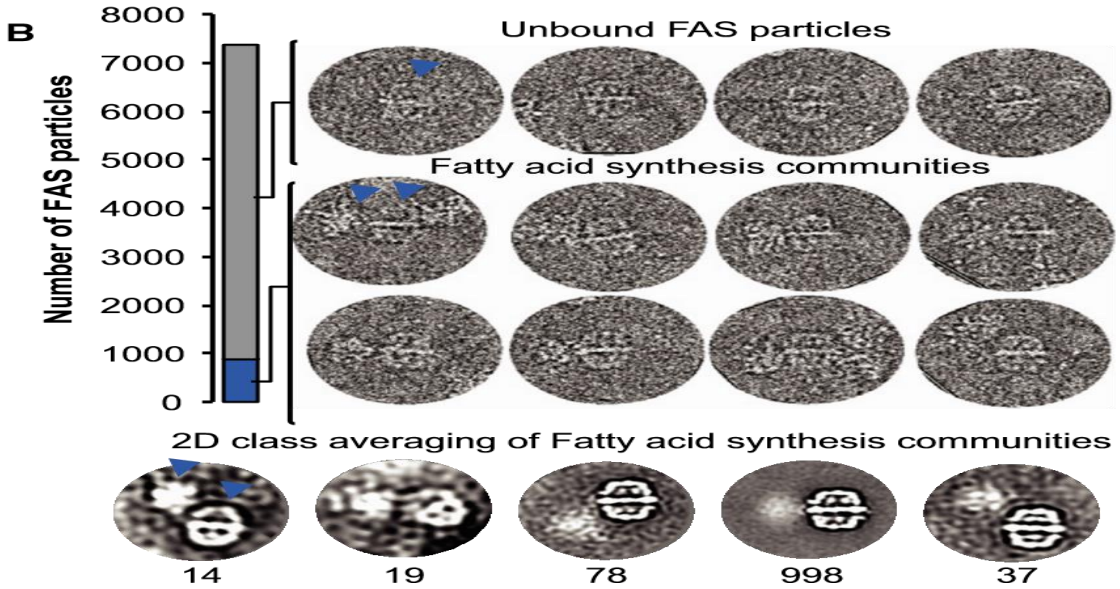
1.We next analyzed one of these structural signatures—fungal FAS—in more detail. In our analysis, FAS is a structurally prominent,2.6-MDa complex that contains six copies of all eight catalytic centers comprising the complete metabolic pathway for 16- and 18-carbon fatty acid production.

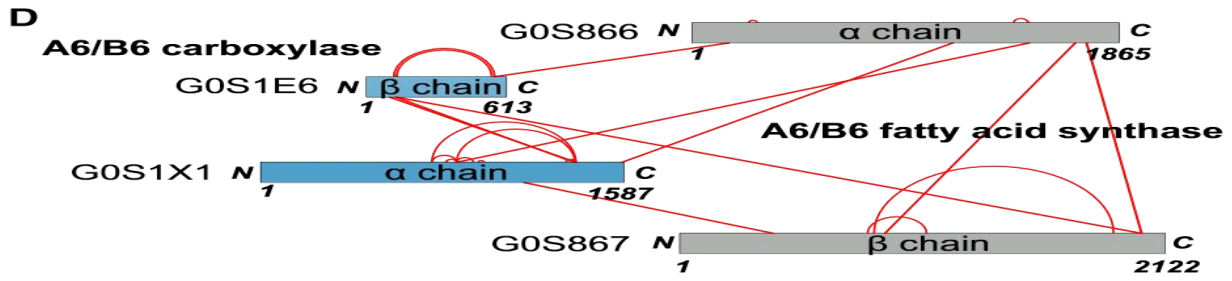
2.接下来我们着手测试这些粗提取物中是否可以进行高分辨率结构测定。X射线晶体学已经确定了独立的FAS的高分辨率结构。

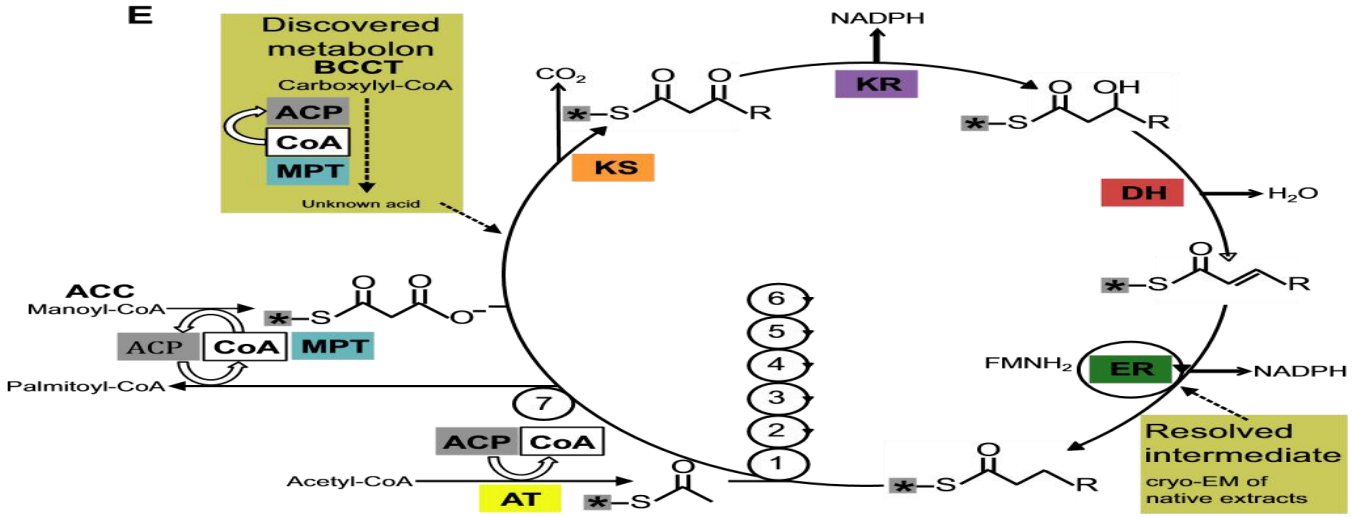
Result


FAS
binder
binder
FAS
FAS
linear elongated arrangements of FAS communities

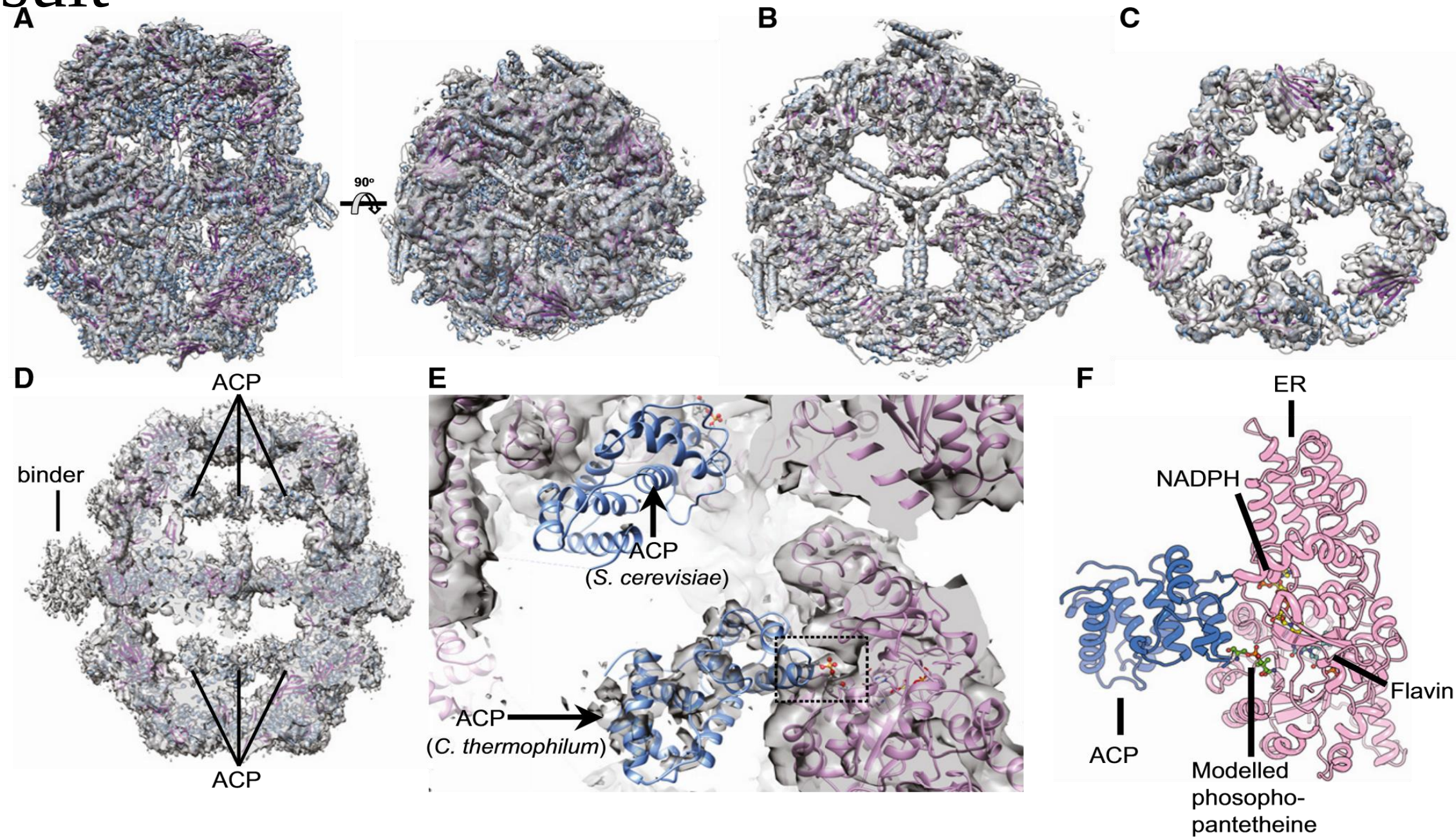

FAS to nearest other particle
N(total distances)=3,898
Randomized distance
supervised picking
FAS to nearest FAS particle
N(total distances)=4,061
Randomized distance
supervised picking
number of particles in each bin (N)
600
500
400
300
200
100
0
0-15
45-60
105-120
165-180
225-240
285-300
345-360
405-420
465-480
Distance (nm)


B
Number of FAS particles
8000
7000
6000
5000
4000
3000
2000
1000
0
Unbound FAS particles
Fatty acid synthesis communities
2D class averaging of Fatty acid synthesis communities
14
19
78
998
37


D
A6/B6 carboxylase
G0S866 N α chain C
1
1865
G0S1E6 N β chain C
1
613
A6/B6 fatty acid synthase
G0S1X1 N α chain C
1
1587
G0S867 N β chain C
1
2122


E
Discovered metabolon
BCCT
Carboxylyl-CoA
ACP
CoA
MPT
Unknown acid
ACC
Manoyl-CoA
ACP
CoA
MPT
Palmitoyl-CoA
ACP
CoA
Acetyl-CoA
AT
KS
KR
NADPH
DH
H2O
ER
FMNH2
NADPH
Resolved intermediate
cryo-EM of native extracts

Result


A
90°
B
C
D
ACP
binder
ACP
E
ACP
(S. cerevisiae)
ACP
(C. thermophilum)
F
ER
NADPH
Flavin
ACP
Modelled
phosopho-
pantetheine

conclusion：

1、确定了27个不同的蛋白质群落，包括108个相互联系的蛋白质复合体。

2、通过cryoEM研究了这些提取物中脂肪酸合成酶的结构，并揭示了这些结构酶的多种灵活状态以适应于与其他复合物的结合。

view:

1.方法框架补充了新兴的单细胞结构生物学方法，提供亚细胞特征的高分辨率快照，但目前无法确定潜在的生物分子实体。