AbSciCon
2022

519: From Amino Acids to the First Proteins II

1
00:00:03,510 --> 00:00:02,310
thank you for staying with us

2
00:00:06,710 --> 00:00:03,520
we're going to be talking about the

3
00:00:07,990 --> 00:00:06,720
evolution of early proteins from amino

4
00:00:09,910 --> 00:00:08,000
acids

5
00:00:11,430 --> 00:00:09,920
and our first speaker

6
00:00:13,270 --> 00:00:11,440
is joanna

7
00:00:15,430 --> 00:00:13,280
i apologize if i mispronounce your name

8
00:00:19,029 --> 00:00:15,440
is it maisel

9
00:00:21,029 --> 00:00:19,039
who is joining us remotely from arizona

10
00:00:22,950 --> 00:00:21,039
um she is a professor at the university

11
00:00:24,710 --> 00:00:22,960
of arizona and she'll be talking to us

12
00:00:30,390 --> 00:00:24,720
about long-term evolution of the

13
00:00:35,030 --> 00:00:33,030

thank you for having me um and thank you

14
00:00:37,590 --> 00:00:35,040
everyone for who's there staying until

15
00:00:40,389 --> 00:00:37,600
the until the last session so i have a

16
00:00:42,470 --> 00:00:40,399
simpler title here um and basically even

17
00:00:44,389 --> 00:00:42,480
in the program what i uh

18
00:00:46,470 --> 00:00:44,399
what we're trying to do is figure out

19
00:00:48,709 --> 00:00:46,480
how the proteome evolves and that way we

20
00:00:51,350 --> 00:00:48,719
can project back and ask what the early

21
00:00:53,110 --> 00:00:51,360
proteome was like so this is a top-down

22
00:00:56,150 --> 00:00:53,120
approach of what we can deduce from

23
00:00:58,389 --> 00:00:56,160
modern proteins and in particular which

24
00:01:00,630 --> 00:00:58,399
amino acids are used and why they're

25
00:01:03,189 --> 00:01:00,640
used and sort of questions about you

26
00:01:06,149 --> 00:01:03,199
know that whether function drives this

27
00:01:08,469 --> 00:01:06,159
or availability drives this

28
00:01:10,789 --> 00:01:08,479
um so before you know we can really

29
00:01:13,510 --> 00:01:10,799
project back it's like how far back are

30
00:01:15,510 --> 00:01:13,520
we projecting um what is the origin of a

31
00:01:16,630 --> 00:01:15,520
given protein-coding gene that we look

32
00:01:18,390 --> 00:01:16,640
at

33
00:01:20,230 --> 00:01:18,400
and the traditional answer of how people

34
00:01:22,870 --> 00:01:20,240
thought about that was that if you look

35
00:01:25,109 --> 00:01:22,880
at some gene then it's diverged from

36
00:01:26,230 --> 00:01:25,119
some duplicate of some other gene but

37
00:01:27,910 --> 00:01:26,240
then you need to know where does that

38
00:01:29,990 --> 00:01:27,920

other gene come from well it must have

39
00:01:31,429 --> 00:01:30,000
diverged from after duplicating from

40
00:01:34,550 --> 00:01:31,439
some other gene

41
00:01:37,270 --> 00:01:34,560
and so the the view was one of some big

42
00:01:39,910 --> 00:01:37,280
ancient big bang of all genes in the

43
00:01:41,910 --> 00:01:39,920
distance past you know that that came

44
00:01:44,950 --> 00:01:41,920
from some primordial ancestor and it's

45
00:01:47,190 --> 00:01:44,960
the same genes sorting out ever since

46
00:01:49,510 --> 00:01:47,200
but that is that that view has recently

47
00:01:51,590 --> 00:01:49,520
been overturned and what we we now have

48
00:01:53,670 --> 00:01:51,600
sort of incontrovertible evidence in

49
00:01:55,510 --> 00:01:53,680
favor of is that at some rate and

50
00:01:58,389 --> 00:01:55,520
dispute what the rate is but at some

51
00:02:01,350 --> 00:01:58,399
rate there is continuous creation that

52
00:02:03,590 --> 00:02:01,360
basically de novo genes um come out of

53
00:02:06,550 --> 00:02:03,600
non-coded or frame shifting dna and they

54
00:02:08,229 --> 00:02:06,560
have no coding ancestor previously so

55
00:02:11,830 --> 00:02:08,239
this is very different to species that

56
00:02:14,229 --> 00:02:11,840
all go back to some sort of luca um

57
00:02:16,710 --> 00:02:14,239
genes all have separate origins

58
00:02:19,750 --> 00:02:16,720
throughout the history of life

59
00:02:21,750 --> 00:02:19,760
so what we then do is we can classify

60
00:02:24,390 --> 00:02:21,760
and a better thing to classify than

61
00:02:26,390 --> 00:02:24,400
genes turns out to be classified protein

62
00:02:28,790 --> 00:02:26,400
domains genes are sort of modular

63
00:02:30,790 --> 00:02:28,800

assortments of different genes

64
00:02:32,710 --> 00:02:30,800
domains that might have different ages

65
00:02:35,110 --> 00:02:32,720
so we classify each

66
00:02:36,949 --> 00:02:35,120
domain in the pfam database according to

67
00:02:38,949 --> 00:02:36,959
when it was born which we can figure out

68
00:02:41,030 --> 00:02:38,959
by when it has homologs

69
00:02:42,710 --> 00:02:41,040
and again we're focusing on homologs not

70
00:02:45,190 --> 00:02:42,720
orthologs a lot of people focus on

71
00:02:48,229 --> 00:02:45,200
orthologs because they're trying to

72
00:02:50,630 --> 00:02:48,239
um deduce function orthologs of the idea

73
00:02:53,830 --> 00:02:50,640
that it's somehow the same gene rather

74
00:02:56,150 --> 00:02:53,840
than some paralog which is also related

75
00:02:57,750 --> 00:02:56,160
but a different gene and that's not

76
00:02:59,430 --> 00:02:57,760
that's not an evolutionary rigorous

77
00:03:01,030 --> 00:02:59,440
distinction but whether or not they're

78
00:03:04,470 --> 00:03:01,040
related to each other by descent with

79
00:03:05,350 --> 00:03:04,480
modification is so include all homologs

80
00:03:08,630 --> 00:03:05,360
um

81
00:03:11,350 --> 00:03:08,640
and then we look at trends as a function

82
00:03:13,350 --> 00:03:11,360
of how long they've had to evolve

83
00:03:15,110 --> 00:03:13,360
and i originally started out very

84
00:03:16,949 --> 00:03:15,120
interested in things like aggregation

85
00:03:18,710 --> 00:03:16,959
propensity and intrinsic structural

86
00:03:20,869 --> 00:03:18,720
disorder and i've become a bit

87
00:03:22,550 --> 00:03:20,879
disenchanted with that over time because

88
00:03:24,949 --> 00:03:22,560

what we found is all the lovely

89
00:03:27,589 --> 00:03:24,959
predictors that take sequences and tell

90
00:03:29,350 --> 00:03:27,599
you what they do it turns out that they

91
00:03:31,270 --> 00:03:29,360
tell you something almost identical if

92
00:03:32,070 --> 00:03:31,280
you take the amino acids and you

93
00:03:34,550 --> 00:03:32,080
ins

94
00:03:36,470 --> 00:03:34,560
and you feed them in random order so the

95
00:03:38,309 --> 00:03:36,480
main predictors tend to be the

96
00:03:41,830 --> 00:03:38,319
frequencies of each of the 20 amino

97
00:03:43,589 --> 00:03:41,840
acids and most stuff follows from that

98
00:03:46,229 --> 00:03:43,599
and so here's an example of that where

99
00:03:48,789 --> 00:03:46,239
we look at the the frequency of proline

100
00:03:50,470 --> 00:03:48,799
um across all the pfam domains we've

101
00:03:53,830 --> 00:03:50,480
looked at and you can see there's a

102
00:03:56,550 --> 00:03:53,840
strong trend in brown among uh domains

103
00:03:59,429 --> 00:03:56,560
that have arisen in animals um it's much

104
00:04:02,550 --> 00:03:59,439
flatter in the green in in domains that

105
00:04:04,630 --> 00:04:02,560
arose in plants and relatively flat also

106
00:04:07,509 --> 00:04:04,640
among ancient domains of different

107
00:04:08,630 --> 00:04:07,519
levels of how ancient

108
00:04:10,789 --> 00:04:08,640
um

109
00:04:13,110 --> 00:04:10,799
and so if we take the slope of each of

110
00:04:15,030 --> 00:04:13,120
these and we plot the 20 slopes for the

111
00:04:17,189 --> 00:04:15,040
20 amino acids and we do this for the

112
00:04:19,110 --> 00:04:17,199
three most ancient groups

113
00:04:21,270 --> 00:04:19,120

what we see is there's a correlation

114
00:04:23,110 --> 00:04:21,280
with the hypothesized order in which the

115
00:04:25,270 --> 00:04:23,120
amino acids were recruited into the

116
00:04:27,830 --> 00:04:25,280
genetic code and what their slope is so

117
00:04:29,749 --> 00:04:27,840
what this is saying that the amino acids

118
00:04:32,870 --> 00:04:29,759
that were

119
00:04:35,110 --> 00:04:32,880
first

120
00:04:38,070 --> 00:04:35,120
are over represented

121
00:04:42,390 --> 00:04:38,080
in domains that date back to luca

122
00:04:43,749 --> 00:04:42,400
relative to other old domains um and we

123
00:04:45,030 --> 00:04:43,759
think the

124
00:04:47,270 --> 00:04:45,040
reason for this is even though the

125
00:04:49,990 --> 00:04:47,280
genetic code we're assuming had sort of

126
00:04:53,110 --> 00:04:50,000
settled down by luca that nevertheless

127
00:04:55,430 --> 00:04:53,120
at that point in time these amino acids

128
00:04:57,350 --> 00:04:55,440
remained much more available

129
00:05:00,070 --> 00:04:57,360
and so they were used more because of

130
00:05:02,150 --> 00:05:00,080
that and some of the other newer amino a

131
00:05:03,909 --> 00:05:02,160
amino acids were still somewhat oddities

132
00:05:06,230 --> 00:05:03,919
that were less available

133
00:05:08,550 --> 00:05:06,240
and we see the same bias towards using

134
00:05:10,629 --> 00:05:08,560
available amino acids

135
00:05:12,629 --> 00:05:10,639
in plants it's just it's a different set

136
00:05:14,950 --> 00:05:12,639
of amino acids that count as available

137
00:05:17,029 --> 00:05:14,960
in plants there's a lot of cysteine

138
00:05:19,670 --> 00:05:17,039

cellularly because it's produced during

139
00:05:21,350 --> 00:05:19,680
sulfur assimilation and it's also pretty

140
00:05:23,350 --> 00:05:21,360
important against reactive oxygen

141
00:05:25,749 --> 00:05:23,360
cysteine is very metabolically available

142
00:05:27,830 --> 00:05:25,759
and around glutamate and aspartate are

143
00:05:29,990 --> 00:05:27,840
also very abundant and those are the

144
00:05:32,710 --> 00:05:30,000
three amino acids that we see

145
00:05:34,870 --> 00:05:32,720
enriched in younger

146
00:05:37,110 --> 00:05:34,880
plant domains that when new stuff gets

147
00:05:38,629 --> 00:05:37,120
invented it tends to use what's most

148
00:05:40,790 --> 00:05:38,639
available

149
00:05:43,350 --> 00:05:40,800
situation is different in

150
00:05:45,830 --> 00:05:43,360
animals where we find more evidence that

151
00:05:47,590 --> 00:05:45,840
function is driving things so we we

152
00:05:49,670 --> 00:05:47,600
estimated um there was a big

153
00:05:52,550 --> 00:05:49,680
experimental evolution done in dtad

154
00:05:54,950 --> 00:05:52,560
towtz's lab where random peptides were

155
00:05:56,870 --> 00:05:54,960
were expressed in plasmids and the

156
00:05:59,670 --> 00:05:56,880
lineages were competed against each

157
00:06:01,749 --> 00:05:59,680
other and we calculated the marginal

158
00:06:03,189 --> 00:06:01,759
effect of having one amino acid versus

159
00:06:03,990 --> 00:06:03,199
another

160
00:06:06,550 --> 00:06:04,000
um

161
00:06:09,909 --> 00:06:06,560
and we found that those marginal effects

162
00:06:12,390 --> 00:06:09,919
correlated uh with um

163
00:06:14,790 --> 00:06:12,400

uh with these phyllo stratigraphy trends

164
00:06:17,590 --> 00:06:14,800
in animals so that young pro animal

165
00:06:19,830 --> 00:06:17,600
proteins tend to be using more harmless

166
00:06:22,550 --> 00:06:19,840
amino

167
00:06:25,350 --> 00:06:22,560
um and we actually see when we have

168
00:06:27,510 --> 00:06:25,360
another technique where we look at which

169
00:06:29,909 --> 00:06:27,520
amino acids are

170
00:06:32,309 --> 00:06:29,919
very slightly preferred basically in

171
00:06:34,230 --> 00:06:32,319
species that have stronger codon bias

172
00:06:36,629 --> 00:06:34,240
compared to species with less stronger

173
00:06:38,550 --> 00:06:36,639
codon bias so species that are able to

174
00:06:39,510 --> 00:06:38,560
make finer distinctions and those that

175
00:06:41,990 --> 00:06:39,520
aren't

176
00:06:44,550 --> 00:06:42,000
uh we find that the same

177
00:06:46,950 --> 00:06:44,560
amino acids are preferred today in

178
00:06:50,870 --> 00:06:46,960
vertebrates as are also

179
00:06:53,830 --> 00:06:50,880
preferred in this e coli experiment

180
00:06:55,830 --> 00:06:53,840
um one trend only we found to be

181
00:06:56,790 --> 00:06:55,840
consistent across the whole history of

182
00:06:59,589 --> 00:06:56,800
life

183
00:07:02,070 --> 00:06:59,599
and that is a a metric if you take the

184
00:07:03,830 --> 00:07:02,080
five most hydrophobic amino acids in

185
00:07:05,510 --> 00:07:03,840
some proteins like in the top here

186
00:07:07,830 --> 00:07:05,520
they're very clustered along the the

187
00:07:09,990 --> 00:07:07,840
primary sequence and in other amino

188
00:07:11,909 --> 00:07:10,000

acids they're very dispersed

189
00:07:14,830 --> 00:07:11,919
and there's a huge trend in this that

190
00:07:17,350 --> 00:07:14,840
goes back basically as far as we can

191
00:07:19,670 --> 00:07:17,360
reconstruct that um

192
00:07:22,150 --> 00:07:19,680
young genes are random and clustering

193
00:07:23,670 --> 00:07:22,160
value of one means it's basically random

194
00:07:26,469 --> 00:07:23,680
and genes that have had a long time to

195
00:07:28,710 --> 00:07:26,479
evolve uh we see this this more

196
00:07:30,790 --> 00:07:28,720
interspersion result where the

197
00:07:33,749 --> 00:07:30,800
hydrophobic and amino acids are less

198
00:07:36,309 --> 00:07:33,759
likely to be near one another

199
00:07:38,790 --> 00:07:36,319
when we see these trends there are sort

200
00:07:40,309 --> 00:07:38,800
of two mechanisms that that we think of

201
00:07:42,390 --> 00:07:40,319
what might be driving them and i think

202
00:07:44,390 --> 00:07:42,400
what most people immediately jump to is

203
00:07:45,830 --> 00:07:44,400
okay if older things

204
00:07:48,550 --> 00:07:45,840
have done something they've had more

205
00:07:50,230 --> 00:07:48,560
time to evolve and the classic process

206
00:07:52,629 --> 00:07:50,240
of evolution by descent with

207
00:07:54,629 --> 00:07:52,639
modification where alleles that are more

208
00:07:56,390 --> 00:07:54,639
in one direction take over from ours

209
00:07:58,390 --> 00:07:56,400
that aren't that this descent with

210
00:08:00,390 --> 00:07:58,400
modification drives it and somehow it's

211
00:08:01,510 --> 00:08:00,400
just so slow that it's taken all this

212
00:08:03,670 --> 00:08:01,520
time

213
00:08:05,830 --> 00:08:03,680

but another hypothesis is that

214
00:08:08,390 --> 00:08:05,840
everything was there originally with

215
00:08:11,029 --> 00:08:08,400
huge diversity but some things have been

216
00:08:13,189 --> 00:08:11,039
differentially lost so what we're seeing

217
00:08:15,350 --> 00:08:13,199
over longer longer periods of time is

218
00:08:17,029 --> 00:08:15,360
the survivors who are always like that

219
00:08:18,710 --> 00:08:17,039
even when they were born

220
00:08:19,830 --> 00:08:18,720
um but they're the ones who made the

221
00:08:22,150 --> 00:08:19,840
distance

222
00:08:23,589 --> 00:08:22,160
so we're currently trying to figure out

223
00:08:25,430 --> 00:08:23,599
which trends are driven by which of

224
00:08:27,029 --> 00:08:25,440
these mechanisms

225
00:08:28,710 --> 00:08:27,039
um and you know this is sort of when

226
00:08:30,629 --> 00:08:28,720
we're trying to think what luca was like

227
00:08:32,149 --> 00:08:30,639
you know just like we all know i think

228
00:08:35,110 --> 00:08:32,159
that most species that ever lived are

229
00:08:37,750 --> 00:08:35,120
now extinct the same is likely true for

230
00:08:40,389 --> 00:08:37,760
lucas protein domains most of them have

231
00:08:42,149 --> 00:08:40,399
no contemporary descendants we only

232
00:08:44,230 --> 00:08:42,159
study the ones that have contemporary

233
00:08:46,949 --> 00:08:44,240
descendants

234
00:08:49,829 --> 00:08:46,959
and so we use this this major maximum

235
00:08:52,630 --> 00:08:49,839
likelihood technique to attempt to uh

236
00:08:55,670 --> 00:08:52,640
quantify the rate of loss along total

237
00:08:57,910 --> 00:08:55,680
loss of a pfam domain across different

238
00:08:59,990 --> 00:08:57,920

lineages and what we find is a

239
00:09:02,150 --> 00:09:00,000
non-linear effect where there is an

240
00:09:04,389 --> 00:09:02,160
optimal value and this is shown here for

241
00:09:06,389 --> 00:09:04,399
the clustering metric and that optimal

242
00:09:09,110 --> 00:09:06,399
value with the lowest level of loss does

243
00:09:11,829 --> 00:09:09,120
indeed match that that you see in the

244
00:09:14,310 --> 00:09:11,839
very oldest pfams and this could help

245
00:09:17,430 --> 00:09:14,320
explain you know on the same lines

246
00:09:19,430 --> 00:09:17,440
what we see is we see greater variation

247
00:09:22,230 --> 00:09:19,440
among the younger domains and less

248
00:09:23,910 --> 00:09:22,240
variation among the older domains so you

249
00:09:25,670 --> 00:09:23,920
know this is really showing so some

250
00:09:28,470 --> 00:09:25,680
evidence that differential loss is

251
00:09:30,550 --> 00:09:28,480
driving some of this

252
00:09:32,550 --> 00:09:30,560
so to ask you know what was the early

253
00:09:34,790 --> 00:09:32,560
protein i'm like well we're still

254
00:09:37,269 --> 00:09:34,800
looking into it but sort of preliminary

255
00:09:39,030 --> 00:09:37,279
conclusions so far

256
00:09:40,310 --> 00:09:39,040
is firstly that the contemporary

257
00:09:41,829 --> 00:09:40,320
descendants are probably

258
00:09:43,990 --> 00:09:41,839
unrepresentative

259
00:09:46,470 --> 00:09:44,000
um they've had more time to evolve and

260
00:09:48,150 --> 00:09:46,480
they're a highly biased set of of

261
00:09:50,389 --> 00:09:48,160
descendants

262
00:09:52,870 --> 00:09:50,399
and so really looking at the field of de

263
00:09:55,509 --> 00:09:52,880

novo genes and what things get invented

264
00:09:57,350 --> 00:09:55,519
from scratch could be informative and we

265
00:09:58,870 --> 00:09:57,360
should consider the likelihood that

266
00:10:01,590 --> 00:09:58,880
there was a lot of that kind of thing

267
00:10:04,550 --> 00:10:01,600
around back in the ancient proteome and

268
00:10:07,670 --> 00:10:04,560
we just no longer see its descendants

269
00:10:09,670 --> 00:10:07,680
um and we also have some kind of hints

270
00:10:11,269 --> 00:10:09,680
that amino acids that were abundant back

271
00:10:13,910 --> 00:10:11,279
then were perhaps a bit more common than

272
00:10:15,190 --> 00:10:13,920
now in particular glycine alanine

273
00:10:16,949 --> 00:10:15,200
invalid

274
00:10:19,190 --> 00:10:16,959
so those are those are the preliminary

275
00:10:20,790 --> 00:10:19,200
conclusions as we continue

276
00:10:23,110 --> 00:10:20,800
to work on this

277
00:10:25,670 --> 00:10:23,120
um so thanks especially to the people in

278
00:10:28,630 --> 00:10:25,680
the top row who did uh this was a lot of

279
00:10:30,310 --> 00:10:28,640
work i compressed into this and also the

280
00:10:32,550 --> 00:10:30,320
people in the bottom row but you know

281
00:10:35,110 --> 00:10:32,560
their contributions and more people who

282
00:10:37,030 --> 00:10:35,120
who aren't even listed here um

283
00:10:40,470 --> 00:10:37,040
i try to put it in in as tightly as

284
00:10:48,150 --> 00:10:46,150
[Applause]

285
00:10:50,870 --> 00:10:48,160
thank you very much joanna we have a

286
00:10:53,190 --> 00:10:50,880
question for you from the audience

287
00:10:54,230 --> 00:10:53,200
hi this is anthony brunetti from georgia

288
00:10:56,630 --> 00:10:54,240

tech i

289
00:10:58,430 --> 00:10:56,640
saw you were seeing differences in the

290
00:11:00,389 --> 00:10:58,440
clustering of

291
00:11:02,870 --> 00:11:00,399
hydrophobicity in

292
00:11:05,269 --> 00:11:02,880
sequences identified as young and old

293
00:11:06,949 --> 00:11:05,279
um could that have anything to do with

294
00:11:10,710 --> 00:11:06,959
preferences for different kinds of

295
00:11:12,870 --> 00:11:10,720
secondary structure in old versus young

296
00:11:15,670 --> 00:11:12,880
domains

297
00:11:17,590 --> 00:11:15,680
um we don't think so so what we think

298
00:11:19,430 --> 00:11:17,600
this drives this

299
00:11:22,150 --> 00:11:19,440
is the uh you know proteins have to do

300
00:11:23,829 --> 00:11:22,160
two things they have to avoid doing harm

301
00:11:25,829 --> 00:11:23,839
they have to avoid aggregating and

302
00:11:27,190 --> 00:11:25,839
misfolding and so on and they also have

303
00:11:29,910 --> 00:11:27,200
to do good

304
00:11:32,630 --> 00:11:29,920
and what we believe is driving this is

305
00:11:34,630 --> 00:11:32,640
the avoidance of harm interestingly we

306
00:11:37,509 --> 00:11:34,640
weren't the first person to observe this

307
00:11:39,430 --> 00:11:37,519
this anti-clustering um but it was

308
00:11:42,790 --> 00:11:39,440
previously attributed

309
00:11:44,710 --> 00:11:42,800
to all proteins as a means of avoiding

310
00:11:46,310 --> 00:11:44,720
harm as a means of avoiding aggregation

311
00:11:48,630 --> 00:11:46,320
you know that having too many in a row

312
00:11:51,269 --> 00:11:48,640
during the translation process is going

313
00:11:53,350 --> 00:11:51,279

to increase the the chance of something

314
00:11:55,430 --> 00:11:53,360
going wrong at that point and what we

315
00:11:57,190 --> 00:11:55,440
found is that the that it's found

316
00:11:58,629 --> 00:11:57,200
different you know only in old proteins

317
00:12:03,670 --> 00:11:58,639
and not in yeah

318
00:12:07,269 --> 00:12:05,670
i think we have um a few minutes so i

319
00:12:10,470 --> 00:12:07,279
actually if you don't mind i want to

320
00:12:11,829 --> 00:12:10,480
maybe ask a question which was um in a

321
00:12:14,629 --> 00:12:11,839
number of your plots when you have an

322
00:12:17,269 --> 00:12:14,639
x-axis labeled age of the p-fam in

323
00:12:20,629 --> 00:12:17,279
billions of years i'm just curious what

324
00:12:22,470 --> 00:12:20,639
what metric do we use to

325
00:12:25,990 --> 00:12:22,480
what's our sort of way of inferring that

326
00:12:29,190 --> 00:12:26,000
or guessing that for a given p-fam

327
00:12:31,110 --> 00:12:29,200
yeah so the method we're using is to

328
00:12:34,710 --> 00:12:31,120
have a big tree of life and to see where

329
00:12:36,150 --> 00:12:34,720
the homologues um are detected and so in

330
00:12:39,350 --> 00:12:36,160
the older

331
00:12:42,710 --> 00:12:39,360
uh you know so for some of these uh

332
00:12:44,470 --> 00:12:42,720
younger pfabs that's relatively good we

333
00:12:46,629 --> 00:12:44,480
you know all of that these all come from

334
00:12:48,470 --> 00:12:46,639
time tree and the somewhat consensus

335
00:12:50,629 --> 00:12:48,480
estimates of when the diver you know

336
00:12:52,550 --> 00:12:50,639
based on the divergence of the species

337
00:12:55,110 --> 00:12:52,560
level um

338
00:12:57,110 --> 00:12:55,120

and then there's a lot more uncertainty

339
00:12:59,990 --> 00:12:57,120
obviously as you all know among these

340
00:13:01,590 --> 00:13:00,000
older age groups but the oldest is

341
00:13:04,629 --> 00:13:01,600
basically those who that have been

342
00:13:06,949 --> 00:13:04,639
attributed to being in in lucca

343
00:13:09,110 --> 00:13:06,959
um the ones up uh

344
00:13:11,430 --> 00:13:09,120
younger among these older ones are ones

345
00:13:14,230 --> 00:13:11,440
that are found both in eukaryote

346
00:13:16,069 --> 00:13:14,240
a fairly basal branch of eukaryotes and

347
00:13:17,670 --> 00:13:16,079
also at least plants and animals because

348
00:13:20,310 --> 00:13:17,680
we were doing a plant and animals

349
00:13:22,230 --> 00:13:20,320
focused study here and in between we

350
00:13:24,470 --> 00:13:22,240
have things that are found only you know

351
00:13:25,750 --> 00:13:24,480
in prokaryotes but aren't believed to be

352
00:13:27,430 --> 00:13:25,760
in luca

353
00:13:29,430 --> 00:13:27,440
and what numbers you want to give to

354
00:13:31,990 --> 00:13:29,440
these uh could definitely be open to

355
00:13:33,590 --> 00:13:32,000
interpretation

356
00:13:35,750 --> 00:13:33,600
lovely thank you

357
00:13:36,800 --> 00:13:35,760
please join me in thanking our first

358
00:13:41,430 --> 00:13:36,810
presenter

359
00:13:45,590 --> 00:13:43,430
um the second presentation will be by

360
00:13:47,990 --> 00:13:45,600
valeria giacobelli

361
00:13:49,670 --> 00:13:48,000
who is visiting us from the charles

362
00:13:52,790 --> 00:13:49,680
university of prague of the czech

363
00:13:55,670 --> 00:13:52,800

republic he is a postdoctoral fellow in

364
00:13:57,990 --> 00:13:55,680
the laboratory of clara huchovo

365
00:13:59,750 --> 00:13:58,000
hey hi everybody i'm valerio from

366
00:14:02,389 --> 00:13:59,760
charles university and i'm really

367
00:14:04,710 --> 00:14:02,399
thrilled today to show our recent work

368
00:14:07,509 --> 00:14:04,720
in vitro evolution reveal non-cationic

369
00:14:09,670 --> 00:14:07,519
protein rna interaction mediated by

370
00:14:11,910 --> 00:14:09,680
metal ions so

371
00:14:13,670 --> 00:14:11,920
briefly introduction i think we're

372
00:14:15,829 --> 00:14:13,680
how was the composition of

373
00:14:17,430 --> 00:14:15,839
probiotic work how we already like

374
00:14:18,389 --> 00:14:17,440
heavily discussed in this conference we

375
00:14:21,030 --> 00:14:18,399
know that

376
00:14:23,829 --> 00:14:21,040
two polymer mostly dominated the scene

377
00:14:25,590 --> 00:14:23,839
and was like peptide and rna

378
00:14:27,750 --> 00:14:25,600
and in some point so we can argue who

379
00:14:29,670 --> 00:14:27,760
was first the irony war peptide word but

380
00:14:32,230 --> 00:14:29,680
what we know that in some point of the

381
00:14:35,350 --> 00:14:32,240
evolution these two pop these two

382
00:14:36,870 --> 00:14:35,360
polymers interact each other and

383
00:14:38,870 --> 00:14:36,880
it's important to notice that the

384
00:14:39,990 --> 00:14:38,880
composition of the ancient peptide there

385
00:14:42,629 --> 00:14:40,000
are different theory about the

386
00:14:45,030 --> 00:14:42,639
composition of the ancient peptide we

387
00:14:47,350 --> 00:14:45,040
know that uh as really the previous talk

388
00:14:49,030 --> 00:14:47,360

described the composition of the amino

389
00:14:50,710 --> 00:14:49,040
acid composition of the peptide was

390
00:14:53,590 --> 00:14:50,720
different so we can

391
00:14:55,509 --> 00:14:53,600
probably was much easier alphabet than

392
00:14:57,350 --> 00:14:55,519
what we have now so we can distinguish

393
00:14:59,990 --> 00:14:57,360
like two classes of amino acid like

394
00:15:00,949 --> 00:15:00,000
early amino acid and late amino acid and

395
00:15:04,150 --> 00:15:00,959
we can

396
00:15:06,069 --> 00:15:04,160
also hear that mostly the amino acids

397
00:15:08,150 --> 00:15:06,079
are like the

398
00:15:10,389 --> 00:15:08,160
there are no positive charge but only

399
00:15:12,870 --> 00:15:10,399
negative charge and aliphatic one so how

400
00:15:14,710 --> 00:15:12,880
it's possible that in probiotic war like

401
00:15:17,990 --> 00:15:14,720
negative charge molecules can interact

402
00:15:20,230 --> 00:15:18,000
each other uh or hypothetical like how

403
00:15:22,310 --> 00:15:20,240
can be this interaction between rna and

404
00:15:25,350 --> 00:15:22,320
ancient peptide there are two

405
00:15:27,430 --> 00:15:25,360
hypothetical like mechanism one it's the

406
00:15:29,749 --> 00:15:27,440
most study in its presence still like in

407
00:15:31,110 --> 00:15:29,759
the modern cells it's like through

408
00:15:32,870 --> 00:15:31,120
electrostatic interaction between

409
00:15:34,710 --> 00:15:32,880
positive and negative the positive

410
00:15:37,509 --> 00:15:34,720
charge of like

411
00:15:39,350 --> 00:15:37,519
arginine lysine and phosphate backbone

412
00:15:40,949 --> 00:15:39,360
and in the case of probiotic board could

413
00:15:42,550 --> 00:15:40,959

be possible that

414
00:15:44,310 --> 00:15:42,560
not arginine was not present there but

415
00:15:47,189 --> 00:15:44,320
was some non-canonic amino acid that

416
00:15:49,509 --> 00:15:47,199
during the evolution just disappear or

417
00:15:51,670 --> 00:15:49,519
another theory that it's what i'm going

418
00:15:53,350 --> 00:15:51,680
to talk today that it's just for the

419
00:15:55,590 --> 00:15:53,360
moment hypothetical that this

420
00:15:58,550 --> 00:15:55,600
interaction between post between

421
00:16:02,389 --> 00:15:58,560
negative charge polymers can be mediated

422
00:16:03,590 --> 00:16:02,399
by metal ions in particularly magnesium

423
00:16:06,150 --> 00:16:03,600
um

424
00:16:07,829 --> 00:16:06,160
how we try to verify this hypothesis

425
00:16:11,189 --> 00:16:07,839
like first of all we select like a

426
00:16:12,629 --> 00:16:11,199
template uh rna binding protein and try

427
00:16:14,470 --> 00:16:12,639
and substitute a

428
00:16:16,550 --> 00:16:14,480
create a library where all the late

429
00:16:18,470 --> 00:16:16,560
amino acids were substituted with the

430
00:16:20,310 --> 00:16:18,480
early amino acid so we have a protein

431
00:16:21,829 --> 00:16:20,320
composed of all the early amino acids

432
00:16:24,870 --> 00:16:21,839
and we will try to understand if he's

433
00:16:27,030 --> 00:16:24,880
still able to bind the rna

434
00:16:29,110 --> 00:16:27,040
so the target we selected was the

435
00:16:31,829 --> 00:16:29,120
ribosomal protein the

436
00:16:35,670 --> 00:16:31,839
c-terminal of the ribosomal protein l11

437
00:16:37,189 --> 00:16:35,680
from geobasiluk geobasilus

438
00:16:39,829 --> 00:16:37,199

we select this target because it was

439
00:16:42,389 --> 00:16:39,839
small domain 80 amino acids so simple to

440
00:16:45,110 --> 00:16:42,399
manage especially from the to manage the

441
00:16:47,509 --> 00:16:45,120
library from from this one uh already

442
00:16:49,670 --> 00:16:47,519
reach of early amino acids more than 74

443
00:16:51,749 --> 00:16:49,680
percent was already early amino acid we

444
00:16:54,470 --> 00:16:51,759
know everything about it it's conserved

445
00:16:57,350 --> 00:16:54,480
we know the crystal structure and uh and

446
00:16:59,670 --> 00:16:57,360
we know we know like the the rna target

447
00:17:03,509 --> 00:16:59,680
so the the target that this protein bind

448
00:17:05,750 --> 00:17:03,519
this rna binding protein bind and

449
00:17:07,829 --> 00:17:05,760
after that we create like um

450
00:17:10,949 --> 00:17:07,839
we generate our library so where every

451
00:17:13,189 --> 00:17:10,959
late amino acid will randomize with the

452
00:17:15,029 --> 00:17:13,199
set of early one here we can see and the

453
00:17:17,029 --> 00:17:15,039
end we obtained like a library of the

454
00:17:19,669 --> 00:17:17,039
size around 10 to the power of 10

455
00:17:21,110 --> 00:17:19,679
variants now with this kind of big

456
00:17:23,110 --> 00:17:21,120
library

457
00:17:25,029 --> 00:17:23,120
we have to select we have to select the

458
00:17:27,270 --> 00:17:25,039
variant and verify if there is something

459
00:17:29,590 --> 00:17:27,280
that's able to still bind dna and the

460
00:17:32,150 --> 00:17:29,600
method that we selected for the for this

461
00:17:34,390 --> 00:17:32,160
purpose but the mrna display

462
00:17:35,990 --> 00:17:34,400
quickly the amino display is like a

463
00:17:38,549 --> 00:17:36,000

technique a selection method that binds

464
00:17:41,029 --> 00:17:38,559
together the genotype and the phenotype

465
00:17:42,870 --> 00:17:41,039
through puromycin so we have like we can

466
00:17:44,870 --> 00:17:42,880
select the function through the protein

467
00:17:47,750 --> 00:17:44,880
that is bind to the

468
00:17:49,430 --> 00:17:47,760
to its own mrna so we can sequence so

469
00:17:52,710 --> 00:17:49,440
once we selected one variant we can

470
00:17:54,789 --> 00:17:52,720
sequencing the sequence uh through mrna

471
00:17:56,630 --> 00:17:54,799
uh here it's described the general

472
00:17:59,510 --> 00:17:56,640
pipeline of the method so we have like

473
00:18:02,230 --> 00:17:59,520
the dna library we in vitro transcribe

474
00:18:04,390 --> 00:18:02,240
and legated to the pulaomycin molecules

475
00:18:06,230 --> 00:18:04,400
and after in cell free so without cells

476
00:18:08,310 --> 00:18:06,240
so i can just in the

477
00:18:10,710 --> 00:18:08,320
in vitro we translate it and we obtain

478
00:18:12,150 --> 00:18:10,720
the protein libraries linked to the rna

479
00:18:15,029 --> 00:18:12,160
and after

480
00:18:17,909 --> 00:18:15,039
we selected the we mobilized the rna

481
00:18:20,230 --> 00:18:17,919
target to to beat to a solid support and

482
00:18:22,310 --> 00:18:20,240
we selected the variant this cycle this

483
00:18:25,350 --> 00:18:22,320
technique it's repeated for like seven

484
00:18:28,630 --> 00:18:25,360
around in this case we perform 60 round

485
00:18:30,950 --> 00:18:28,640
and on the right we can

486
00:18:33,430 --> 00:18:30,960
we can see we're sequencing every round

487
00:18:35,830 --> 00:18:33,440
and we can see the arrangement of every

488
00:18:37,190 --> 00:18:35,840

in every position of the mutagenesis and

489
00:18:38,870 --> 00:18:37,200
we can see

490
00:18:41,830 --> 00:18:38,880
in every position of the library and we

491
00:18:44,470 --> 00:18:41,840
can see that step by step we selected

492
00:18:45,270 --> 00:18:44,480
the the population was a reached of like

493
00:18:47,029 --> 00:18:45,280
uh

494
00:18:48,870 --> 00:18:47,039
negative charge amino acid we can see

495
00:18:51,029 --> 00:18:48,880
like how the presence of aspartating and

496
00:18:52,710 --> 00:18:51,039
glutamine increase during the selection

497
00:18:55,110 --> 00:18:52,720
till we arrive to the last end when we

498
00:18:57,830 --> 00:18:55,120
select one variant the most abundant in

499
00:18:59,750 --> 00:18:57,840
the in the mix and um

500
00:19:01,750 --> 00:18:59,760
we select this one let's go like

501
00:19:04,070 --> 00:19:01,760
e variant

502
00:19:06,310 --> 00:19:04,080
uh after that that we have our variant

503
00:19:08,470 --> 00:19:06,320
we need to prove it so we express in

504
00:19:09,510 --> 00:19:08,480
nicolite purify it and verify the

505
00:19:11,669 --> 00:19:09,520
binding

506
00:19:14,710 --> 00:19:11,679
comparison to the wall type protein so

507
00:19:16,950 --> 00:19:14,720
we have a scale we have like

508
00:19:18,870 --> 00:19:16,960
a comparison and we perform different

509
00:19:21,350 --> 00:19:18,880
technique to verify the binding one of

510
00:19:23,510 --> 00:19:21,360
them was the amsa the electrophoretic

511
00:19:26,549 --> 00:19:23,520
mobility shift ac where we load on a

512
00:19:28,310 --> 00:19:26,559
native gel page gel like the free rna

513
00:19:30,950 --> 00:19:28,320

and the error in potential in the

514
00:19:33,270 --> 00:19:30,960
complex and we can just

515
00:19:34,390 --> 00:19:33,280
see the shift between these two

516
00:19:36,150 --> 00:19:34,400
um

517
00:19:37,990 --> 00:19:36,160
between both the complex and the free

518
00:19:40,150 --> 00:19:38,000
rna and we can observe that the

519
00:19:42,470 --> 00:19:40,160
e-variant compared to wall type showed

520
00:19:44,230 --> 00:19:42,480
the same binding at least it might

521
00:19:45,350 --> 00:19:44,240
uh after that we were curious to know

522
00:19:47,350 --> 00:19:45,360
how is the

523
00:19:49,909 --> 00:19:47,360
the structure the general structure of

524
00:19:52,230 --> 00:19:49,919
the protein uh in solution not binding

525
00:19:54,549 --> 00:19:52,240
and we can see that this mutation the er

526
00:19:56,310 --> 00:19:54,559
e-variants uh lost completely the

527
00:19:58,310 --> 00:19:56,320
secondary structure compared to the wall

528
00:20:01,430 --> 00:19:58,320
type that was like most of filix and how

529
00:20:03,029 --> 00:20:01,440
we can see the e-library the e-variants

530
00:20:05,149 --> 00:20:03,039
show like a

531
00:20:07,510 --> 00:20:05,159
peak around 200 nanometer in the

532
00:20:10,789 --> 00:20:07,520
circulation technique that show like

533
00:20:12,789 --> 00:20:10,799
that it's like highly disorder

534
00:20:15,590 --> 00:20:12,799
after that we try to quantify give some

535
00:20:18,149 --> 00:20:15,600
number about the binding so we perform

536
00:20:21,190 --> 00:20:18,159
the spr as sulfate plasmas resonance

537
00:20:23,830 --> 00:20:21,200
technique where we immobilize the

538
00:20:25,750 --> 00:20:23,840

the target rna on a chip and just pass

539
00:20:27,990 --> 00:20:25,760
on it like the protein the two protein

540
00:20:31,830 --> 00:20:28,000
dual type invariant and we calculated

541
00:20:35,110 --> 00:20:31,840
the association and dissociation binding

542
00:20:38,549 --> 00:20:35,120
constant and we can see that the

543
00:20:41,029 --> 00:20:38,559
e-variant uh bind much slower to the

544
00:20:44,149 --> 00:20:41,039
target but on the other hand compared to

545
00:20:46,710 --> 00:20:44,159
the wall type but once the the

546
00:20:48,470 --> 00:20:46,720
the protein bind the rna

547
00:20:51,110 --> 00:20:48,480
it's more stable the complex it's more

548
00:20:54,470 --> 00:20:51,120
stable the overall kde that is duration

549
00:20:56,310 --> 00:20:54,480
between on and off it's mostly similar

550
00:20:58,149 --> 00:20:56,320
to the to wall type but the difference

551
00:20:59,990 --> 00:20:58,159
is mostly in the dissociation and

552
00:21:01,750 --> 00:21:00,000
actually this is a it suggests that

553
00:21:04,390 --> 00:21:01,760
maybe the evolution during the evolution

554
00:21:06,149 --> 00:21:04,400
like something so stable on the rna it's

555
00:21:08,070 --> 00:21:06,159
not so advantageous if we imagine like a

556
00:21:10,070 --> 00:21:08,080
ribosome or whatever or every mechanism

557
00:21:11,830 --> 00:21:10,080
in the cell it's something dynamic but

558
00:21:13,830 --> 00:21:11,840
here we have something that once it's

559
00:21:15,909 --> 00:21:13,840
bind it stay there so maybe the

560
00:21:18,230 --> 00:21:15,919
evolution also select this one to

561
00:21:21,110 --> 00:21:18,240
towards something more dynamic

562
00:21:23,669 --> 00:21:21,120
uh another a fire characterization was

563
00:21:26,549 --> 00:21:23,679

done by uh pull down technique so we

564
00:21:28,390 --> 00:21:26,559
mobilized the complex on a bit support

565
00:21:31,190 --> 00:21:28,400
and changing the parameter like

566
00:21:33,590 --> 00:21:31,200
temperature ph or the presence of iron

567
00:21:35,430 --> 00:21:33,600
we can destabilize or not the complex if

568
00:21:37,510 --> 00:21:35,440
the complex is destabilized the protein

569
00:21:39,270 --> 00:21:37,520
get released and we have a signal on

570
00:21:41,190 --> 00:21:39,280
western blot

571
00:21:43,350 --> 00:21:41,200
we notice that compared to the wall type

572
00:21:44,390 --> 00:21:43,360
the e-variant is much sensitive to

573
00:21:46,950 --> 00:21:44,400
temperature

574
00:21:49,270 --> 00:21:46,960
and it's pretty extreme ph but what was

575
00:21:50,789 --> 00:21:49,280
really interesting it was uh

576
00:21:53,350 --> 00:21:50,799
really interesting was that in the

577
00:21:55,750 --> 00:21:53,360
absence of completely iron or metal ions

578
00:21:58,710 --> 00:21:55,760
so in the buffer was just buffer

579
00:22:01,029 --> 00:21:58,720
the complex was destabilized

580
00:22:03,270 --> 00:22:01,039
but this did happen in the case of the

581
00:22:05,830 --> 00:22:03,280
wall type so it means that these these

582
00:22:10,310 --> 00:22:05,840
ions were involved in somehow in

583
00:22:15,510 --> 00:22:13,110
to give fire suggestions like proof to

584
00:22:17,270 --> 00:22:15,520
this theory uh we perform in

585
00:22:18,470 --> 00:22:17,280
collaboration with the academia of

586
00:22:20,789 --> 00:22:18,480
science of czech republic in czech

587
00:22:21,990 --> 00:22:20,799
republic uh the molecular dynamics

588
00:22:24,149 --> 00:22:22,000

simulation

589
00:22:25,830 --> 00:22:24,159
uh we use as template the the crystal

590
00:22:27,990 --> 00:22:25,840
structure of the complex of the wall

591
00:22:28,870 --> 00:22:28,000
type that it's available it's a pdb it's

592
00:23:17,990 --> 00:22:28,880
a

593
00:23:20,070 --> 00:23:18,000
diet given like further proof to this

594
00:23:21,350 --> 00:23:20,080
experimental data that metal ion

595
00:23:23,430 --> 00:23:21,360
actually

596
00:23:27,590 --> 00:23:23,440
help to the the interface to bind

597
00:23:31,990 --> 00:23:29,830
so in conclusion uh

598
00:23:34,070 --> 00:23:32,000
first of all we demonstrate that an

599
00:23:37,750 --> 00:23:34,080
early protein composite of only early

600
00:23:39,830 --> 00:23:37,760
amino acid is still able to bind the rna

601
00:23:41,510 --> 00:23:39,840
and second for the first time we give

602
00:23:44,310 --> 00:23:41,520
like for the first the first

603
00:23:46,470 --> 00:23:44,320
experimental indication that cat on ion

604
00:23:48,870 --> 00:23:46,480
like magnesium can really help the

605
00:23:51,269 --> 00:23:48,880
interaction between rna and protein that

606
00:23:53,830 --> 00:23:51,279
can also be possible in the in modern

607
00:23:56,230 --> 00:23:53,840
world maybe just we didn't look at it

608
00:23:57,510 --> 00:23:56,240
but it's still possible it's another way

609
00:23:59,510 --> 00:23:57,520
of interaction

610
00:24:01,669 --> 00:23:59,520
and third one third

611
00:24:04,230 --> 00:24:01,679
third we can say that

612
00:24:07,350 --> 00:24:04,240
a word the probiotic word without late

613
00:24:10,149 --> 00:24:07,360

amino acid was possible and probably the

614
00:24:12,230 --> 00:24:10,159
they were inserted inside the evolution

615
00:24:15,190 --> 00:24:12,240
because just to help to fine-tune the

616
00:24:16,870 --> 00:24:15,200
interaction between rna and protein just

617
00:24:19,830 --> 00:24:16,880
to make everything more dynamic but

618
00:24:21,909 --> 00:24:19,840
anyway their absence still like

619
00:24:24,149 --> 00:24:21,919
even without them the the rna was still

620
00:24:26,230 --> 00:24:24,159
possible to interact with protein

621
00:24:28,230 --> 00:24:26,240
this work was published on molecular

622
00:24:30,390 --> 00:24:28,240
biology at evolution journal where we

623
00:24:32,310 --> 00:24:30,400
also got the cover and

624
00:24:34,070 --> 00:24:32,320
in this query code you can find the the

625
00:24:36,149 --> 00:24:34,080
paper if you want to read there are much

626
00:24:37,990 --> 00:24:36,159
more detail like scientific detail about

627
00:24:40,230 --> 00:24:38,000
experiment about the binding the

628
00:24:42,710 --> 00:24:40,240
structure and whatever and i want to

629
00:24:45,269 --> 00:24:42,720
really thanks like my colleague clara

630
00:24:47,110 --> 00:24:45,279
okova groups and

631
00:24:50,210 --> 00:24:47,120
and an ola collaborator and your for

632
00:24:56,310 --> 00:24:50,220
your attention thank you

633
00:24:59,750 --> 00:24:58,390
brilliant thank you valeria

634
00:25:06,950 --> 00:24:59,760
do you have any questions from the

635
00:25:11,190 --> 00:25:08,950
hi uh jessica bowman from georgia tech

636
00:25:12,710 --> 00:25:11,200
that was a super interesting talk um i'm

637
00:25:15,430 --> 00:25:12,720
from the lab

638
00:25:19,909 --> 00:25:18,070

and we are frequently looking at

639
00:25:22,630 --> 00:25:19,919
protein and rna interactions

640
00:25:25,190 --> 00:25:22,640
specifically rna from the ribosome

641
00:25:26,789 --> 00:25:25,200
one of your conclusions indicated that

642
00:25:29,590 --> 00:25:26,799
this is the first known

643
00:25:32,870 --> 00:25:29,600
interaction between a protein and

644
00:25:35,590 --> 00:25:32,880
ribosomal rna that's magnesium mediated

645
00:25:38,149 --> 00:25:35,600
if i recall correctly chelang shaw of

646
00:25:41,750 --> 00:25:38,159
our group published

647
00:25:46,789 --> 00:25:44,230
uh ribosomal protein

648
00:25:48,470 --> 00:25:46,799
in the rna that is magnesium mediated by

649
00:25:49,830 --> 00:25:48,480
an am n

650
00:25:51,990 --> 00:25:49,840
conserved

651
00:25:53,830 --> 00:25:52,000
region in that ult protein

652
00:25:56,310 --> 00:25:53,840
just a comment

653
00:25:57,750 --> 00:25:56,320
yeah actually we we also working on it

654
00:25:59,110 --> 00:25:57,760
like it's a parallel pro it's not my

655
00:26:01,110 --> 00:25:59,120
project but our colleague we are

656
00:26:02,710 --> 00:26:01,120
studying about this and like also yeah

657
00:26:04,549 --> 00:26:02,720
we noticed that in especially in the

658
00:26:06,549 --> 00:26:04,559
ribosome the presence of magnesium it's

659
00:26:09,430 --> 00:26:06,559
important to stabilize this

660
00:26:11,669 --> 00:26:09,440
so we can also fit this this model in in

661
00:26:13,350 --> 00:26:11,679
the recent world like of the ribosome so

662
00:26:15,750 --> 00:26:13,360
yeah

663
00:26:17,830 --> 00:26:15,760

thank you just one other comment

664
00:26:19,669 --> 00:26:17,840
interestingly in that case we have a

665
00:26:20,950 --> 00:26:19,679
later paper also i think chao longsha

666
00:26:23,029 --> 00:26:20,960
was the first author

667
00:26:24,710 --> 00:26:23,039
um that demonstrated

668
00:26:27,510 --> 00:26:24,720
interactions between

669
00:26:30,230 --> 00:26:27,520
a proposed ancestral ribosomal rna and

670
00:26:32,390 --> 00:26:30,240
some of these um ancestral peptides or

671
00:26:36,310 --> 00:26:32,400
hypothesized ancestral peptides one of

672
00:26:37,269 --> 00:26:36,320
which was ul2 we looked at uo2 ul3 ul4

673
00:26:39,669 --> 00:26:37,279
and

674
00:26:41,430 --> 00:26:39,679
what was interesting is that most of

675
00:26:43,430 --> 00:26:41,440
those um

676
00:26:45,350 --> 00:26:43,440
interactions were not magnesium was

677
00:26:47,350 --> 00:26:45,360
shown to disrupt the interaction between

678
00:26:49,269 --> 00:26:47,360
the protein and the rna in the case of

679
00:26:50,549 --> 00:26:49,279
ul2

680
00:26:52,549 --> 00:26:50,559
so just

681
00:26:56,950 --> 00:26:52,559
we can talk afterwards yeah sure sure

682
00:27:01,350 --> 00:26:59,510
chris may or bacon um university of

683
00:27:03,350 --> 00:27:01,360
maryland baltimore county very

684
00:27:06,230 --> 00:27:03,360
interesting talk

685
00:27:09,430 --> 00:27:06,240
in one of your slides you mentioned that

686
00:27:11,669 --> 00:27:09,440
the absence of magnesium

687
00:27:13,750 --> 00:27:11,679
or potassium

688
00:27:17,669 --> 00:27:13,760

disrupted the

689
00:27:20,710 --> 00:27:17,679
rna the rna binding and you showed md

690
00:27:22,789 --> 00:27:20,720
simulations about the role of magnesium

691
00:27:26,070 --> 00:27:22,799
i'm curious if

692
00:27:27,590 --> 00:27:26,080
i'm curious where the role of potassium

693
00:27:30,230 --> 00:27:27,600
ions come in

694
00:27:31,990 --> 00:27:30,240
in a stabilizing the rna protein

695
00:27:34,389 --> 00:27:32,000
interaction

696
00:27:36,310 --> 00:27:34,399
yeah okay i didn't perform like by

697
00:27:38,470 --> 00:27:36,320
myself was in collaboration but i know

698
00:27:41,029 --> 00:27:38,480
that like actually the first simulation

699
00:27:43,350 --> 00:27:41,039
was through uh potassium ion actually

700
00:27:45,350 --> 00:27:43,360
and they were like stuck there and after

701
00:27:47,669 --> 00:27:45,360
uh it substitute like the potassium with

702
00:27:50,070 --> 00:27:47,679
magnesium and it like confirmed this

703
00:27:51,909 --> 00:27:50,080
data so also the potassium ion were like

704
00:27:54,389 --> 00:27:51,919
present in the in the in the first

705
00:27:56,789 --> 00:27:54,399
simulation in the um in the structure in

706
00:27:58,310 --> 00:27:56,799
the interface

707
00:27:59,669 --> 00:27:58,320
all right interesting thank you all

708
00:28:01,350 --> 00:27:59,679
right i'm afraid we might we need to

709
00:28:03,669 --> 00:28:01,360
move on but we do have some time at the

710
00:28:06,470 --> 00:28:03,679
end for extra discussion so i apologize

711
00:28:09,510 --> 00:28:06,480
to the third questioner thank you um our

712
00:28:10,389 --> 00:28:09,520
third presentation is going to be from

713
00:28:13,190 --> 00:28:10,399

um

714
00:28:15,590 --> 00:28:13,200
uh dr pratik vias who

715
00:28:19,430 --> 00:28:15,600
is joining us remotely from the weizmann

716
00:28:20,870 --> 00:28:19,440
institute of science in rehovot israel

717
00:28:34,149 --> 00:28:20,880
he is a

718
00:28:37,510 --> 00:28:36,549
hi hi stephen and thank you for

719
00:28:38,950 --> 00:28:37,520
um

720
00:28:41,590 --> 00:28:38,960
having me

721
00:28:42,950 --> 00:28:41,600
um so basically my research like the

722
00:28:45,430 --> 00:28:42,960
broad goal of my research is to

723
00:28:46,630 --> 00:28:45,440
understand how did the first enzymes

724
00:28:47,430 --> 00:28:46,640
evolve

725
00:28:50,230 --> 00:28:47,440
and

726
00:28:52,149 --> 00:28:50,240
if like enzyme evolution basically it

727
00:28:54,789 --> 00:28:52,159
relates to recruitment of pre-existing

728
00:28:57,430 --> 00:28:54,799
enzymes to perform new function by a

729
00:28:59,190 --> 00:28:57,440
series of mutations and selections

730
00:29:01,669 --> 00:28:59,200
this is synonymous to teaching an old

731
00:29:02,950 --> 00:29:01,679
dog new tricks like enzymes being the

732
00:29:04,950 --> 00:29:02,960
old dogs

733
00:29:06,950 --> 00:29:04,960
but the key question in the field is

734
00:29:09,269 --> 00:29:06,960
that how and where did the old dog come

735
00:29:10,630 --> 00:29:09,279
about in the first place

736
00:29:12,070 --> 00:29:10,640
because if you look at the modern day

737
00:29:14,549 --> 00:29:12,080
proteins we know that they are

738
00:29:17,029 --> 00:29:14,559

incredibly complex and yet it tends to

739
00:29:20,310 --> 00:29:17,039
reason that in the pre-luca world these

740
00:29:22,630 --> 00:29:20,320
complex proteins likely emerged from

741
00:29:24,310 --> 00:29:22,640
precursors that were much more simpler

742
00:29:25,350 --> 00:29:24,320
both in terms of the sequence and the

743
00:29:26,950 --> 00:29:25,360
structure

744
00:29:29,350 --> 00:29:26,960
so what were these precursors of these

745
00:29:31,190 --> 00:29:29,360
complex proteins what kind of function

746
00:29:33,430 --> 00:29:31,200
did they possess what kind of structure

747
00:29:34,870 --> 00:29:33,440
did they possess and can we relate their

748
00:29:36,230 --> 00:29:34,880
structure and function to their modern

749
00:29:37,430 --> 00:29:36,240
day counterparts

750
00:29:40,230 --> 00:29:37,440
these are the questions that i'm trying

751
00:29:42,470 --> 00:29:40,240
to address in my in my work and

752
00:29:43,830 --> 00:29:42,480
specifically i'm trying to understand

753
00:29:45,590 --> 00:29:43,840
experimentally

754
00:29:48,950 --> 00:29:45,600
what were the precursors of this family

755
00:29:51,510 --> 00:29:48,960
of enzymes known as the t-loop ntp asses

756
00:29:53,669 --> 00:29:51,520
so the p-loop and the phases are one of

757
00:29:55,269 --> 00:29:53,679
the most diverse and abundant protein

758
00:29:56,950 --> 00:29:55,279
families that we know of

759
00:29:59,350 --> 00:29:56,960
these include

760
00:30:01,350 --> 00:29:59,360
uh complex macromolecular machines such

761
00:30:03,669 --> 00:30:01,360
as the atp synthesis

762
00:30:05,269 --> 00:30:03,679
regulatory combinations helicases and

763
00:30:07,269 --> 00:30:05,279

many other proteins that are implicated

764
00:30:10,230 --> 00:30:07,279
in essential life processes

765
00:30:12,149 --> 00:30:10,240
and also p loop ntps are one of the most

766
00:30:14,389 --> 00:30:12,159
ancient protein families that we know of

767
00:30:16,870 --> 00:30:14,399
and these are unambiguously assigned the

768
00:30:19,350 --> 00:30:16,880
last universal common ancestor

769
00:30:21,590 --> 00:30:19,360
so both these attributes make the p loop

770
00:30:23,669 --> 00:30:21,600
ndpas is attractive candidates to study

771
00:30:25,830 --> 00:30:23,679
protein evolution

772
00:30:28,149 --> 00:30:25,840
so in in all the peel panty pages the

773
00:30:29,909 --> 00:30:28,159
critical element is the walker a moti

774
00:30:32,230 --> 00:30:29,919
for the p-loop motif which is

775
00:30:34,950 --> 00:30:32,240
essentially a glycine rich loop

776
00:30:37,430 --> 00:30:34,960
that is embedded in a beta loop alpha

777
00:30:39,590 --> 00:30:37,440
element and the glycine which look

778
00:30:41,510 --> 00:30:39,600
mainly via the g k and the t that

779
00:30:44,870 --> 00:30:41,520
resides at the tip of the helix and also

780
00:30:47,350 --> 00:30:44,880
by the glycines it interacts with ntp

781
00:30:49,510 --> 00:30:47,360
the phosphates of ntp such as atp and

782
00:30:51,750 --> 00:30:49,520
gtp and mediates the transfer of the

783
00:30:55,269 --> 00:30:51,760
terminal phosphoryl group in reactions

784
00:30:58,310 --> 00:30:55,279
such as atp hydrolysis or atp synthesis

785
00:31:00,710 --> 00:30:58,320
so this extended beta philu palpa motif

786
00:31:03,350 --> 00:31:00,720
underlines all the enzymes that belong

787
00:31:05,990 --> 00:31:03,360
to the p loop ntps family

788
00:31:09,190 --> 00:31:06,000

and structurally the the core domain of

789
00:31:11,509 --> 00:31:09,200
p-loop ntps comprises of a three-layer

790
00:31:14,310 --> 00:31:11,519
alpha beta alpha sandwich-like

791
00:31:17,029 --> 00:31:14,320
architecture and almost always the

792
00:31:18,470 --> 00:31:17,039
p-loop is a part of the first beta loop

793
00:31:20,070 --> 00:31:18,480
and alpha element

794
00:31:22,310 --> 00:31:20,080
so with this background and with this

795
00:31:23,509 --> 00:31:22,320
structural information it brings me back

796
00:31:26,070 --> 00:31:23,519
to the question that i'm trying to

797
00:31:28,389 --> 00:31:26,080
address is that what were the precursors

798
00:31:29,990 --> 00:31:28,399
of e-loop ntps

799
00:31:31,990 --> 00:31:30,000
so to answer this question we need to

800
00:31:33,669 --> 00:31:32,000
first understand or we need to first ask

801
00:31:35,269 --> 00:31:33,679
is that what do these precursors

802
00:31:36,710 --> 00:31:35,279
actually do what kind of functions would

803
00:31:38,789 --> 00:31:36,720
they possess

804
00:31:41,669 --> 00:31:38,799
so previously it has been shown by liam

805
00:31:43,190 --> 00:31:41,679
longo who is also one of the speakers in

806
00:31:44,149 --> 00:31:43,200
today's session

807
00:31:46,549 --> 00:31:44,159
is that

808
00:31:48,310 --> 00:31:46,559
binding to phosphate containing ligands

809
00:31:50,789 --> 00:31:48,320
was one of the founding function or one

810
00:31:53,509 --> 00:31:50,799
of the ancient functions of not only the

811
00:31:55,509 --> 00:31:53,519
p loop entities but also by also of many

812
00:31:57,029 --> 00:31:55,519
other evolutionary ancient

813
00:31:59,029 --> 00:31:57,039

families such as the rosmans and the

814
00:32:01,430 --> 00:31:59,039
plebitoxin

815
00:32:03,590 --> 00:32:01,440
and in all these ancient families

816
00:32:06,310 --> 00:32:03,600
phosphate binding is realized by a

817
00:32:08,630 --> 00:32:06,320
stretch of simple abiotic amino acids

818
00:32:10,389 --> 00:32:08,640
such as glycine serine and threonine

819
00:32:12,710 --> 00:32:10,399
that reside at the end terminal tip of

820
00:32:15,509 --> 00:32:12,720
the helix and this interaction this

821
00:32:17,350 --> 00:32:15,519
phosphate binding interaction is via a

822
00:32:19,190 --> 00:32:17,360
wide-ended backbone interaction as well

823
00:32:21,029 --> 00:32:19,200
as a side chin interaction

824
00:32:22,870 --> 00:32:21,039
so we established that we concluded that

825
00:32:25,430 --> 00:32:22,880
phosphate binding functions was one of

826
00:32:28,310 --> 00:32:25,440
the ancient founding functions of of p

827
00:32:30,870 --> 00:32:28,320
loop n tps and this is what we set to

828
00:32:33,909 --> 00:32:30,880
assess if the ancient precursors of p

829
00:32:36,470 --> 00:32:33,919
loop and t phases can bind phosphates

830
00:32:37,509 --> 00:32:36,480
so before that our our our hypothesis

831
00:32:39,590 --> 00:32:37,519
was that

832
00:32:41,750 --> 00:32:39,600
the the beta p loop alpha motif that i

833
00:32:45,029 --> 00:32:41,760
just mentioned was one of the or was

834
00:32:46,950 --> 00:32:45,039
rather the the earliest standalone seed

835
00:32:49,509 --> 00:32:46,960
segment which then underwent

836
00:32:52,470 --> 00:32:49,519
self-assembly duplication and fusion to

837
00:32:54,710 --> 00:32:52,480
give rise to modern day peru ntpas

838
00:32:56,470 --> 00:32:54,720

and to test this strategy

839
00:32:58,630 --> 00:32:56,480
we use the static uh to test this

840
00:33:00,870 --> 00:32:58,640
hypothesis we use a strategy where we

841
00:33:03,110 --> 00:33:00,880
construct uh prototypes which are

842
00:33:04,950 --> 00:33:03,120
essentially mimics of ancient p loop

843
00:33:06,789 --> 00:33:04,960
entities

844
00:33:07,830 --> 00:33:06,799
so essentially what we do over here is

845
00:33:09,909 --> 00:33:07,840
we take

846
00:33:11,830 --> 00:33:09,919
uh the ancestrally reconstructed copies

847
00:33:15,029 --> 00:33:11,840
of the beta philip alpha from all the

848
00:33:17,990 --> 00:33:15,039
p-loop ntps and graphed it onto a very

849
00:33:20,310 --> 00:33:18,000
rudimentary scaffold that mimics the

850
00:33:21,909 --> 00:33:20,320
core domain of the p-loop ntps that is

851
00:33:23,509 --> 00:33:21,919
the three-layered alpha beta alpha

852
00:33:25,269 --> 00:33:23,519
sandwich architecture

853
00:33:27,590 --> 00:33:25,279
but it does not have any of the other

854
00:33:29,350 --> 00:33:27,600
active site residues that modern day p

855
00:33:31,750 --> 00:33:29,360
loop and tps have

856
00:33:34,950 --> 00:33:31,760
and then we see if these prototypes the

857
00:33:35,909 --> 00:33:34,960
simple prototypes can function

858
00:33:38,070 --> 00:33:35,919
so we

859
00:33:40,950 --> 00:33:38,080
interestingly we did see that the spiel

860
00:33:43,269 --> 00:33:40,960
of prototypes are bound to atp as shown

861
00:33:45,509 --> 00:33:43,279
here in an spr method that was just

862
00:33:48,070 --> 00:33:45,519
discussed by the previous speaker

863
00:33:50,389 --> 00:33:48,080

but what was more interesting was that

864
00:33:52,470 --> 00:33:50,399
these fragments of these proto proteins

865
00:33:54,789 --> 00:33:52,480
also bound single-stranded dna as you

866
00:33:56,470 --> 00:33:54,799
can see here by higher signal

867
00:33:58,710 --> 00:33:56,480
relative to the double-stranded dna in

868
00:33:59,990 --> 00:33:58,720
analyzer-based method

869
00:34:01,909 --> 00:34:00,000
so it was

870
00:34:03,909 --> 00:34:01,919
it was great that these prototypes by

871
00:34:05,669 --> 00:34:03,919
bound both ntps and single-standard dna

872
00:34:07,509 --> 00:34:05,679
and i must add that they bind to both

873
00:34:09,190 --> 00:34:07,519
these ligands via the same phosphate

874
00:34:11,669 --> 00:34:09,200
binding loop

875
00:34:14,629 --> 00:34:11,679
we now wanted to see if we can extend

876
00:34:16,950 --> 00:34:14,639
from the realm of just ligand binding

877
00:34:18,389 --> 00:34:16,960
and ask if these prototypes or these

878
00:34:20,069 --> 00:34:18,399
protoproteins

879
00:34:23,349 --> 00:34:20,079
does have any function which is of

880
00:34:24,950 --> 00:34:23,359
greater evolutionary relevance so we

881
00:34:26,470 --> 00:34:24,960
asked if these p look prototypes can

882
00:34:29,270 --> 00:34:26,480
remodel nucleic acid or more

883
00:34:31,510 --> 00:34:29,280
specifically if they can unwind dna

884
00:34:33,430 --> 00:34:31,520
given that they bind preferably to

885
00:34:35,430 --> 00:34:33,440
single-stranded dna can they shift the

886
00:34:38,069 --> 00:34:35,440
equilibrium from a double stranded bound

887
00:34:40,310 --> 00:34:38,079
form to a single standard bound form

888
00:34:42,230 --> 00:34:40,320

and we were basically guided by the

889
00:34:45,109 --> 00:34:42,240
observation that many of the luca p loop

890
00:34:47,109 --> 00:34:45,119
ntps were helicases recombinases

891
00:34:48,950 --> 00:34:47,119
and translocasis

892
00:34:51,109 --> 00:34:48,960
and it goes without saying that in the

893
00:34:53,510 --> 00:34:51,119
piluka world composed of nucleic acids

894
00:34:54,629 --> 00:34:53,520
and and proteins the ability to remodel

895
00:34:56,550 --> 00:34:54,639
nucleic acid would have been an

896
00:34:58,390 --> 00:34:56,560
important function

897
00:34:59,990 --> 00:34:58,400
our second guiding observation was that

898
00:35:01,190 --> 00:35:00,000
although in most of the contemporary

899
00:35:02,630 --> 00:35:01,200
heli cases

900
00:35:04,470 --> 00:35:02,640
the phosphate binding loop does not

901
00:35:07,030 --> 00:35:04,480
interact with the single standard dna

902
00:35:09,430 --> 00:35:07,040
and yet in the pdb we were able to find

903
00:35:10,630 --> 00:35:09,440
certain instances or certain vestiges

904
00:35:12,550 --> 00:35:10,640
where we see

905
00:35:14,470 --> 00:35:12,560
that the phosphate binding loop does

906
00:35:15,829 --> 00:35:14,480
interact with the

907
00:35:17,109 --> 00:35:15,839
the phosphate backbone of the single

908
00:35:20,069 --> 00:35:17,119
standard dna

909
00:35:21,670 --> 00:35:20,079
especially in xpd helicases so given

910
00:35:23,990 --> 00:35:21,680
like with both these observations we

911
00:35:26,310 --> 00:35:24,000
then wanted to test if the

912
00:35:28,630 --> 00:35:26,320
prototypes can unwind dna

913
00:35:30,550 --> 00:35:28,640

and to test our hypothesis we used an

914
00:35:32,470 --> 00:35:30,560
assay known as the molecular beacon

915
00:35:34,150 --> 00:35:32,480
assay where you have a double-stranded

916
00:35:35,670 --> 00:35:34,160
piece of dna

917
00:35:37,349 --> 00:35:35,680
the top strand of pitch has a

918
00:35:39,430 --> 00:35:37,359
fluorophore and a venture and opposite

919
00:35:41,750 --> 00:35:39,440
ends and if the dna strands were to be

920
00:35:43,750 --> 00:35:41,760
unwound it can form a beacon-like

921
00:35:45,109 --> 00:35:43,760
structure due to self-complementary ends

922
00:35:46,950 --> 00:35:45,119
and resulting in the loss of

923
00:35:49,750 --> 00:35:46,960
fluorescence

924
00:35:51,829 --> 00:35:49,760
and indeed the intact prototype as soon

925
00:35:53,349 --> 00:35:51,839
as you add it to a fluorescent dna as

926
00:35:56,390 --> 00:35:53,359
you can see here you see a drop in

927
00:35:57,990 --> 00:35:56,400
fluorescence that reaches the baseline

928
00:36:00,230 --> 00:35:58,000
in a two two-hour time scale and the

929
00:36:02,069 --> 00:36:00,240
baseline over here basically represents

930
00:36:03,430 --> 00:36:02,079
a completely twin state

931
00:36:05,829 --> 00:36:03,440
so it was great that the impact

932
00:36:07,510 --> 00:36:05,839
prototype mediates dna unwinding or

933
00:36:09,190 --> 00:36:07,520
stand separation

934
00:36:11,589 --> 00:36:09,200
but we wanted to see

935
00:36:13,270 --> 00:36:11,599
how small can we go while still

936
00:36:14,870 --> 00:36:13,280
retaining the function

937
00:36:17,109 --> 00:36:14,880
so here

938
00:36:19,270 --> 00:36:17,119

by a series of truncation and

939
00:36:20,950 --> 00:36:19,280
circular permutation we narrowed down or

940
00:36:23,109 --> 00:36:20,960
to shorten down the intact prototype

941
00:36:25,349 --> 00:36:23,119
from 110 amino acid to something which

942
00:36:27,030 --> 00:36:25,359
is less than 40 amino acid and this

943
00:36:28,950 --> 00:36:27,040
construct which we call as the n alpha

944
00:36:32,870 --> 00:36:28,960
beta alpha construct just has an alpha

945
00:36:35,589 --> 00:36:32,880
helix and the beta pulo pulpa motif

946
00:36:38,150 --> 00:36:35,599
so this an alpha beta alpha construct

947
00:36:39,990 --> 00:36:38,160
not only does it unwind dna it is the

948
00:36:42,230 --> 00:36:40,000
most efficient at dna unwinding as you

949
00:36:44,150 --> 00:36:42,240
can see by sharp dropping fluorescence

950
00:36:45,829 --> 00:36:44,160
indicating strand separation by the

951
00:36:47,109 --> 00:36:45,839
molecular beacon assay and it reaches

952
00:36:50,069 --> 00:36:47,119
the baseline

953
00:36:52,790 --> 00:36:50,079
so overall it suggests that the the

954
00:36:55,109 --> 00:36:52,800
basic beta palpa motif demonstrates

955
00:36:57,109 --> 00:36:55,119
significant structure plasticity in that

956
00:36:58,950 --> 00:36:57,119
you can put it in a variety of reduced

957
00:37:01,349 --> 00:36:58,960
complexity structural complexity

958
00:37:02,790 --> 00:37:01,359
scaffolds and it still not only retains

959
00:37:04,870 --> 00:37:02,800
the function but it can also show

960
00:37:06,950 --> 00:37:04,880
enhanced activity and this structural

961
00:37:09,910 --> 00:37:06,960
plasticity would have been crucial for

962
00:37:11,750 --> 00:37:09,920
primordial peptides to function

963
00:37:13,589 --> 00:37:11,760

so overall the helicase-like activity

964
00:37:15,109 --> 00:37:13,599
that i just showed you provides a

965
00:37:18,230 --> 00:37:15,119
plausible solution to the rna

966
00:37:20,550 --> 00:37:18,240
replication problem which is once the

967
00:37:22,069 --> 00:37:20,560
rna molecules have been replicated and

968
00:37:24,630 --> 00:37:22,079
once they have formed a double standard

969
00:37:26,470 --> 00:37:24,640
structure for them to unwind or for them

970
00:37:28,069 --> 00:37:26,480
to open up it requires an unwinding

971
00:37:29,990 --> 00:37:28,079
polypeptide for the second round of

972
00:37:32,470 --> 00:37:30,000
replication to occur and this is where

973
00:37:34,550 --> 00:37:32,480
the p-loop prototypes of proto-peptides

974
00:37:37,750 --> 00:37:34,560
like the ones which i've shown you would

975
00:37:40,870 --> 00:37:37,760
have provided a solution to this problem

976
00:37:43,430 --> 00:37:40,880
okay so i mentioned earlier uh that

977
00:37:45,349 --> 00:37:43,440
these fragments bind to ntps and single

978
00:37:47,910 --> 00:37:45,359
standard dna both by the phosphate

979
00:37:49,510 --> 00:37:47,920
binding loop if that is the case can we

980
00:37:51,510 --> 00:37:49,520
have some kind of an exchange between

981
00:37:54,150 --> 00:37:51,520
the two ligands

982
00:37:55,829 --> 00:37:54,160
and it turns out it we can so what you

983
00:37:57,430 --> 00:37:55,839
see over here is the same molecule as we

984
00:37:59,430 --> 00:37:57,440
can say where you see a decrease in

985
00:38:01,190 --> 00:37:59,440
fluorescence upon addition of protein

986
00:38:02,550 --> 00:38:01,200
and at this point

987
00:38:04,710 --> 00:38:02,560
when the dna molecules have been

988
00:38:07,670 --> 00:38:04,720

completely unbound if we add ligands

989
00:38:08,950 --> 00:38:07,680
like gtp and atp you see that the bound

990
00:38:11,030 --> 00:38:08,960
proteins release

991
00:38:13,349 --> 00:38:11,040
allowing the dna to revert back to its

992
00:38:15,109 --> 00:38:13,359
initial unwound state as you can see by

993
00:38:17,270 --> 00:38:15,119
increasing fluorescence

994
00:38:19,349 --> 00:38:17,280
therefore resembling some kind of a

995
00:38:21,109 --> 00:38:19,359
rudimentary helical cycle

996
00:38:22,710 --> 00:38:21,119
whereas modern day helicases what they

997
00:38:25,109 --> 00:38:22,720
do is what they would use the energy of

998
00:38:26,870 --> 00:38:25,119
atp hydrolysis unwind the dna and

999
00:38:28,870 --> 00:38:26,880
release from the dna so we see that

1000
00:38:32,150 --> 00:38:28,880
these prototypes also have some helicase

1001
00:38:33,750 --> 00:38:32,160
like activity or helicals like cycles

1002
00:38:36,150 --> 00:38:33,760
but what was the most interesting part

1003
00:38:38,470 --> 00:38:36,160
which i'm going to talk now is that

1004
00:38:39,990 --> 00:38:38,480
inorganic polyphosphates that is

1005
00:38:41,750 --> 00:38:40,000
long-chain polyphosphates and

1006
00:38:43,430 --> 00:38:41,760
hexamethylphosphate which is cyclic form

1007
00:38:45,990 --> 00:38:43,440
of phosphate was the most efficient in

1008
00:38:48,230 --> 00:38:46,000
releasing the proteins from the dna as

1009
00:38:50,870 --> 00:38:48,240
you can see here this 5.6 micromolar of

1010
00:38:53,589 --> 00:38:50,880
hexameter phosphate can release

1011
00:38:54,550 --> 00:38:53,599
almost 50 of the proteins bound to the

1012
00:38:57,510 --> 00:38:54,560
dna

1013
00:38:59,109 --> 00:38:57,520

whereas atp requires three point

1014
00:39:00,550 --> 00:38:59,119
almost three millimolar concentration to

1015
00:39:03,910 --> 00:39:00,560
have the same effect

1016
00:39:05,829 --> 00:39:03,920
so that these primordial proteins bind

1017
00:39:07,750 --> 00:39:05,839
favorably to inorganic polyphosphate

1018
00:39:09,589 --> 00:39:07,760
which i have also been proposed to be

1019
00:39:11,109 --> 00:39:09,599
the ancient precursor

1020
00:39:13,510 --> 00:39:11,119
of ntps

1021
00:39:15,829 --> 00:39:13,520
we can say that the mode of action of

1022
00:39:18,390 --> 00:39:15,839
these prototypes is quite tailored to

1023
00:39:20,870 --> 00:39:18,400
the needs of the primordial world

1024
00:39:23,670 --> 00:39:20,880
so basically now you you can ask me that

1025
00:39:25,990 --> 00:39:23,680
how can such a short fragment

1026
00:39:27,990 --> 00:39:26,000
demonstrate such complex function and i

1027
00:39:30,310 --> 00:39:28,000
think and and we know that the key to

1028
00:39:32,470 --> 00:39:30,320
function is that the ability of these

1029
00:39:33,670 --> 00:39:32,480
short proteins to oligomerize or to

1030
00:39:36,390 --> 00:39:33,680
self-assemble

1031
00:39:39,510 --> 00:39:36,400
by native mass spec we have shown that

1032
00:39:42,630 --> 00:39:39,520
the n-alpha beta alpha peptide can form

1033
00:39:44,950 --> 00:39:42,640
large oligomers up to 30 more complexes

1034
00:39:46,550 --> 00:39:44,960
and this is the key for it to function

1035
00:39:48,630 --> 00:39:46,560
otherwise a short peptide cannot

1036
00:39:50,069 --> 00:39:48,640
function by itself in a solvent exposed

1037
00:39:51,589 --> 00:39:50,079
group

1038
00:39:52,950 --> 00:39:51,599

so to summarize

1039
00:39:54,630 --> 00:39:52,960
uh the ancient p loop was a

1040
00:39:56,230 --> 00:39:54,640
multifunctional p loop which that one

1041
00:39:59,510 --> 00:39:56,240
which had to do multiple functions such

1042
00:40:00,630 --> 00:39:59,520
as dna binding single uh ntp binding dna

1043
00:40:02,630 --> 00:40:00,640
unwinding

1044
00:40:05,030 --> 00:40:02,640
and such multi-functional prototypes

1045
00:40:07,109 --> 00:40:05,040
then underwent self-assembly duplication

1046
00:40:09,109 --> 00:40:07,119
and fusion to give rise to modern day

1047
00:40:12,150 --> 00:40:09,119
proteins which had specialized domains

1048
00:40:13,829 --> 00:40:12,160
that carry out specialized functions

1049
00:40:16,309 --> 00:40:13,839
and to end i would

1050
00:40:18,069 --> 00:40:16,319
say that these fragments these p-loop

1051
00:40:19,990 --> 00:40:18,079
prototypes

1052
00:40:21,910 --> 00:40:20,000
satisfy the basic postulates regarding

1053
00:40:24,150 --> 00:40:21,920
the emergence of earliest proteins in

1054
00:40:25,349 --> 00:40:24,160
that they are relatively short the

1055
00:40:27,430 --> 00:40:25,359
compose of

1056
00:40:30,309 --> 00:40:27,440
almost a minimal abiotic amino acid

1057
00:40:31,910 --> 00:40:30,319
alphabet these prototypes have a lysine

1058
00:40:33,270 --> 00:40:31,920
and i is tagged but we know that if you

1059
00:40:35,589 --> 00:40:33,280
remove the haystack and even if you

1060
00:40:37,270 --> 00:40:35,599
mutate the lysine with a glycine they

1061
00:40:39,829 --> 00:40:37,280
still retain function and they are

1062
00:40:42,069 --> 00:40:39,839
incredibly tolerant to mutations

1063
00:40:44,550 --> 00:40:42,079

and the last type the last postulate is

1064
00:40:47,430 --> 00:40:44,560
that they tend to self-assemble which

1065
00:40:49,109 --> 00:40:47,440
allows them to form a larger structural

1066
00:40:51,990 --> 00:40:49,119
you know configuration

1067
00:40:54,390 --> 00:40:52,000
that is crucial for function

1068
00:40:56,390 --> 00:40:54,400
so to conclude i would say that the p

1069
00:40:58,950 --> 00:40:56,400
loop prototype despite the simplicity

1070
00:41:00,550 --> 00:40:58,960
they relate to contemporary p loop n tps

1071
00:41:03,510 --> 00:41:00,560
in terms of their sequence structure and

1072
00:41:05,109 --> 00:41:03,520
function and that they serve as starting

1073
00:41:07,910 --> 00:41:05,119
points or evolutionary starting points

1074
00:41:09,990 --> 00:41:07,920
for enzymes with more complex activity

1075
00:41:12,309 --> 00:41:10,000
and it is only app that i end the

1076
00:41:14,390 --> 00:41:12,319
presentation by this quote from darwin

1077
00:41:16,309 --> 00:41:14,400
which was also one of dhani's favorite

1078
00:41:17,270 --> 00:41:16,319
quote is that from so simple the

1079
00:41:18,950 --> 00:41:17,280
beginning

1080
00:41:20,550 --> 00:41:18,960
endless forms most beautiful and most

1081
00:41:21,829 --> 00:41:20,560
wonderful have been and are being

1082
00:41:23,430 --> 00:41:21,839
involved

1083
00:41:25,910 --> 00:41:23,440
and

1084
00:41:26,790 --> 00:41:25,920
i would like to thank the people from my

1085
00:41:28,630 --> 00:41:26,800
lab

1086
00:41:30,550 --> 00:41:28,640
i would like to thank sarah fleischmann

1087
00:41:32,470 --> 00:41:30,560
who is my new supervisor

1088
00:41:33,990 --> 00:41:32,480

stephen and the organized organizing

1089
00:41:35,109 --> 00:41:34,000
committee for giving me this opportunity

1090
00:41:36,630 --> 00:41:35,119
once again

1091
00:41:38,790 --> 00:41:36,640
and the volkswagen foundation and the

1092
00:41:40,309 --> 00:41:38,800
weizmann institute for the generous

1093
00:41:48,069 --> 00:41:40,319
funding

1094
00:41:51,589 --> 00:41:50,150
you pratik for stimulating talk and in

1095
00:41:53,430 --> 00:41:51,599
the interest of time we're going to

1096
00:41:55,030 --> 00:41:53,440
suppress questions but we do have extra

1097
00:41:57,430 --> 00:41:55,040
time at the end for questions i'm sure

1098
00:42:00,069 --> 00:41:57,440
there will be many i'm going to

1099
00:42:02,710 --> 00:42:00,079
go ahead and introduce our next speaker

1100
00:42:05,670 --> 00:42:02,720
who is claudia alvarez who is a

1101
00:42:06,790 --> 00:42:05,680
postdoctoral scholar in the laboratory

1102
00:42:12,069 --> 00:42:06,800
of

1103
00:42:16,550 --> 00:42:13,829
thank you

1104
00:42:17,829 --> 00:42:16,560
i'm going to talk about protein fold

1105
00:42:18,630 --> 00:42:17,839
evolution

1106
00:42:19,589 --> 00:42:18,640
or

1107
00:42:24,309 --> 00:42:19,599
how

1108
00:42:29,349 --> 00:42:26,870
so in this work we wanted to understand

1109
00:42:31,670 --> 00:42:29,359
the evolutionary mechanisms that led to

1110
00:42:33,750 --> 00:42:31,680
the diversity of protein falls in

1111
00:42:36,950 --> 00:42:33,760
contemporary biology

1112
00:42:40,470 --> 00:42:36,960
so for example in a human cell or in a

1113
00:42:41,750 --> 00:42:40,480

human proteome there are around 20 000

1114
00:42:47,510 --> 00:42:41,760
proteins

1115
00:42:50,309 --> 00:42:47,520
000 unique units

1116
00:42:53,030 --> 00:42:50,319
so 1000 is a very small number when

1117
00:42:57,109 --> 00:42:53,040
compared to the total number of proteins

1118
00:42:58,790 --> 00:42:57,119
that are present in a single human cell

1119
00:43:00,390 --> 00:42:58,800
and i think it's also a very small

1120
00:43:03,510 --> 00:43:00,400
number when you think

1121
00:43:07,670 --> 00:43:03,520
that these are the product of 3.8

1122
00:43:12,870 --> 00:43:10,309
but we can see the same question with a

1123
00:43:14,069 --> 00:43:12,880
different perspective

1124
00:43:16,069 --> 00:43:14,079
so the

1125
00:43:18,430 --> 00:43:16,079
emergence of

1126
00:43:20,630 --> 00:43:18,440
folding competent sequences is a

1127
00:43:23,589 --> 00:43:20,640
multi-layer problem

1128
00:43:26,710 --> 00:43:23,599
so first we have the problem of the

1129
00:43:28,950 --> 00:43:26,720
amino acid sequences being very

1130
00:43:31,910 --> 00:43:28,960
there are many combinations

1131
00:43:34,630 --> 00:43:31,920
so for the extant genetic code there are

1132
00:43:36,950 --> 00:43:34,640
far more possible amino acid sequences

1133
00:43:38,309 --> 00:43:36,960
than there are stars in the universe

1134
00:43:41,270 --> 00:43:38,319
actually for

1135
00:43:43,990 --> 00:43:41,280
a sequence of 100 residues there are

1136
00:43:46,950 --> 00:43:44,000
more combinations that are possible than

1137
00:43:48,309 --> 00:43:46,960
atoms in the universe so

1138
00:43:51,190 --> 00:43:48,319

um

1139
00:43:54,069 --> 00:43:51,200
it's not it's unlikely that all of these

1140
00:43:55,910 --> 00:43:54,079
sequences can be sampled

1141
00:43:58,069 --> 00:43:55,920
the next problem is that not all

1142
00:44:00,950 --> 00:43:58,079
combinations will result in a stable

1143
00:44:04,309 --> 00:44:00,960
fold and then when you finally find a

1144
00:44:05,990 --> 00:44:04,319
combination that can fold stably

1145
00:44:07,670 --> 00:44:06,000
it's not

1146
00:44:08,470 --> 00:44:07,680
like you can

1147
00:44:11,670 --> 00:44:08,480
move

1148
00:44:13,270 --> 00:44:11,680
from fold to fall just by simply

1149
00:44:14,870 --> 00:44:13,280
modifying

1150
00:44:16,630 --> 00:44:14,880
the sequence

1151
00:44:20,309 --> 00:44:16,640
step by step

1152
00:44:23,349 --> 00:44:20,319
so there are not many examples of

1153
00:44:25,430 --> 00:44:23,359
sequences that can transition from one

1154
00:44:28,950 --> 00:44:25,440
fold to the other

1155
00:44:32,390 --> 00:44:28,960
but what we do find is many examples of

1156
00:44:35,030 --> 00:44:32,400
sequences that share similarity between

1157
00:44:37,910 --> 00:44:35,040
different folds the similarity in this

1158
00:44:40,950 --> 00:44:37,920
case is not overall in the entire

1159
00:44:42,950 --> 00:44:40,960
sequence but just a small fragment

1160
00:44:45,990 --> 00:44:42,960
these are called crossfall sequence

1161
00:44:49,109 --> 00:44:46,000
similarities and they suggest fault

1162
00:44:51,190 --> 00:44:49,119
evolution so once you find this

1163
00:44:53,510 --> 00:44:51,200

crossfall sequence similarities you can

1164
00:44:55,589 --> 00:44:53,520
assume there's a evolutionary history

1165
00:44:58,950 --> 00:44:55,599
that is shared but you still don't

1166
00:45:02,069 --> 00:44:58,960
understand how these came to be and we

1167
00:45:04,790 --> 00:45:02,079
wanted to know the step-by-step process

1168
00:45:07,750 --> 00:45:04,800
of how this happened so we started

1169
00:45:10,790 --> 00:45:07,760
looking at examples we we thought do we

1170
00:45:14,390 --> 00:45:10,800
really know of a case of false evolution

1171
00:45:16,670 --> 00:45:14,400
that we completely understand

1172
00:45:18,390 --> 00:45:16,680
and it turns out that there is a

1173
00:45:20,550 --> 00:45:18,400
paradigmatic case

1174
00:45:23,030 --> 00:45:20,560
that is circular permutation

1175
00:45:26,150 --> 00:45:23,040
so circular permutation is a

1176
00:45:27,829 --> 00:45:26,160
relationship between two proteins

1177
00:45:29,430 --> 00:45:27,839
or two topologies

1178
00:45:31,670 --> 00:45:29,440
that have a very similar

1179
00:45:33,430 --> 00:45:31,680
three-dimensional structure but the

1180
00:45:35,510 --> 00:45:33,440
secondary structural elements are

1181
00:45:39,829 --> 00:45:35,520
rearranged

1182
00:45:42,870 --> 00:45:39,839
so how do you get from fold a to fold b

1183
00:45:45,910 --> 00:45:42,880
simply by circularizing the

1184
00:45:48,230 --> 00:45:45,920
fold a and then you can cleave at

1185
00:45:49,829 --> 00:45:48,240
whichever point in the

1186
00:45:52,710 --> 00:45:49,839
protein structure

1187
00:45:55,510 --> 00:45:52,720
you will get the circular permutant of

1188
00:45:57,510 --> 00:45:55,520

the fold a

1189
00:45:59,829 --> 00:45:57,520
but this is not what happens in

1190
00:46:02,470 --> 00:45:59,839
evolution so there are many examples of

1191
00:46:04,230 --> 00:46:02,480
circular permutation but the mechanism

1192
00:46:06,470 --> 00:46:04,240
is not this

1193
00:46:10,470 --> 00:46:06,480
what happens in evolution is that you

1194
00:46:13,030 --> 00:46:10,480
get one gene that is duplicated in line

1195
00:46:16,390 --> 00:46:13,040
so usually a duplication of a domain

1196
00:46:18,390 --> 00:46:16,400
gives a repeat of the same fold so you

1197
00:46:21,270 --> 00:46:18,400
have the same fold twice in a single

1198
00:46:23,510 --> 00:46:21,280
protein but when you have circular

1199
00:46:26,550 --> 00:46:23,520
permutation this is not what happens the

1200
00:46:27,750 --> 00:46:26,560
duplication opens a new folding

1201
00:46:30,230 --> 00:46:27,760
landscape

1202
00:46:33,109 --> 00:46:30,240
for this protein and then a new fold

1203
00:46:35,030 --> 00:46:33,119
emerges and this new fold will have some

1204
00:46:37,109 --> 00:46:35,040
secondary structural elements from the

1205
00:46:39,109 --> 00:46:37,119
first copy of the repeat and some

1206
00:46:40,710 --> 00:46:39,119
secondary elements from the second copy

1207
00:46:43,109 --> 00:46:40,720
of the repeat

1208
00:46:46,069 --> 00:46:43,119
so the last step in the maturation of

1209
00:46:48,150 --> 00:46:46,079
the circular permutant is the loss of

1210
00:46:50,309 --> 00:46:48,160
the terminal segment

1211
00:46:52,309 --> 00:46:50,319
and that way you have a daughter fold

1212
00:46:53,910 --> 00:46:52,319
that is very similar to the ancestral

1213
00:46:56,790 --> 00:46:53,920

fold but the secondary structural

1214
00:46:59,030 --> 00:46:56,800
elements are rearranged

1215
00:47:01,190 --> 00:46:59,040
so what do we learn from the study of

1216
00:47:03,750 --> 00:47:01,200
circular permutation well we learned

1217
00:47:04,790 --> 00:47:03,760
that if we take many homologs to the

1218
00:47:07,750 --> 00:47:04,800
first

1219
00:47:09,349 --> 00:47:07,760
ancestral copy

1220
00:47:11,750 --> 00:47:09,359
sometimes we will have

1221
00:47:13,990 --> 00:47:11,760
sequences that are more similar to the

1222
00:47:16,230 --> 00:47:14,000
n-terminus of the daughter fold and

1223
00:47:18,710 --> 00:47:16,240
sometimes we will have

1224
00:47:22,230 --> 00:47:18,720
other sequences that are more similar to

1225
00:47:25,829 --> 00:47:22,240
the second health of the dodger fold

1226
00:47:28,150 --> 00:47:25,839
so if we sample a long enough uh list of

1227
00:47:30,790 --> 00:47:28,160
sequences and then we align them we will

1228
00:47:33,030 --> 00:47:30,800
get a pattern of cross-fold sequence

1229
00:47:35,430 --> 00:47:33,040
similarities that will look

1230
00:47:37,589 --> 00:47:35,440
like these

1231
00:47:40,069 --> 00:47:37,599
and this is what we want to look for so

1232
00:47:42,870 --> 00:47:40,079
now we have a strategy we know what we

1233
00:47:44,549 --> 00:47:42,880
want to look for and we can interpret

1234
00:47:45,589 --> 00:47:44,559
this pattern

1235
00:47:48,390 --> 00:47:45,599
next

1236
00:47:51,430 --> 00:47:48,400
where do we start and we started by one

1237
00:47:53,670 --> 00:47:51,440
of the ribosomal proteins of course

1238
00:47:55,990 --> 00:47:53,680

this is universal ribosomal protein two

1239
00:47:57,589 --> 00:47:56,000
and this is very interesting

1240
00:47:59,510 --> 00:47:57,599
this is a very interesting protein

1241
00:48:01,990 --> 00:47:59,520
because it's one of the few universal

1242
00:48:04,549 --> 00:48:02,000
ribosomal proteins that has more than

1243
00:48:07,109 --> 00:48:04,559
one domain it's a multi-domain protein

1244
00:48:10,470 --> 00:48:07,119
the two domains in ul2 are distinct

1245
00:48:12,549 --> 00:48:10,480
these are called sh3 and oe

1246
00:48:14,870 --> 00:48:12,559
and these two folds are present

1247
00:48:17,349 --> 00:48:14,880
everywhere in the translation machinery

1248
00:48:20,069 --> 00:48:17,359
so from other ribosomal proteins to

1249
00:48:23,430 --> 00:48:20,079
amino acid rna synthetases and

1250
00:48:25,829 --> 00:48:23,440
initiation and elongation factors

1251
00:48:28,309 --> 00:48:25,839
so we took ul2

1252
00:48:30,950 --> 00:48:28,319
build multiple sequence alignments

1253
00:48:32,069 --> 00:48:30,960
search the evolutionary classification

1254
00:48:33,430 --> 00:48:32,079
of domains

1255
00:48:35,109 --> 00:48:33,440
for

1256
00:48:38,710 --> 00:48:35,119
sequence similarities and we were

1257
00:48:41,030 --> 00:48:38,720
looking for this characteristic pattern

1258
00:48:44,069 --> 00:48:41,040
so these are the results of our search

1259
00:48:46,470 --> 00:48:44,079
for crossfall sequence similarities in

1260
00:48:49,270 --> 00:48:46,480
the orange squares i'm showing you the

1261
00:48:51,430 --> 00:48:49,280
region where we would expect to see um

1262
00:48:53,990 --> 00:48:51,440
these crossfall sequence similarities

1263
00:48:55,349 --> 00:48:54,000

and you can see that i have divided the

1264
00:48:57,990 --> 00:48:55,359
results into

1265
00:48:59,270 --> 00:48:58,000
different panels

1266
00:49:00,790 --> 00:48:59,280
so the first

1267
00:49:03,670 --> 00:49:00,800
one shows the crossfull sequence

1268
00:49:06,390 --> 00:49:03,680
similarities between ob and sh3 and the

1269
00:49:07,990 --> 00:49:06,400
second one between ob and cradle loop

1270
00:49:09,910 --> 00:49:08,000
barrels

1271
00:49:11,430 --> 00:49:09,920
so these folds are

1272
00:49:13,670 --> 00:49:11,440
in the field of

1273
00:49:16,470 --> 00:49:13,680
protein fold evolution like rock stars

1274
00:49:18,710 --> 00:49:16,480
of the protein fold evolution this has

1275
00:49:19,430 --> 00:49:18,720
been this have been very well studied

1276
00:49:22,950 --> 00:49:19,440
and

1277
00:49:27,589 --> 00:49:25,270
have been very interesting

1278
00:49:30,630 --> 00:49:27,599
for this field

1279
00:49:33,750 --> 00:49:30,640
so for the first one sh3 and ob

1280
00:49:35,190 --> 00:49:33,760
i am showing you here in color

1281
00:49:37,270 --> 00:49:35,200
i have mapped

1282
00:49:41,190 --> 00:49:37,280
the region across fault sequence

1283
00:49:44,309 --> 00:49:41,200
similarity into 3d and into 1d

1284
00:49:47,589 --> 00:49:44,319
representations so on the left we have

1285
00:49:50,230 --> 00:49:47,599
one pair of sh3 and ob that share one

1286
00:49:52,150 --> 00:49:50,240
region in this region we can see that

1287
00:49:54,870 --> 00:49:52,160
the cross fall sequence similarity also

1288
00:49:57,430 --> 00:49:54,880

corresponds to a very similar structure

1289
00:49:59,109 --> 00:49:57,440
and then for the pair on the right

1290
00:50:01,589 --> 00:49:59,119
the region of crossfall sequence

1291
00:50:04,549 --> 00:50:01,599
similarity is also similar in structure

1292
00:50:08,549 --> 00:50:04,559
but there is a variation

1293
00:50:10,790 --> 00:50:08,559
there's a different turn between them

1294
00:50:12,630 --> 00:50:10,800
and now the next thing that we can do

1295
00:50:15,589 --> 00:50:12,640
because these two ob folds are

1296
00:50:18,150 --> 00:50:15,599
homologous we can align one to the other

1297
00:50:20,470 --> 00:50:18,160
and then bring their respective sh3

1298
00:50:23,349 --> 00:50:20,480
pairs to the alignment and when we do

1299
00:50:25,349 --> 00:50:23,359
that we find the characteristic pattern

1300
00:50:27,430 --> 00:50:25,359
that is similar to to the circular

1301
00:50:30,309 --> 00:50:27,440
permutation case

1302
00:50:32,710 --> 00:50:30,319
now if we study the case of obn cradle

1303
00:50:35,510 --> 00:50:32,720
loop barrel we have the same we have one

1304
00:50:37,430 --> 00:50:35,520
pair on the left that has one region of

1305
00:50:40,069 --> 00:50:37,440
cross-full sequence similarity and then

1306
00:50:42,630 --> 00:50:40,079
one pair on the right that shows a

1307
00:50:45,510 --> 00:50:42,640
different region so we did exactly the

1308
00:50:47,510 --> 00:50:45,520
same we aligned these two cradle loop

1309
00:50:49,829 --> 00:50:47,520
parallels together

1310
00:50:52,549 --> 00:50:49,839
and then brought the obs

1311
00:50:53,510 --> 00:50:52,559
and this is the pattern that we observe

1312
00:50:57,109 --> 00:50:53,520
so

1313
00:50:58,950 --> 00:50:57,119

what do we think happened

1314
00:51:01,510 --> 00:50:58,960
so what we think that

1315
00:51:06,390 --> 00:51:01,520
can be said about these relationships is

1316
00:51:09,190 --> 00:51:06,400
that possibly one ob fold ancestor

1317
00:51:12,549 --> 00:51:09,200
duplicated so usually you would get a

1318
00:51:14,790 --> 00:51:12,559
repeat of the ob fold but in this case

1319
00:51:17,589 --> 00:51:14,800
the repeat didn't give rise to our

1320
00:51:20,829 --> 00:51:17,599
repeat of the structure we got a new

1321
00:51:24,309 --> 00:51:20,839
fold with new hydrogen bonds and new

1322
00:51:27,349 --> 00:51:24,319
interactions so these fold matured and

1323
00:51:29,750 --> 00:51:27,359
transformed into what we know now as a

1324
00:51:32,390 --> 00:51:29,760
cradle loop barrel so in this crayola

1325
00:51:34,710 --> 00:51:32,400
barrel we have some secondary structural

1326
00:51:36,870 --> 00:51:34,720
elements and motifs that are very

1327
00:51:41,030 --> 00:51:36,880
similar to the ancestor

1328
00:51:43,430 --> 00:51:41,040
and some others that are new

1329
00:51:44,870 --> 00:51:43,440
what we can say is that relationships

1330
00:51:48,069 --> 00:51:44,880
between ob

1331
00:51:50,630 --> 00:51:48,079
sh3 and cradle loop barrels illustrate a

1332
00:51:53,190 --> 00:51:50,640
process that generates new fall

1333
00:51:55,589 --> 00:51:53,200
topologies from within

1334
00:51:58,390 --> 00:51:55,599
and we would say incessantly destroying

1335
00:51:59,990 --> 00:51:58,400
the old one incessantly creating a new

1336
00:52:03,910 --> 00:52:00,000
one so

1337
00:52:05,910 --> 00:52:03,920
here for example two sh3's form one of b

1338
00:52:08,790 --> 00:52:05,920

two o these form one crayola loop

1339
00:52:13,589 --> 00:52:11,349
so we called this process creative

1340
00:52:16,710 --> 00:52:13,599
destruction and this is the idea that

1341
00:52:19,750 --> 00:52:16,720
once you have one fold you can create

1342
00:52:22,870 --> 00:52:19,760
many from that one so maybe you don't

1343
00:52:25,750 --> 00:52:22,880
need to create many faults many times

1344
00:52:27,910 --> 00:52:25,760
you just need to create one and then you

1345
00:52:30,470 --> 00:52:27,920
can generate many

1346
00:52:33,349 --> 00:52:30,480
so creative destructions acts on the

1347
00:52:36,230 --> 00:52:33,359
level of domains depends on false

1348
00:52:38,790 --> 00:52:36,240
plasticity and resolves crossfall

1349
00:52:41,430 --> 00:52:38,800
similarities by a biologically plausible

1350
00:52:43,990 --> 00:52:41,440
mechanism suggesting that the universe

1351
00:52:46,549 --> 00:52:44,000
of protein folds is better described as

1352
00:52:48,790 --> 00:52:46,559
a network than as a tree

1353
00:52:49,829 --> 00:52:48,800
so i want to thank everyone in this

1354
00:52:53,190 --> 00:52:49,839
slide

1355
00:52:54,540 --> 00:52:53,200
and we have a preprint for more details

1356
00:53:09,190 --> 00:52:54,550
thank you

1357
00:53:14,790 --> 00:53:11,670
hey anthony brunetti here also from

1358
00:53:16,710 --> 00:53:14,800
georgia tech and i was uh wondering so

1359
00:53:19,190 --> 00:53:16,720
this so this

1360
00:53:22,390 --> 00:53:19,200
work that you showed is looking at these

1361
00:53:25,190 --> 00:53:22,400
incredibly ancient incredibly deeply

1362
00:53:27,270 --> 00:53:25,200
important uh really common uh folds and

1363
00:53:29,510 --> 00:53:27,280

things i was wondering

1364
00:53:31,349 --> 00:53:29,520
could another way of looking at this be

1365
00:53:32,470 --> 00:53:31,359
trying to find

1366
00:53:33,510 --> 00:53:32,480
uh

1367
00:53:36,470 --> 00:53:33,520
newer

1368
00:53:39,109 --> 00:53:36,480
proteins because there are new proteins

1369
00:53:40,870 --> 00:53:39,119
being generated especially in like giant

1370
00:53:42,470 --> 00:53:40,880
virus genomes and things like that and i

1371
00:53:44,230 --> 00:53:42,480
wonder if

1372
00:53:46,870 --> 00:53:44,240
even though those aren't extraordinarily

1373
00:53:48,470 --> 00:53:46,880
well established or i'm extraordinarily

1374
00:53:50,470 --> 00:53:48,480
well understood i wonder if that might

1375
00:53:53,349 --> 00:53:50,480
be a place to see

1376
00:53:56,950 --> 00:53:53,359
rapid rates of this happening if this if

1377
00:53:58,870 --> 00:53:56,960
this is uh going on there yeah so this

1378
00:54:02,630 --> 00:53:58,880
would be a process that can auto

1379
00:54:05,430 --> 00:54:02,640
propagate and this example here

1380
00:54:06,710 --> 00:54:05,440
is actually of a protein that is present

1381
00:54:07,990 --> 00:54:06,720
in

1382
00:54:11,349 --> 00:54:08,000
humans

1383
00:54:14,069 --> 00:54:11,359
so this pdb code comes from an

1384
00:54:16,470 --> 00:54:14,079
sequence that very recently suffered

1385
00:54:22,590 --> 00:54:16,480
this creative destruction

1386
00:54:22,600 --> 00:54:37,030
[Applause]

1387
00:54:41,589 --> 00:54:39,670
wonderful well thank you for joining us

1388
00:54:44,309 --> 00:54:41,599

for this session

1389
00:54:46,870 --> 00:54:44,319
um i'm going to join i'm going to

1390
00:54:48,309 --> 00:54:46,880
begin by acknowledging the individuals

1391
00:54:50,150 --> 00:54:48,319
and organizations that have made this

1392
00:54:51,349 --> 00:54:50,160
research possible

1393
00:54:53,990 --> 00:54:51,359
i'm going to be talking about the work

1394
00:54:55,670 --> 00:54:54,000
of three of my uh students to our

1395
00:54:57,190 --> 00:54:55,680
graduate students philip toe and haley

1396
00:54:59,910 --> 00:54:57,200
moran and one is a very talented

1397
00:55:02,870 --> 00:54:59,920
undergraduate atarva bhagwat and we've

1398
00:55:04,309 --> 00:55:02,880
received support from hfsp the nih and

1399
00:55:06,630 --> 00:55:04,319
the nsf

1400
00:55:08,710 --> 00:55:06,640
so let me start with a question

1401
00:55:11,349 --> 00:55:08,720
how do we know which proteins are the

1402
00:55:13,190 --> 00:55:11,359
most ancient well we can do our best to

1403
00:55:15,349 --> 00:55:13,200
try to answer this difficult and tangled

1404
00:55:17,349 --> 00:55:15,359
up question one way is we can sort of

1405
00:55:19,030 --> 00:55:17,359
infer that proteins are probably ancient

1406
00:55:21,349 --> 00:55:19,040
if they're extremely important and if we

1407
00:55:23,670 --> 00:55:21,359
can infer their presence in some

1408
00:55:25,190 --> 00:55:23,680
primordial organisms such as luca but

1409
00:55:27,190 --> 00:55:25,200
the problem of this is that of course a

1410
00:55:29,109 --> 00:55:27,200
lot of protein evolution occurred before

1411
00:55:30,870 --> 00:55:29,119
luca especially of the most fundamental

1412
00:55:32,230 --> 00:55:30,880
domains such as the ones that claudia

1413
00:55:34,069 --> 00:55:32,240

was speaking about

1414
00:55:35,589 --> 00:55:34,079
we can also try to address this question

1415
00:55:38,309 --> 00:55:35,599
by looking at the phylogeny or the

1416
00:55:40,470 --> 00:55:38,319
distribution by creating the trees but

1417
00:55:42,309 --> 00:55:40,480
the problem is that this trees can de

1418
00:55:44,549 --> 00:55:42,319
expansion events whereby rapid

1419
00:55:47,430 --> 00:55:44,559
diversification of a certain fold class

1420
00:55:48,789 --> 00:55:47,440
can decouple the distribution and the

1421
00:55:51,030 --> 00:55:48,799
actual

1422
00:55:51,990 --> 00:55:51,040
order of incorporation into the protein

1423
00:55:53,750 --> 00:55:52,000
universe

1424
00:55:56,710 --> 00:55:53,760
so i'll sort of repeat the question can

1425
00:55:59,589 --> 00:55:56,720
any experimentally observable property

1426
00:56:02,230 --> 00:55:59,599
of a protein speak to the antiquity of

1427
00:56:04,230 --> 00:56:02,240
its provenance and we think the answer

1428
00:56:06,630 --> 00:56:04,240
to that question is yes and it's it's

1429
00:56:08,470 --> 00:56:06,640
refoldability

1430
00:56:10,390 --> 00:56:08,480
so um my background is as a

1431
00:56:12,069 --> 00:56:10,400
protein-folding biophysicist and so we

1432
00:56:14,309 --> 00:56:12,079
think a lot about this remarkable

1433
00:56:16,549 --> 00:56:14,319
property of proteins whereby they can

1434
00:56:18,710 --> 00:56:16,559
spontaneously self-assemble into complex

1435
00:56:20,230 --> 00:56:18,720
three-dimensional architectures and this

1436
00:56:22,390 --> 00:56:20,240
is a property that is frequently

1437
00:56:23,990 --> 00:56:22,400
explored by biophysicists through

1438
00:56:25,510 --> 00:56:24,000

experiments in which proteins are

1439
00:56:27,349 --> 00:56:25,520
unfolded either by increasing

1440
00:56:30,150 --> 00:56:27,359
temperature or adding chaotropes like

1441
00:56:32,309 --> 00:56:30,160
urea or guanidine and then removing that

1442
00:56:34,870 --> 00:56:32,319
condition to return to the physiological

1443
00:56:36,950 --> 00:56:34,880
conditions under which some particular

1444
00:56:39,430 --> 00:56:36,960
proteins have this capacity to return to

1445
00:56:42,390 --> 00:56:39,440
their native fold unassisted so we call

1446
00:56:44,390 --> 00:56:42,400
this property of a protein refoldability

1447
00:56:46,309 --> 00:56:44,400
now of course the physical basis by

1448
00:56:48,470 --> 00:56:46,319
which this is generally explained is by

1449
00:56:50,470 --> 00:56:48,480
positing this so-called free energy

1450
00:56:53,349 --> 00:56:50,480
landscape in which we hypothesize that

1451
00:56:55,190 --> 00:56:53,359
native states reflect um thermodynamic

1452
00:56:56,870 --> 00:56:55,200
minima so that is the conformation that

1453
00:56:59,510 --> 00:56:56,880
lowers the gibbs free energy of the

1454
00:57:01,670 --> 00:56:59,520
system and if you um can posit that then

1455
00:57:03,190 --> 00:57:01,680
it's easy to imagine why you could do

1456
00:57:05,190 --> 00:57:03,200
whatever you want to this protein and

1457
00:57:07,510 --> 00:57:05,200
it's going to find safe passage back

1458
00:57:09,030 --> 00:57:07,520
home to its native fold because that's

1459
00:57:10,549 --> 00:57:09,040
basically what thermodynamics says it

1460
00:57:11,990 --> 00:57:10,559
has to do

1461
00:57:13,190 --> 00:57:12,000
but it's worth pointing out that even

1462
00:57:15,190 --> 00:57:13,200
though this is a property that we

1463
00:57:18,150 --> 00:57:15,200

frequently study it's by no means

1464
00:57:19,190 --> 00:57:18,160
universal it's basically a property of

1465
00:57:20,870 --> 00:57:19,200
small

1466
00:57:23,030 --> 00:57:20,880
single domain proteins the type that

1467
00:57:24,309 --> 00:57:23,040
biophysicists like to study but there

1468
00:57:26,069 --> 00:57:24,319
are lots of proteins that are

1469
00:57:27,750 --> 00:57:26,079
extraordinarily important for biology

1470
00:57:29,349 --> 00:57:27,760
that are complicated that involve lots

1471
00:57:31,510 --> 00:57:29,359
of moving parts that are embedded in

1472
00:57:33,349 --> 00:57:31,520
mechanisms and they use all sorts of

1473
00:57:35,670 --> 00:57:33,359
other machineries like chaperones in

1474
00:57:37,990 --> 00:57:35,680
order to be able to assemble so our

1475
00:57:40,710 --> 00:57:38,000
hypothesis is that by looking at what

1476
00:57:42,950 --> 00:57:40,720
classes of proteins are capable of

1477
00:57:45,829 --> 00:57:42,960
refolding themselves autonomously we're

1478
00:57:48,069 --> 00:57:45,839
in essence asking a biophysical basis of

1479
00:57:50,069 --> 00:57:48,079
antiquity because we don't think that

1480
00:57:52,230 --> 00:57:50,079
during the origin of life a complex

1481
00:57:54,549 --> 00:57:52,240
chaperone network or quality control was

1482
00:57:57,109 --> 00:57:54,559
available in essence the only quality

1483
00:57:59,670 --> 00:57:57,119
control that was available for or 4.2

1484
00:58:01,349 --> 00:57:59,680
billion years ago was thermodynamics and

1485
00:58:04,390 --> 00:58:01,359
so as a consequence the intrinsic

1486
00:58:05,910 --> 00:58:04,400
refoldability of a protein is a bit of a

1487
00:58:08,309 --> 00:58:05,920
way of thinking about which ones were

1488
00:58:10,789 --> 00:58:08,319

probably easier to access before we have

1489
00:58:12,150 --> 00:58:10,799
more complex metabolism

1490
00:58:13,750 --> 00:58:12,160
in this note i'll point out that of

1491
00:58:16,069 --> 00:58:13,760
course one property of proteins that

1492
00:58:17,750 --> 00:58:16,079
makes makes it very different than rna

1493
00:58:19,270 --> 00:58:17,760
is that protein folding has a puzzle

1494
00:58:21,750 --> 00:58:19,280
like quality in which there's only

1495
00:58:23,990 --> 00:58:21,760
really one or small number of possible

1496
00:58:26,230 --> 00:58:24,000
solutions to minimize the energy which

1497
00:58:28,069 --> 00:58:26,240
is very different than rna where there

1498
00:58:29,589 --> 00:58:28,079
are many many different possible

1499
00:58:31,349 --> 00:58:29,599
near-degenerate

1500
00:58:33,510 --> 00:58:31,359
combinations that normally also have

1501
00:58:35,510 --> 00:58:33,520
reasonably low free energy and this is a

1502
00:58:37,510 --> 00:58:35,520
consideration that makes rna generally

1503
00:58:38,870 --> 00:58:37,520
less re-foldable than protein and

1504
00:58:40,870 --> 00:58:38,880
perhaps something that we should think

1505
00:58:42,309 --> 00:58:40,880
more about in the context of origins of

1506
00:58:44,390 --> 00:58:42,319
life

1507
00:58:46,069 --> 00:58:44,400
but with that point aside i want to

1508
00:58:47,750 --> 00:58:46,079
briefly illustrate an experiment that

1509
00:58:49,510 --> 00:58:47,760
our team has been developing to try to

1510
00:58:50,470 --> 00:58:49,520
explore refoldability on the proteome

1511
00:58:51,990 --> 00:58:50,480
scale

1512
00:58:54,069 --> 00:58:52,000
so the way this experiment works is we

1513
00:58:56,150 --> 00:58:54,079

start with cells we lyse them using

1514
00:58:57,829 --> 00:58:56,160
cryogenic pulverization which retains

1515
00:58:59,190 --> 00:58:57,839
the vast majority of proteins in their

1516
00:59:01,109 --> 00:58:59,200
native structure

1517
00:59:02,630 --> 00:59:01,119
we divide that sample in half so the one

1518
00:59:04,470 --> 00:59:02,640
half will do nothing we'll call that the

1519
00:59:06,390 --> 00:59:04,480
native sample to the other half we

1520
00:59:08,390 --> 00:59:06,400
globally unfold the entire proteome

1521
00:59:09,990 --> 00:59:08,400
using six molar guanidine and then

1522
00:59:11,910 --> 00:59:10,000
re-fold it by removing that guanidine

1523
00:59:14,390 --> 00:59:11,920
with a hundred full dilution

1524
00:59:16,789 --> 00:59:14,400
now the key part of this experiment is

1525
00:59:19,030 --> 00:59:16,799
that we then expose these two samples to

1526
00:59:21,910 --> 00:59:19,040
pulse proteolysis with this enzyme

1527
00:59:23,589 --> 00:59:21,920
called proteinase k now proteinase k is

1528
00:59:25,270 --> 00:59:23,599
a protease that has virtually no

1529
00:59:27,510 --> 00:59:25,280
sequence specificity so it can cut

1530
00:59:29,190 --> 00:59:27,520
between any two amino acids but it does

1531
00:59:30,789 --> 00:59:29,200
have a very strong preference to cut

1532
00:59:32,390 --> 00:59:30,799
regions that are more susceptible or

1533
00:59:34,789 --> 00:59:32,400
solvent exposed

1534
00:59:37,589 --> 00:59:34,799
so as a consequence protein sk allows us

1535
00:59:39,430 --> 00:59:37,599
to encode structural information about

1536
00:59:42,069 --> 00:59:39,440
what the conformational ensemble of the

1537
00:59:43,910 --> 00:59:42,079
protein looks like into cleavage events

1538
00:59:45,750 --> 00:59:43,920

and of course since we are ultimately a

1539
00:59:47,910 --> 00:59:45,760
mass spectrometry proteomics lab what we

1540
00:59:50,309 --> 00:59:47,920
are very good at doing is sequencing and

1541
00:59:52,950 --> 00:59:50,319
quantifying tens if not 20 000 different

1542
00:59:55,670 --> 00:59:52,960
peptides in one sample so by identifying

1543
00:59:57,750 --> 00:59:55,680
the um different peptidic fragments that

1544
00:59:59,430 --> 00:59:57,760
come from these digest we can address

1545
01:00:01,670 --> 00:59:59,440
the question of whether or not a protein

1546
01:00:03,829 --> 01:00:01,680
was conformationally identical in the

1547
01:00:05,430 --> 01:00:03,839
refolded sample in which case you'd

1548
01:00:07,589 --> 01:00:05,440
expect to get the same pattern of

1549
01:00:09,349 --> 01:00:07,599
fragments or for whatever reason

1550
01:00:12,230 --> 01:00:09,359
non-refoldable in which case we would

1551
01:00:14,789 --> 01:00:12,240
expect novel cleavage sites to appear

1552
01:00:18,150 --> 01:00:14,799
that were not available in the protein

1553
01:00:19,829 --> 01:00:18,160
when it was in its native folded form

1554
01:00:21,510 --> 01:00:19,839
so what do we get when we do this

1555
01:00:24,309 --> 01:00:21,520
experiment to e coli we find that

1556
01:00:25,829 --> 01:00:24,319
roughly 60 of e coli proteins are

1557
01:00:27,190 --> 01:00:25,839
refoldable

1558
01:00:28,789 --> 01:00:27,200
whether or not you consider that a lot

1559
01:00:30,789 --> 01:00:28,799
or a little sort of a glass half empty

1560
01:00:31,990 --> 01:00:30,799
glass half full the data set that

1561
01:00:33,430 --> 01:00:32,000
actually i'm going to be talking more

1562
01:00:35,750 --> 01:00:33,440
about in this presentation is when we

1563
01:00:37,510 --> 01:00:35,760

did the same experiment in yeast which

1564
01:00:39,349 --> 01:00:37,520
surprisingly actually has a higher

1565
01:00:40,789 --> 01:00:39,359
refoldability index and that's something

1566
01:00:42,710 --> 01:00:40,799
that we think there's a lot of really

1567
01:00:44,549 --> 01:00:42,720
interesting molecular biology associated

1568
01:00:45,990 --> 01:00:44,559
with but for the purpose of this talk

1569
01:00:47,910 --> 01:00:46,000
i'm just going to talk about our yeast

1570
01:00:49,589 --> 01:00:47,920
data set because the trends in it happen

1571
01:00:51,829 --> 01:00:49,599
to be cleaner because there's no there's

1572
01:00:53,990 --> 01:00:51,839
very little if not any aggregation in

1573
01:00:55,829 --> 01:00:54,000
these experiments

1574
01:00:57,349 --> 01:00:55,839
so what can we say about what types of

1575
01:00:59,190 --> 01:00:57,359
proteins are good at folding on their

1576
01:01:01,270 --> 01:00:59,200
own well one thing that we can do that's

1577
01:01:03,430 --> 01:01:01,280
very very simple is just divide up these

1578
01:01:05,430 --> 01:01:03,440
proteins into the number of domains that

1579
01:01:07,430 --> 01:01:05,440
they have and one thing that we find

1580
01:01:09,750 --> 01:01:07,440
very cleanly is that the more domains

1581
01:01:11,910 --> 01:01:09,760
that a protein has the harder it is at

1582
01:01:13,750 --> 01:01:11,920
folding and this makes a lot of sense

1583
01:01:16,710 --> 01:01:13,760
because it's long been hypothesized that

1584
01:01:18,630 --> 01:01:16,720
multi-domain proteins rely more on

1585
01:01:20,870 --> 01:01:18,640
folding on the ribosome or so-called

1586
01:01:22,069 --> 01:01:20,880
co-translational folding and the reason

1587
01:01:24,470 --> 01:01:22,079
why that is is because when you're

1588
01:01:26,710 --> 01:01:24,480

folding on the ribosome the first domain

1589
01:01:28,950 --> 01:01:26,720
can fold before the second domain has

1590
01:01:30,710 --> 01:01:28,960
even been formed and the second domain

1591
01:01:32,870 --> 01:01:30,720
can fold after the first domain has

1592
01:01:35,109 --> 01:01:32,880
already folded so it acts as a

1593
01:01:37,349 --> 01:01:35,119
convenient way of decoupling the folding

1594
01:01:39,589 --> 01:01:37,359
of complex objects which of course is

1595
01:01:41,589 --> 01:01:39,599
not available when you're doing

1596
01:01:43,670 --> 01:01:41,599
refolding of a completely denatured

1597
01:01:45,190 --> 01:01:43,680
chain

1598
01:01:47,589 --> 01:01:45,200
now the other thing that we can do is we

1599
01:01:49,670 --> 01:01:47,599
can look at these individual domains and

1600
01:01:51,910 --> 01:01:49,680
classify them into an evolutionary

1601
01:01:53,670 --> 01:01:51,920
lineage and to do that we make use of

1602
01:01:55,910 --> 01:01:53,680
the e-cod system that you've heard about

1603
01:01:57,670 --> 01:01:55,920
from claudia as well as from liam this

1604
01:02:02,150 --> 01:01:57,680
is a way of classifying the protein

1605
01:02:07,109 --> 01:02:05,190
fold groups that have a common ancestor

1606
01:02:09,270 --> 01:02:07,119
and one thing that we find is that the

1607
01:02:10,789 --> 01:02:09,280
types of protein or the types of domains

1608
01:02:12,950 --> 01:02:10,799
rather i should say that are

1609
01:02:14,950 --> 01:02:12,960
extraordinarily refoldable have a lot of

1610
01:02:17,829 --> 01:02:14,960
traits in common they are generally

1611
01:02:19,029 --> 01:02:17,839
small they are generally all alpha or

1612
01:02:21,990 --> 01:02:19,039
all beta

1613
01:02:24,390 --> 01:02:22,000

and they are highly represented amongst

1614
01:02:26,710 --> 01:02:24,400
folds that bind to nucleotides and small

1615
01:02:29,109 --> 01:02:26,720
peptides and in that group it's both the

1616
01:02:31,430 --> 01:02:29,119
sh3 fold and the ob fold that claudia

1617
01:02:33,430 --> 01:02:31,440
was telling us a lot about what we find

1618
01:02:35,510 --> 01:02:33,440
is that the worst refolding in every

1619
01:02:37,430 --> 01:02:35,520
organism that we've looked at so far is

1620
01:02:39,910 --> 01:02:37,440
always found amongst folds that are

1621
01:02:42,549 --> 01:02:39,920
associated with the aminoacyl trna

1622
01:02:44,710 --> 01:02:42,559
synthetases as well as tin barrels with

1623
01:02:46,390 --> 01:02:44,720
rosmans and p-loops not being so far

1624
01:02:47,990 --> 01:02:46,400
behind

1625
01:02:49,589 --> 01:02:48,000
so just to sort of put a picture onto

1626
01:02:50,950 --> 01:02:49,599
some of these domains if you're not uh

1627
01:02:53,270 --> 01:02:50,960
used to looking at lots of different

1628
01:02:55,670 --> 01:02:53,280
protein structures again also reinforced

1629
01:02:57,829 --> 01:02:55,680
the sh3 and ob are these small albedo

1630
01:02:59,910 --> 01:02:57,839
folds the helix turn helix is a small

1631
01:03:01,990 --> 01:02:59,920
alpha fold and of course tim barrels and

1632
01:03:03,670 --> 01:03:02,000
rosemans are alpha slash beta folds that

1633
01:03:07,270 --> 01:03:03,680
tend to be larger and more topologically

1634
01:03:08,870 --> 01:03:07,280
complex and have a greater contact order

1635
01:03:10,390 --> 01:03:08,880
another thing that we can do is organize

1636
01:03:12,549 --> 01:03:10,400
these proteins on the basis of their

1637
01:03:14,710 --> 01:03:12,559
acidity what we find is that the worst

1638
01:03:15,910 --> 01:03:14,720

three folders are mildly acidic so that

1639
01:03:18,710 --> 01:03:15,920
means that these are things that have a

1640
01:03:21,190 --> 01:03:18,720
isoelectric point between five and seven

1641
01:03:22,950 --> 01:03:21,200
very acidic proteins tend to be pretty

1642
01:03:26,069 --> 01:03:22,960
good refolders and that i think bodes

1643
01:03:27,589 --> 01:03:26,079
well for hypotheses about um the ancient

1644
01:03:29,349 --> 01:03:27,599
proteins that were of course highly

1645
01:03:31,589 --> 01:03:29,359
acidic and would have had pis less than

1646
01:03:33,990 --> 01:03:31,599
five but we also find that very basic

1647
01:03:36,069 --> 01:03:34,000
proteins also tend to refold very well

1648
01:03:37,589 --> 01:03:36,079
and here our hypothesis is possibly that

1649
01:03:40,069 --> 01:03:37,599
these are proteins whose folding is

1650
01:03:42,789 --> 01:03:40,079
chaperoned by rna

1651
01:03:44,789 --> 01:03:42,799
now on that topic we can also look um

1652
01:03:46,390 --> 01:03:44,799
closely at the ribosomal proteins and

1653
01:03:48,870 --> 01:03:46,400
when we did that we found a truly

1654
01:03:51,349 --> 01:03:48,880
shocking discovery and that is that in

1655
01:03:53,990 --> 01:03:51,359
both e coli and in yeast the large

1656
01:03:55,910 --> 01:03:54,000
subunit is almost entirely refoldable in

1657
01:03:57,430 --> 01:03:55,920
yeast it's completely refoldable and

1658
01:03:59,589 --> 01:03:57,440
i'll remind you that this was not in

1659
01:04:01,349 --> 01:03:59,599
some pre-ordained biochemical reaction

1660
01:04:04,309 --> 01:04:01,359
this was literally refolding entire

1661
01:04:06,549 --> 01:04:04,319
extracts so lots of components very

1662
01:04:09,109 --> 01:04:06,559
messy the small subunit on the other

1663
01:04:10,549 --> 01:04:09,119

hand tends to be much less refoldable

1664
01:04:12,549 --> 01:04:10,559
and we think that this this is an

1665
01:04:14,549 --> 01:04:12,559
interesting finding that possibly points

1666
01:04:16,789 --> 01:04:14,559
to the antiquity of the large samina or

1667
01:04:18,470 --> 01:04:16,799
lisa's function in relation to the small

1668
01:04:19,750 --> 01:04:18,480
subunit

1669
01:04:21,349 --> 01:04:19,760
the final result that i'll share with

1670
01:04:23,270 --> 01:04:21,359
you is that we did this same refolding

1671
01:04:25,190 --> 01:04:23,280
reaction in thermos thermophilus which

1672
01:04:27,029 --> 01:04:25,200
is a model thermophile and we were

1673
01:04:28,230 --> 01:04:27,039
actually very struck by the finding that

1674
01:04:30,630 --> 01:04:28,240
actually in contrast to what we

1675
01:04:32,710 --> 01:04:30,640
hypothesized proteins from thermists

1676
01:04:35,589 --> 01:04:32,720
were miserable refolders they were much

1677
01:04:37,990 --> 01:04:35,599
worse than e coli and yeast

1678
01:04:40,910 --> 01:04:38,000
so why do we think this is we think that

1679
01:04:43,589 --> 01:04:40,920
the way that evolution is able to create

1680
01:04:45,349 --> 01:04:43,599
thermo-tolerant proteins is maybe not

1681
01:04:48,549 --> 01:04:45,359
through this classical mechanism of

1682
01:04:50,549 --> 01:04:48,559
having a very stable protein with a low

1683
01:04:53,990 --> 01:04:50,559
gibbs free energy but rather through a

1684
01:04:56,470 --> 01:04:54,000
kinetic trapping mechanism whereby the

1685
01:04:59,029 --> 01:04:56,480
barriers to exit the native state become

1686
01:05:00,870 --> 01:04:59,039
very high thereby trapping the protein

1687
01:05:03,670 --> 01:05:00,880
preventing thermal fluctuations from

1688
01:05:05,270 --> 01:05:03,680

unfolding it but by that same token it

1689
01:05:07,270 --> 01:05:05,280
means that if you wanted to refold that

1690
01:05:09,270 --> 01:05:07,280
protein after it was unfolded you'd be

1691
01:05:12,309 --> 01:05:09,280
in trouble because now those barriers

1692
01:05:14,789 --> 01:05:12,319
are going to act in both directions

1693
01:05:16,870 --> 01:05:14,799
so i'll summarize by trying to let you

1694
01:05:18,789 --> 01:05:16,880
know some of our current thinking about

1695
01:05:20,390 --> 01:05:18,799
how refoldability has affected the way

1696
01:05:22,390 --> 01:05:20,400
that at least in our lab we think about

1697
01:05:24,069 --> 01:05:22,400
the origins of life first of all we

1698
01:05:26,309 --> 01:05:24,079
think that the best three folders were

1699
01:05:29,349 --> 01:05:26,319
small topologically simple proteins that

1700
01:05:31,670 --> 01:05:29,359
bind peptides and nucleosides explicitly

1701
01:05:33,510 --> 01:05:31,680
not the synthetase folds now in some

1702
01:05:35,349 --> 01:05:33,520
ways this is maybe almost obvious you

1703
01:05:37,750 --> 01:05:35,359
know once you say it because synthetases

1704
01:05:39,109 --> 01:05:37,760
tend to be large multi-domain proteins

1705
01:05:40,870 --> 01:05:39,119
but i think it's worth pointing out that

1706
01:05:43,109 --> 01:05:40,880
this sort of notion that these represent

1707
01:05:45,109 --> 01:05:43,119
the most ancient proteins probably

1708
01:05:47,910 --> 01:05:45,119
represents a ripple from the an

1709
01:05:50,230 --> 01:05:47,920
implausibly strong rna world hypothesis

1710
01:05:52,470 --> 01:05:50,240
in which it has been positive by some

1711
01:05:53,990 --> 01:05:52,480
that proteins only became important once

1712
01:05:55,829 --> 01:05:54,000
you could encode them with an rna

1713
01:05:56,950 --> 01:05:55,839

template and of course in that train of

1714
01:05:59,109 --> 01:05:56,960
thought you couldn't even create

1715
01:06:00,870 --> 01:05:59,119
proteins until you had synthetases we

1716
01:06:03,270 --> 01:06:00,880
think the evidence from refoldability is

1717
01:06:05,349 --> 01:06:03,280
not consistent with that point of view

1718
01:06:07,270 --> 01:06:05,359
secondly we think that the large subunit

1719
01:06:09,270 --> 01:06:07,280
predated the small subunit so we think

1720
01:06:11,510 --> 01:06:09,280
the early life benefited from a catalyst

1721
01:06:13,430 --> 01:06:11,520
that could make peptide bonds before you

1722
01:06:15,589 --> 01:06:13,440
were able to encode that information in

1723
01:06:17,750 --> 01:06:15,599
a nucleic acid template we think that

1724
01:06:19,910 --> 01:06:17,760
that thinking nicely coheres with the

1725
01:06:21,510 --> 01:06:19,920
evolutionary and structural analysis

1726
01:06:23,910 --> 01:06:21,520
that the williams group has been working

1727
01:06:25,750 --> 01:06:23,920
on for several decades

1728
01:06:27,430 --> 01:06:25,760
we think that one thing that kind of

1729
01:06:29,510 --> 01:06:27,440
struck to us is that tim barrels

1730
01:06:31,109 --> 01:06:29,520
actually are pretty miserable refolders

1731
01:06:33,910 --> 01:06:31,119
we think that's because these like key

1732
01:06:36,549 --> 01:06:33,920
metabolic processes co-evolved with

1733
01:06:38,150 --> 01:06:36,559
translation so essentially once you have

1734
01:06:40,150 --> 01:06:38,160
translation you can start to create

1735
01:06:42,230 --> 01:06:40,160
proteins that are addicted to

1736
01:06:43,750 --> 01:06:42,240
translation in order to be able to fold

1737
01:06:45,750 --> 01:06:43,760
properly and so we think that

1738
01:06:48,230 --> 01:06:45,760

translation and glycolysis and the

1739
01:06:49,750 --> 01:06:48,240
synthetases by um const by consequence

1740
01:06:51,910 --> 01:06:49,760
co-evolve together

1741
01:06:53,829 --> 01:06:51,920
and finally we think that it would have

1742
01:06:55,510 --> 01:06:53,839
been actually relatively difficult to

1743
01:06:57,430 --> 01:06:55,520
initially evolve proteins in a

1744
01:06:59,270 --> 01:06:57,440
thermophilic setting because it seems

1745
01:07:01,589 --> 01:06:59,280
that thermophilic proteins are more

1746
01:07:03,750 --> 01:07:01,599
reliant on a robust translational

1747
01:07:06,150 --> 01:07:03,760
apparatus in order to create these

1748
01:07:07,670 --> 01:07:06,160
kinetically trapped folds so in essence

1749
01:07:09,990 --> 01:07:07,680
if we had seen that thermophilic

1750
01:07:12,710 --> 01:07:10,000
proteins refold very easily we might

1751
01:07:14,630 --> 01:07:12,720
have been able to accept the hypothesis

1752
01:07:17,510 --> 01:07:14,640
that these were ancient proteins that

1753
01:07:19,990 --> 01:07:17,520
were more easily able to assemble before

1754
01:07:22,150 --> 01:07:20,000
the advent of translation but that's not

1755
01:07:23,910 --> 01:07:22,160
exactly what our results show i'll put

1756
01:07:25,670 --> 01:07:23,920
some asterisks there because i think we

1757
01:07:27,430 --> 01:07:25,680
need to test the hypothesis on more

1758
01:07:29,670 --> 01:07:27,440
thermophiles first but that is where our

1759
01:07:32,309 --> 01:07:29,680
current evidence is taking us

1760
01:07:34,950 --> 01:07:32,319
so with that i want to conclude just by

1761
01:07:36,789 --> 01:07:34,960
acknowledging the extreme

1762
01:07:38,549 --> 01:07:36,799
importance that dan toffee has had in

1763
01:07:40,230 --> 01:07:38,559

shaping the thinking i think of a lot of

1764
01:07:41,430 --> 01:07:40,240
the people in this room as well as the

1765
01:07:43,029 --> 01:07:41,440
speakers

1766
01:07:44,710 --> 01:07:43,039
he's dearly missed and i'm glad that

1767
01:07:46,950 --> 01:07:44,720
we're able to

1768
01:07:48,870 --> 01:07:46,960
have a number of his trainees and

1769
01:07:56,470 --> 01:07:48,880
collaborators able to with to speak with

1770
01:08:01,029 --> 01:07:58,870
unfortunately we don't have time for

1771
01:08:08,150 --> 01:08:01,039
questions but at the end we will have

1772
01:08:13,510 --> 01:08:12,230
so our next speaker will be giving

1773
01:08:14,950 --> 01:08:13,520
a talk

1774
01:08:15,910 --> 01:08:14,960
remotely

1775
01:08:17,829 --> 01:08:15,920
it's

1776
01:08:22,229 --> 01:08:17,839
liam longo

1777
01:08:27,749 --> 01:08:24,709
i don't have the information uh from the

1778
01:08:42,390 --> 01:08:27,759
tokyo uh lc in tokyo

1779
01:08:42,400 --> 01:08:48,149
uh we don't have sound

1780
01:08:56,149 --> 01:08:51,669
well dna and rna i'm going to replay it

1781
01:09:00,229 --> 01:08:58,309
hello the title of my talk today is

1782
01:09:02,630 --> 01:09:00,239
through the looking glass functional

1783
01:09:06,470 --> 01:09:02,640
ambidexterity in an ancient nucleic acid

1784
01:09:08,390 --> 01:09:06,480
binding protein and i'm liam longo from

1785
01:09:10,950 --> 01:09:08,400
elsie at the tokyo institute of

1786
01:09:13,030 --> 01:09:10,960
technology and this is a joint project

1787
01:09:16,229 --> 01:09:13,040
with norman matanis at the hebrew

1788
01:09:18,229 --> 01:09:16,239

university of jerusalem

1789
01:09:21,030 --> 01:09:18,239
biopolymers as we all know are

1790
01:09:22,470 --> 01:09:21,040
exquisitely homochiral proteins use l

1791
01:09:25,349 --> 01:09:22,480
amino acids

1792
01:09:27,349 --> 01:09:25,359
while dna and rna are derived from

1793
01:09:28,229 --> 01:09:27,359
d-ribose

1794
01:09:30,709 --> 01:09:28,239
and so

1795
01:09:33,510 --> 01:09:30,719
while homochirality is the rule in

1796
01:09:34,950 --> 01:09:33,520
biology its origins are actually quite

1797
01:09:37,189 --> 01:09:34,960
mysterious

1798
01:09:39,269 --> 01:09:37,199
i think everyone here would agree that

1799
01:09:41,749 --> 01:09:39,279
homochirality probably predates the

1800
01:09:43,990 --> 01:09:41,759
leuka the exact point of emergence of

1801
01:09:46,390 --> 01:09:44,000
homochirality is unclear

1802
01:09:49,430 --> 01:09:46,400
and it's also unclear to what extent the

1803
01:09:50,550 --> 01:09:49,440
emergence of chomo chirality and rna was

1804
01:09:52,309 --> 01:09:50,560
coupled

1805
01:09:53,910 --> 01:09:52,319
to the emergence of homochiraldine

1806
01:09:56,390 --> 01:09:53,920
protein

1807
01:09:57,830 --> 01:09:56,400
and so although there are some very

1808
01:10:01,030 --> 01:09:57,840
interesting mechanisms that have been

1809
01:10:03,669 --> 01:10:01,040
proposed that can result in enantiomeric

1810
01:10:05,830 --> 01:10:03,679
excess in chemical systems

1811
01:10:08,310 --> 01:10:05,840
i think it's safe to say that the

1812
01:10:11,430 --> 01:10:08,320
question of homochirology and biology is

1813
01:10:14,870 --> 01:10:13,189

the veil between enantiomers is the

1814
01:10:16,709 --> 01:10:14,880
result of billions of years of

1815
01:10:19,110 --> 01:10:16,719
biological evolution

1816
01:10:22,310 --> 01:10:19,120
and the consequences of this veil were

1817
01:10:24,870 --> 01:10:22,320
first demonstrated by milton and kent

1818
01:10:27,270 --> 01:10:24,880
what milton and kent did is they

1819
01:10:29,750 --> 01:10:27,280
inverted the chirality of either hiv

1820
01:10:32,229 --> 01:10:29,760
protease or its substrate and they

1821
01:10:36,630 --> 01:10:32,239
showed that if you use the unnatural

1822
01:10:39,669 --> 01:10:36,640
couple so either lnd or dnl

1823
01:10:42,550 --> 01:10:39,679
you abolished activity but if you used

1824
01:10:44,630 --> 01:10:42,560
the natural couple or its mirror image

1825
01:10:46,550 --> 01:10:44,640
you actually had near equivalent

1826
01:10:49,270 --> 01:10:46,560
activity

1827
01:10:51,189 --> 01:10:49,280
and since then several technologies like

1828
01:10:53,990 --> 01:10:51,199
mirror image phase display and

1829
01:10:55,750 --> 01:10:54,000
sphegelmers have been developed to take

1830
01:10:58,790 --> 01:10:55,760
advantage of the properties of mirror

1831
01:11:01,189 --> 01:10:58,800
image molecules spiegelmers for example

1832
01:11:03,270 --> 01:11:01,199
are aftermers with high plasma stability

1833
01:11:05,189 --> 01:11:03,280
and low immunogenicity and this is

1834
01:11:07,590 --> 01:11:05,199
because they don't interact strongly

1835
01:11:09,189 --> 01:11:07,600
with nucleases or nucleic acid binding

1836
01:11:12,070 --> 01:11:09,199
proteins in the cell

1837
01:11:14,310 --> 01:11:12,080
but we wondered do the same truths hold

1838
01:11:16,310 --> 01:11:14,320

for the most ancient proteins

1839
01:11:17,590 --> 01:11:16,320
are they also highly sensitive to chiral

1840
01:11:20,550 --> 01:11:17,600
inversion

1841
01:11:23,430 --> 01:11:20,560
to ask this question we turn to a motif

1842
01:11:26,229 --> 01:11:23,440
called the helix herpen helix motif

1843
01:11:28,870 --> 01:11:26,239
and vikram alva and andre lupus have

1844
01:11:31,590 --> 01:11:28,880
shown that this is one of the most

1845
01:11:33,990 --> 01:11:31,600
ancient peptides and was at the origin

1846
01:11:36,070 --> 01:11:34,000
of folded proteins

1847
01:11:38,229 --> 01:11:36,080
what we've done previously

1848
01:11:40,149 --> 01:11:38,239
is we've used a combination of ancestor

1849
01:11:41,350 --> 01:11:40,159
reconstruction techniques and protein

1850
01:11:44,229 --> 01:11:41,360
engineering

1851
01:11:46,950 --> 01:11:44,239
to simplify the sequence of this motif

1852
01:11:49,750 --> 01:11:46,960
so that we can track its evolution from

1853
01:11:52,550 --> 01:11:49,760
a relatively unstructured peptide that

1854
01:11:54,870 --> 01:11:52,560
phase separates with dna into a folded

1855
01:11:56,950 --> 01:11:54,880
domain with specific double strand dna

1856
01:11:59,030 --> 01:11:56,960
binding activity

1857
01:12:00,950 --> 01:11:59,040
here is that model in a little bit more

1858
01:12:02,709 --> 01:12:00,960
detail

1859
01:12:05,990 --> 01:12:02,719
a long long time ago

1860
01:12:08,550 --> 01:12:06,000
we had flexible peptides probably with a

1861
01:12:11,590 --> 01:12:08,560
poly basic sequence composition that

1862
01:12:13,990 --> 01:12:11,600
formed coastavates with rna

1863
01:12:16,229 --> 01:12:14,000

over time those peptides became

1864
01:12:18,709 --> 01:12:16,239
more complicated and they were able to

1865
01:12:20,870 --> 01:12:18,719
adopt compact structures

1866
01:12:22,870 --> 01:12:20,880
these compact structures in the case of

1867
01:12:25,110 --> 01:12:22,880
the helix herpen helix motif could

1868
01:12:27,030 --> 01:12:25,120
potentially dimerize and these dimers

1869
01:12:29,430 --> 01:12:27,040
could promote the formation of more

1870
01:12:30,790 --> 01:12:29,440
stable coast surveys or phase separated

1871
01:12:32,070 --> 01:12:30,800
droplets

1872
01:12:33,830 --> 01:12:32,080
eventually

1873
01:12:34,950 --> 01:12:33,840
upon duplication and fusion of this

1874
01:12:38,229 --> 01:12:34,960
motif

1875
01:12:40,390 --> 01:12:38,239
we could achieve what is now observed as

1876
01:12:42,790 --> 01:12:40,400
an independently folding double strand

1877
01:12:45,830 --> 01:12:42,800
dna binding domain

1878
01:12:48,709 --> 01:12:45,840
remarkably we've been able to track

1879
01:12:50,630 --> 01:12:48,719
every one of these stages experimentally

1880
01:12:52,870 --> 01:12:50,640
in the laboratory

1881
01:12:55,189 --> 01:12:52,880
and so we recently submitted an article

1882
01:12:57,669 --> 01:12:55,199
in collaboration with daniela goldfarb

1883
01:13:00,470 --> 01:12:57,679
and the nasil where we characterize the

1884
01:13:01,990 --> 01:13:00,480
presence of these dimers inside the

1885
01:13:04,470 --> 01:13:02,000
coast survey

1886
01:13:06,709 --> 01:13:04,480
and with this model system in hand we

1887
01:13:08,229 --> 01:13:06,719
ask the question at what stage does

1888
01:13:10,149 --> 01:13:08,239

chirality matter

1889
01:13:12,950 --> 01:13:10,159
does it matter at the stage of forming

1890
01:13:15,910 --> 01:13:12,960
coesurvates by a simple dimerizing

1891
01:13:17,350 --> 01:13:15,920
peptide or does it matter at the level

1892
01:13:19,830 --> 01:13:17,360
of an independently folding

1893
01:13:21,910 --> 01:13:19,840
double-stranded dna binding domain

1894
01:13:24,070 --> 01:13:21,920
we started off by testing whether or not

1895
01:13:26,149 --> 01:13:24,080
coastervation or phase separation was

1896
01:13:29,030 --> 01:13:26,159
sensitive to chiral inversion and so

1897
01:13:32,870 --> 01:13:29,040
we'd previously shown that the l-peptide

1898
01:13:34,630 --> 01:13:32,880
coastervates strongly with polyu

1899
01:13:36,950 --> 01:13:34,640
when we inverted the chirality of the

1900
01:13:39,669 --> 01:13:36,960
l-peptide to form the d-peptide the

1901
01:13:42,229 --> 01:13:39,679
mir-image peptide we found that it still

1902
01:13:44,950 --> 01:13:42,239
formed coastervades with polyu

1903
01:13:46,790 --> 01:13:44,960
the differences here are because of the

1904
01:13:48,950 --> 01:13:46,800
cover slip we're using

1905
01:13:50,709 --> 01:13:48,960
it's not it's not a fundamental property

1906
01:13:53,030 --> 01:13:50,719
of the system

1907
01:13:55,910 --> 01:13:53,040
nevertheless using a nano site we were

1908
01:13:58,790 --> 01:13:55,920
able to see that actually the l-peptide

1909
01:14:03,110 --> 01:13:58,800
formed slightly more droplets than the

1910
01:14:05,590 --> 01:14:03,120
d-peptide at identical concentrations

1911
01:14:09,270 --> 01:14:05,600
now if you'll remember we previously

1912
01:14:11,430 --> 01:14:09,280
showed that inside the droplets there is

1913
01:14:13,669 --> 01:14:11,440

some folding of our peptide so we wanted

1914
01:14:16,070 --> 01:14:13,679
to test whether or not that folding was

1915
01:14:18,390 --> 01:14:16,080
important for coastervation

1916
01:14:22,229 --> 01:14:18,400
and to do that we generated a peptide

1917
01:14:24,310 --> 01:14:22,239
that had alternating d and l amino acids

1918
01:14:27,510 --> 01:14:24,320
such a peptide is unable to fold it's

1919
01:14:29,189 --> 01:14:27,520
unable to form alpha helices so

1920
01:14:31,510 --> 01:14:29,199
we tested whether or not this could

1921
01:14:34,149 --> 01:14:31,520
coast and indeed it could also

1922
01:14:36,870 --> 01:14:34,159
coastervate but it did so

1923
01:14:40,070 --> 01:14:36,880
with a lower propensity than either the

1924
01:14:42,790 --> 01:14:40,080
d-peptide or the l-peptide both of which

1925
01:14:45,990 --> 01:14:42,800
have the ability to fold and so we must

1926
01:14:49,990 --> 01:14:46,000
conclude that coastervation and face

1927
01:14:52,070 --> 01:14:50,000
separation is robust to chiral inversion

1928
01:14:54,390 --> 01:14:52,080
and this isn't perhaps very surprising

1929
01:14:56,470 --> 01:14:54,400
because it's already been shown that

1930
01:14:59,030 --> 01:14:56,480
largely unstructured peptides

1931
01:15:02,310 --> 01:14:59,040
can phase separate with double-stranded

1932
01:15:04,310 --> 01:15:02,320
or single-stranded dna or rna

1933
01:15:06,070 --> 01:15:04,320
this is perhaps because the nature of

1934
01:15:10,550 --> 01:15:06,080
the interactions that drive face

1935
01:15:12,390 --> 01:15:10,560
separation tend to be transient and weak

1936
01:15:14,310 --> 01:15:12,400
this is not the case for an

1937
01:15:17,030 --> 01:15:14,320
independently folding domain binding to

1938
01:15:18,709 --> 01:15:17,040

double-strand dna so how do we expect

1939
01:15:21,189 --> 01:15:18,719
this domain to withstand chiral

1940
01:15:23,189 --> 01:15:21,199
inversion to answer this question we

1941
01:15:25,990 --> 01:15:23,199
synthesized the full-length

1942
01:15:27,990 --> 01:15:26,000
double-strand dna binding domain in both

1943
01:15:29,510 --> 01:15:28,000
the mirror image chirality and the

1944
01:15:31,350 --> 01:15:29,520
natural chirality

1945
01:15:33,270 --> 01:15:31,360
and so this is the circular dichroism

1946
01:15:35,590 --> 01:15:33,280
spectra which reports on the secondary

1947
01:15:37,669 --> 01:15:35,600
structure of our domain we can see that

1948
01:15:41,270 --> 01:15:37,679
both domains are alpha-helical so they

1949
01:15:42,630 --> 01:15:41,280
have peaks at about 208 and 222 but that

1950
01:15:44,790 --> 01:15:42,640
they have an inverted circular

1951
01:15:47,669 --> 01:15:44,800
dichroisin spectrum because they have

1952
01:15:48,630 --> 01:15:47,679
helices of opposite handedness

1953
01:15:51,270 --> 01:15:48,640
now

1954
01:15:53,590 --> 01:15:51,280
using these two proteins we tested their

1955
01:15:54,950 --> 01:15:53,600
ability to bind double-strand dna using

1956
01:15:56,950 --> 01:15:54,960
spr

1957
01:16:00,870 --> 01:15:56,960
and we tested their ability to bind not

1958
01:16:02,149 --> 01:16:00,880
just the natural dna but we also used

1959
01:16:04,709 --> 01:16:02,159
ldna

1960
01:16:06,470 --> 01:16:04,719
this makes it so that our experiment has

1961
01:16:08,709 --> 01:16:06,480
a natural control

1962
01:16:11,430 --> 01:16:08,719
embedded in it because we expect that

1963
01:16:13,830 --> 01:16:11,440

the mere universe should have similar

1964
01:16:15,669 --> 01:16:13,840
affinities to our universe

1965
01:16:18,550 --> 01:16:15,679
and so as you can see here

1966
01:16:21,430 --> 01:16:18,560
both the l protein binding to the d dna

1967
01:16:24,630 --> 01:16:21,440
and the d protein binding to the l dna

1968
01:16:27,510 --> 01:16:24,640
they have a similar interaction affinity

1969
01:16:30,149 --> 01:16:27,520
surprisingly when we looked at l protein

1970
01:16:32,950 --> 01:16:30,159
binding to ldna or d protein binding to

1971
01:16:34,790 --> 01:16:32,960
d dna that is the case where only one of

1972
01:16:36,550 --> 01:16:34,800
the binding partners has an inverted

1973
01:16:39,110 --> 01:16:36,560
chirality

1974
01:16:41,430 --> 01:16:39,120
we still saw significant evidence of

1975
01:16:43,510 --> 01:16:41,440
binding and even in the tens of

1976
01:16:46,630 --> 01:16:43,520
micromolar concentration we have

1977
01:16:49,030 --> 01:16:46,640
unambiguous evidence of binding of our

1978
01:16:50,790 --> 01:16:49,040
protein to the dna

1979
01:16:52,470 --> 01:16:50,800
and we wanted to assess whether or not

1980
01:16:54,870 --> 01:16:52,480
this was the result of the background

1981
01:16:58,229 --> 01:16:54,880
binding of the fold itself and not the

1982
01:17:01,189 --> 01:16:58,239
result of specific binding to our domain

1983
01:17:02,790 --> 01:17:01,199
to do this we mutated the canonical

1984
01:17:05,910 --> 01:17:02,800
pgigp

1985
01:17:07,910 --> 01:17:05,920
binding loops to five glycines this is

1986
01:17:10,070 --> 01:17:07,920
in a sense an entropy mutation because

1987
01:17:12,229 --> 01:17:10,080
it doesn't change the overall charge of

1988
01:17:14,950 --> 01:17:12,239

the protein it just makes it so that

1989
01:17:16,709 --> 01:17:14,960
these loops are more flexible and thus

1990
01:17:18,870 --> 01:17:16,719
less likely to adopt the correct

1991
01:17:20,950 --> 01:17:18,880
confirmation for binding

1992
01:17:23,830 --> 01:17:20,960
when we do this we observe that the l

1993
01:17:25,030 --> 01:17:23,840
primordial rh protein with the five

1994
01:17:28,070 --> 01:17:25,040
glycines

1995
01:17:30,550 --> 01:17:28,080
actually binds worse than total chiral

1996
01:17:33,590 --> 01:17:30,560
inversion of the protein domain

1997
01:17:36,149 --> 01:17:33,600
on 29 base pair double stranded dna in

1998
01:17:39,189 --> 01:17:36,159
the natural chiral conformation we can

1999
01:17:42,070 --> 01:17:39,199
see that the d mere protein binds better

2000
01:17:44,470 --> 01:17:42,080
than the l primordial arch protein with

2001
01:17:46,229 --> 01:17:44,480
the 5g mutation

2002
01:17:47,910 --> 01:17:46,239
when you look at 101 base pair

2003
01:17:50,229 --> 01:17:47,920
double-stranded dna

2004
01:17:52,229 --> 01:17:50,239
we see the difference is even larger and

2005
01:17:55,030 --> 01:17:52,239
this is because in our system we've

2006
01:17:57,270 --> 01:17:55,040
observed that the longer the dna strand

2007
01:17:59,510 --> 01:17:57,280
the higher the binding affinity perhaps

2008
01:18:02,070 --> 01:17:59,520
due to some cooperativity

2009
01:18:05,030 --> 01:18:02,080
it's relatively easy to understand how a

2010
01:18:07,110 --> 01:18:05,040
single helix hairpin helix motif could

2011
01:18:08,790 --> 01:18:07,120
bind to double-stranded dna or

2012
01:18:11,510 --> 01:18:08,800
single-stranded dna

2013
01:18:14,550 --> 01:18:11,520

regardless of its chirality

2014
01:18:17,030 --> 01:18:14,560
what's harder to understand is how when

2015
01:18:19,270 --> 01:18:17,040
you have a duplicated domain and these

2016
01:18:20,229 --> 01:18:19,280
two loops are juxtaposed relative to

2017
01:18:22,390 --> 01:18:20,239
each other

2018
01:18:24,550 --> 01:18:22,400
how they could correctly insert into the

2019
01:18:26,470 --> 01:18:24,560
minor groove without a significant

2020
01:18:28,149 --> 01:18:26,480
rearrangement this is a question that

2021
01:18:31,590 --> 01:18:28,159
we're currently addressing with md

2022
01:18:35,910 --> 01:18:33,510
but now we have to grapple with the

2023
01:18:38,229 --> 01:18:35,920
question which is why would an ancient

2024
01:18:40,550 --> 01:18:38,239
nucleic acid binding domain be

2025
01:18:43,270 --> 01:18:40,560
ambidextrous why should an ancient

2026
01:18:45,750 --> 01:18:43,280
domain be able to bind in effectively

2027
01:18:47,669 --> 01:18:45,760
both chiral forms

2028
01:18:49,350 --> 01:18:47,679
and i would like to acknowledge right

2029
01:18:51,590 --> 01:18:49,360
out of the gate that this could be the

2030
01:18:54,550 --> 01:18:51,600
result of chance

2031
01:18:56,550 --> 01:18:54,560
some domains are surely ambidextrous

2032
01:18:58,709 --> 01:18:56,560
just by chance and that this has nothing

2033
01:19:00,550 --> 01:18:58,719
to do with the early history of the fold

2034
01:19:03,110 --> 01:19:00,560
and so if this was the case it would

2035
01:19:05,350 --> 01:19:03,120
predict that as we test more domains for

2036
01:19:07,110 --> 01:19:05,360
this property of ambidexterity the

2037
01:19:08,790 --> 01:19:07,120
ancient domains will have no greater

2038
01:19:10,950 --> 01:19:08,800

preference for amber dexterity than any

2039
01:19:13,510 --> 01:19:10,960
other fold so i want to acknowledge this

2040
01:19:15,830 --> 01:19:13,520
possibility right at the outset i think

2041
01:19:17,669 --> 01:19:15,840
it's a very reasonable one

2042
01:19:19,350 --> 01:19:17,679
but i'd also like to lean into the

2043
01:19:21,590 --> 01:19:19,360
result a bit more

2044
01:19:24,310 --> 01:19:21,600
what would it mean if the history of

2045
01:19:26,630 --> 01:19:24,320
homo chirality was written into the most

2046
01:19:29,590 --> 01:19:26,640
ancient domains and if this history was

2047
01:19:31,830 --> 01:19:29,600
somehow observable by their ability to

2048
01:19:33,350 --> 01:19:31,840
be ambidextrous

2049
01:19:35,590 --> 01:19:33,360
what would that mean

2050
01:19:38,950 --> 01:19:35,600
and could that be a relic of a time when

2051
01:19:40,709 --> 01:19:38,960
amino acid preferences were emerging in

2052
01:19:42,229 --> 01:19:40,719
a complex community of competing

2053
01:19:44,390 --> 01:19:42,239
organisms

2054
01:19:47,910 --> 01:19:44,400
in the model i've got here

2055
01:19:50,790 --> 01:19:47,920
we have an ancient ribosome a primitive

2056
01:19:53,189 --> 01:19:50,800
rna-based translation machine and it has

2057
01:19:54,550 --> 01:19:53,199
no preference for either l or d amino

2058
01:19:56,470 --> 01:19:54,560
acids

2059
01:19:58,310 --> 01:19:56,480
the resulting peptide would likely be

2060
01:20:00,630 --> 01:19:58,320
unstructured but it would still be able

2061
01:20:02,390 --> 01:20:00,640
to perform some simple function kind of

2062
01:20:04,790 --> 01:20:02,400
like the phase separating peptide we saw

2063
01:20:07,110 --> 01:20:04,800

at the beginning of the talk

2064
01:20:09,590 --> 01:20:07,120
over time however this primitive

2065
01:20:12,229 --> 01:20:09,600
rna-based translation machinery

2066
01:20:15,510 --> 01:20:12,239
would eventually develop some chiral

2067
01:20:16,229 --> 01:20:15,520
preference for either d or l amino acids

2068
01:20:19,709 --> 01:20:16,239
if

2069
01:20:22,390 --> 01:20:19,719
a community of these d and l preferring

2070
01:20:27,270 --> 01:20:22,400
proto-ribosomes existed along with

2071
01:20:28,709 --> 01:20:27,280
ribozyme aminoacyl trna synthetases

2072
01:20:29,750 --> 01:20:28,719
any gene

2073
01:20:32,629 --> 01:20:29,760
that could

2074
01:20:34,550 --> 01:20:32,639
operate in either chirality

2075
01:20:35,669 --> 01:20:34,560
would have an advantage in that

2076
01:20:37,590 --> 01:20:35,679
community

2077
01:20:39,750 --> 01:20:37,600
in other words an ancient preference for

2078
01:20:42,070 --> 01:20:39,760
ambidextrous protein domains could be

2079
01:20:45,590 --> 01:20:42,080
the result of a competition between a

2080
01:20:47,830 --> 01:20:45,600
complex community of early life that had

2081
01:20:50,310 --> 01:20:47,840
different amino acid preferences but

2082
01:20:52,149 --> 01:20:50,320
were sharing genes in any gene that

2083
01:20:54,629 --> 01:20:52,159
could have functioned in either chiral

2084
01:20:57,030 --> 01:20:54,639
form would have had a distinct advantage

2085
01:20:59,430 --> 01:20:57,040
in this complex community and it's from

2086
01:21:04,629 --> 01:20:59,440
this that we came up with the idea of an

2087
01:21:08,470 --> 01:21:06,550
and so with that i would like to thank

2088
01:21:10,229 --> 01:21:08,480

you for your attention i would like to

2089
01:21:11,510 --> 01:21:10,239
thank my wonderful collaborators for

2090
01:21:14,070 --> 01:21:11,520
their hard work

2091
01:21:15,910 --> 01:21:14,080
and if this theory sounds too crazy or

2092
01:21:22,410 --> 01:21:15,920
just crazy enough and you'd like to talk

2093
01:21:28,310 --> 01:21:26,629
[Applause]

2094
01:21:29,830 --> 01:21:28,320
thank you liam i'm sorry that we won't

2095
01:21:32,070 --> 01:21:29,840
be able to chat with you more here but

2096
01:21:33,430 --> 01:21:32,080
hopefully some of us will do offline or

2097
01:21:35,669 --> 01:21:33,440
by email

2098
01:21:38,310 --> 01:21:35,679
um and with that i'd like to introduce

2099
01:21:40,950 --> 01:21:38,320
our final presenter who's also coming to

2100
01:21:43,510 --> 01:21:40,960
us remotely from the charles university

2101
01:21:44,229 --> 01:21:43,520
of prague in the czech republic and this

2102
01:21:45,350 --> 01:21:44,239
is

2103
01:21:50,950 --> 01:21:45,360
um

2104
01:21:57,430 --> 01:21:54,070
good evening good evening from israel

2105
01:21:59,189 --> 01:21:57,440
i'm going to present part of the work

2106
01:22:00,149 --> 01:21:59,199
which i did at charles university in

2107
01:22:01,350 --> 01:22:00,159
prague

2108
01:22:04,550 --> 01:22:01,360
and

2109
01:22:08,070 --> 01:22:04,560
now i'm residing at weitzman institute

2110
01:22:10,390 --> 01:22:08,080
so in our lab we were considering this

2111
01:22:11,510 --> 01:22:10,400
peculiar disparity just mentioned by

2112
01:22:15,830 --> 01:22:11,520
claudia

2113
01:22:18,390 --> 01:22:15,840

that with only 100 residue protein

2114
01:22:20,070 --> 01:22:18,400
we can construct 20 to 100 possible

2115
01:22:22,709 --> 01:22:20,080
protein sequences

2116
01:22:24,790 --> 01:22:22,719
but approximately only 10 to 15

2117
01:22:25,830 --> 01:22:24,800
different protein sequences are used by

2118
01:22:27,910 --> 01:22:25,840
nature

2119
01:22:28,870 --> 01:22:27,920
so why is that and what is hidden in

2120
01:22:31,510 --> 01:22:28,880
this

2121
01:22:35,189 --> 01:22:31,520
dark protein space was exactly what we

2122
01:22:38,950 --> 01:22:35,199
were interesting interested

2123
01:22:41,510 --> 01:22:38,960
so long story short we made in vitro

2124
01:22:42,790 --> 01:22:41,520
random libraries

2125
01:22:45,030 --> 01:22:42,800
and

2126
01:22:47,110 --> 01:22:45,040
for doing so we used two different amino

2127
01:22:49,910 --> 01:22:47,120
acid alphabet full alphabet consisting

2128
01:22:51,830 --> 01:22:49,920
of all 20 amino acids and so-called

2129
01:22:54,790 --> 01:22:51,840
early alphabet

2130
01:22:56,470 --> 01:22:54,800
which used only periodically available

2131
01:23:00,229 --> 01:22:56,480
amino acids

2132
01:23:02,950 --> 01:23:00,239
length

2133
01:23:05,430 --> 01:23:02,960
consisting of these randomized parts and

2134
01:23:08,550 --> 01:23:05,440
we introduced the thrombin

2135
01:23:10,629 --> 01:23:08,560
protease cleavage site in the middle

2136
01:23:12,310 --> 01:23:10,639
so the first essay which we tried was

2137
01:23:14,550 --> 01:23:12,320
the solubility

2138
01:23:16,550 --> 01:23:14,560

section of the library

2139
01:23:19,669 --> 01:23:16,560
so that was assessed simply by

2140
01:23:20,870 --> 01:23:19,679
expression of our randomized billions of

2141
01:23:23,510 --> 01:23:20,880
sequences

2142
01:23:25,590 --> 01:23:23,520
in a cell free expression system and

2143
01:23:28,070 --> 01:23:25,600
western lotting and for solubility we

2144
01:23:31,270 --> 01:23:28,080
just spinned the mixture and took the

2145
01:23:32,870 --> 01:23:31,280
supernatant in supernatant to assess the

2146
01:23:33,669 --> 01:23:32,880
soluble fraction

2147
01:23:35,350 --> 01:23:33,679
so

2148
01:23:39,189 --> 01:23:35,360
upon the expression in three different

2149
01:23:41,350 --> 01:23:39,199
temperatures 25 30 and 37 degrees

2150
01:23:42,629 --> 01:23:41,360
we've seen a monitoring increase in

2151
01:23:44,709 --> 01:23:42,639
expression

2152
01:23:46,830 --> 01:23:44,719
in early and full

2153
01:23:49,590 --> 01:23:46,840
amino acid alphabet libraries as

2154
01:23:51,350 --> 01:23:49,600
expected but

2155
01:23:53,910 --> 01:23:51,360
the solubility of these two libraries

2156
01:23:57,189 --> 01:23:53,920
showed that while early alphabet

2157
01:23:59,590 --> 01:23:57,199
proteins are essentially fully soluble

2158
01:24:02,470 --> 01:23:59,600
in all temperatures the full alphabet

2159
01:24:04,870 --> 01:24:02,480
library is only partially soluble and

2160
01:24:06,390 --> 01:24:04,880
its solubility remains approximately

2161
01:24:08,950 --> 01:24:06,400
constant

2162
01:24:11,750 --> 01:24:08,960
within our temperature range

2163
01:24:15,110 --> 01:24:11,760

so next i tried to add the chaperone dna

2164
01:24:18,709 --> 01:24:15,120
k into the sulfury mixture and again i

2165
01:24:20,790 --> 01:24:18,719
seen no effect in early alphabet library

2166
01:24:23,669 --> 01:24:20,800
so supplementation of chaperone did not

2167
01:24:25,430 --> 01:24:23,679
improve the expression anyhow

2168
01:24:27,750 --> 01:24:25,440
but in the full amino acid alpha

2169
01:24:31,270 --> 01:24:27,760
alphabet library i see small deviation

2170
01:24:33,189 --> 01:24:31,280
however the difference is not large

2171
01:24:35,750 --> 01:24:33,199
the interesting part is that

2172
01:24:38,870 --> 01:24:35,760
the soluble part of the libraries

2173
01:24:40,950 --> 01:24:38,880
of sharper and supplemented libraries

2174
01:24:43,350 --> 01:24:40,960
showed interesting trends that earlier

2175
01:24:47,669 --> 01:24:43,360
amino acid level library remained

2176
01:24:50,870 --> 01:24:47,679
soluble as shown before but the full

2177
01:24:52,790 --> 01:24:50,880
amino acid alphabet library got

2178
01:24:54,790 --> 01:24:52,800
completely solubilized in the presence

2179
01:24:56,709 --> 01:24:54,800
of chaperone which means that chaperone

2180
01:24:58,550 --> 01:24:56,719
can actually act on

2181
01:24:59,830 --> 01:24:58,560
proteins without any evolutionary

2182
01:25:02,629 --> 01:24:59,840
background

2183
01:25:05,350 --> 01:25:02,639
so the next essay after our

2184
01:25:07,830 --> 01:25:05,360
centrifugation solubility essay was the

2185
01:25:09,510 --> 01:25:07,840
proteolysis assay which allowed us to

2186
01:25:11,910 --> 01:25:09,520
separate

2187
01:25:13,189 --> 01:25:11,920
the whole combinatorial library into

2188
01:25:15,669 --> 01:25:13,199

four parts

2189
01:25:17,510 --> 01:25:15,679
the soluble and degradable degradable

2190
01:25:20,149 --> 01:25:17,520
and soluble degradable and degradable

2191
01:25:22,149 --> 01:25:20,159
which corresponds to the more structured

2192
01:25:23,990 --> 01:25:22,159
parts of soluble proteins

2193
01:25:25,750 --> 01:25:24,000
and more

2194
01:25:27,110 --> 01:25:25,760
disordered parts of soluble and

2195
01:25:28,950 --> 01:25:27,120
insoluble

2196
01:25:30,790 --> 01:25:28,960
fraction of the library

2197
01:25:33,590 --> 01:25:30,800
so

2198
01:25:35,669 --> 01:25:33,600
these are the results the this figure is

2199
01:25:38,229 --> 01:25:35,679
quite complicated and i have no chance

2200
01:25:39,750 --> 01:25:38,239
to describe all the juicy details which

2201
01:25:40,709 --> 01:25:39,760
are contained with them

2202
01:25:42,790 --> 01:25:40,719
but

2203
01:25:45,830 --> 01:25:42,800
let's consider only the

2204
01:25:47,430 --> 01:25:45,840
dark blue part of all these

2205
01:25:49,430 --> 01:25:47,440
results of

2206
01:25:52,070 --> 01:25:49,440
full amino acid alphabet and early amino

2207
01:25:54,870 --> 01:25:52,080
acid altered libraries without and with

2208
01:25:56,870 --> 01:25:54,880
chaperones we see that structured

2209
01:25:58,149 --> 01:25:56,880
fraction is prevalent in all four

2210
01:26:01,350 --> 01:25:58,159
conditions

2211
01:26:03,669 --> 01:26:01,360
and upon the addition of chaperones we

2212
01:26:05,669 --> 01:26:03,679
do not see any induction of the

2213
01:26:09,990 --> 01:26:05,679

structure that means that

2214
01:26:15,030 --> 01:26:13,110
coded within its primaries

2215
01:26:17,830 --> 01:26:15,040
so in conclusion

2216
01:26:20,070 --> 01:26:17,840
we think that early alphabet is soluble

2217
01:26:21,990 --> 01:26:20,080
and chaperone independent

2218
01:26:24,310 --> 01:26:22,000
that full alphabet is solubilized by

2219
01:26:27,110 --> 01:26:24,320
chaperones we observed similar compacted

2220
01:26:29,030 --> 01:26:27,120
structure frequency in both libraries uh

2221
01:26:29,750 --> 01:26:29,040
i've seen that chaperones do not promote

2222
01:26:34,229 --> 01:26:29,760
the

2223
01:26:35,910 --> 01:26:34,239
possible structure formation in a

2224
01:26:36,950 --> 01:26:35,920
prebiotically plausible alphabet

2225
01:26:39,430 --> 01:26:36,960
libraries

2226
01:26:41,270 --> 01:26:39,440
and we showed that chaperones do

2227
01:26:42,470 --> 01:26:41,280
positively interact with the random

2228
01:26:44,950 --> 01:26:42,480
sequences

2229
01:26:47,430 --> 01:26:44,960
so with all of that i

2230
01:26:49,510 --> 01:26:47,440
recommend you to look at our paper where

2231
01:26:50,470 --> 01:26:49,520
we describe many other interesting

2232
01:26:53,110 --> 01:26:50,480
things

2233
01:26:56,229 --> 01:26:53,120
uh on how we made shop more homemade

2234
01:26:58,550 --> 01:26:56,239
libraries how the library is

2235
01:27:00,470 --> 01:26:58,560
behaving upon the heat shock different

2236
01:27:01,669 --> 01:27:00,480
protease essay and bioinformatic

2237
01:27:05,030 --> 01:27:01,679
predictions

2238
01:27:07,430 --> 01:27:05,040

and with all that thank you thank clara

2239
01:27:09,180 --> 01:27:07,440
and thank organizers to you of the

2240
01:27:15,350 --> 01:27:09,190
conference

2241
01:27:15,360 --> 01:27:21,110
lovely thank you

2242
01:27:25,990 --> 01:27:24,149
so um let's have maybe a few minutes of

2243
01:27:28,229 --> 01:27:26,000
discussion with all the speakers so

2244
01:27:30,390 --> 01:27:28,239
speakers who are in person um you can

2245
01:27:32,470 --> 01:27:30,400
maybe join us on the panel those who are

2246
01:27:34,070 --> 01:27:32,480
online maybe stay in the room

2247
01:27:36,070 --> 01:27:34,080
if you have a question for any of the

2248
01:27:38,470 --> 01:27:36,080
speakers please uh

2249
01:27:40,229 --> 01:27:38,480
line up behind one of the mics and

2250
01:27:41,830 --> 01:27:40,239
we'll probably spend more time hanging

2251
01:27:56,310 --> 01:27:41,840
out after this because there's nothing

2252
01:28:00,149 --> 01:27:58,390
hello there shelby osborne university of

2253
01:28:03,350 --> 01:28:00,159
arkansas center for planetary and space

2254
01:28:05,669 --> 01:28:03,360
sciences this is a question for dr freud

2255
01:28:06,950 --> 01:28:05,679
is that how you pronounce it oh sorry

2256
01:28:08,950 --> 01:28:06,960
it's free dude

2257
01:28:10,870 --> 01:28:08,960
well i'm from arkansas so we just say

2258
01:28:13,830 --> 01:28:10,880
fried all the time

2259
01:28:18,070 --> 01:28:13,840
so i was just going to ask you e coli

2260
01:28:21,110 --> 01:28:18,080
and yeast have a lot of similar enzymes

2261
01:28:22,629 --> 01:28:21,120
and generally we study those in tandem

2262
01:28:25,590 --> 01:28:22,639
anyways

2263
01:28:28,390 --> 01:28:25,600

what would the approach be if you had a

2264
01:28:30,310 --> 01:28:28,400
protein or a ribonuclease sequence

2265
01:28:32,470 --> 01:28:30,320
and you wanted to know what that

2266
01:28:34,629 --> 01:28:32,480
sequence was like before

2267
01:28:37,110 --> 01:28:34,639
the modern folding but you don't know

2268
01:28:38,310 --> 01:28:37,120
what the original or analogous structure

2269
01:28:40,470 --> 01:28:38,320
was

2270
01:28:41,510 --> 01:28:40,480
yeah cool that's a great question

2271
01:28:42,950 --> 01:28:41,520
um

2272
01:28:45,030 --> 01:28:42,960
so

2273
01:28:46,950 --> 01:28:45,040
the the trends that we see in e coli and

2274
01:28:48,950 --> 01:28:46,960
the trends that we see in yeast are

2275
01:28:50,790 --> 01:28:48,960
basically the same so like whatever is

2276
01:28:52,870 --> 01:28:50,800
more refillable in e coli is also more

2277
01:28:54,950 --> 01:28:52,880
refillable in yeast it's just that in

2278
01:28:57,510 --> 01:28:54,960
any given category the yeast ortholog is

2279
01:28:59,750 --> 01:28:57,520
generally more reflectable on average by

2280
01:29:01,669 --> 01:28:59,760
about 15 to 20 percent and we've

2281
01:29:04,229 --> 01:29:01,679
recently i think come up with a pretty

2282
01:29:06,149 --> 01:29:04,239
um convincing explanation for why that

2283
01:29:07,830 --> 01:29:06,159
is and it can be basically explained in

2284
01:29:09,990 --> 01:29:07,840
terms of the fact that yeast proteins

2285
01:29:12,870 --> 01:29:10,000
are more disordered so the extra

2286
01:29:15,430 --> 01:29:12,880
disorder that tends to punctuate between

2287
01:29:17,430 --> 01:29:15,440
the folded domains and yeast proteins

2288
01:29:18,870 --> 01:29:17,440

seems to make it easier to refold them

2289
01:29:21,270 --> 01:29:18,880
off the ribosome because they're sort of

2290
01:29:23,990 --> 01:29:21,280
less likely to get in each other's way

2291
01:29:26,629 --> 01:29:24,000
whereas the e coli proteins tend to have

2292
01:29:29,110 --> 01:29:26,639
very short if any disordered linkers at

2293
01:29:32,070 --> 01:29:29,120
all and that in our opinion or at least

2294
01:29:32,790 --> 01:29:32,080
our hypothesis is that that destined to

2295
01:29:35,830 --> 01:29:32,800
be

2296
01:29:38,550 --> 01:29:35,840
dependent on translation to fold

2297
01:29:40,629 --> 01:29:38,560
and if we didn't know for example that e

2298
01:29:42,950 --> 01:29:40,639
coli and yeast were correlated how would

2299
01:29:44,470 --> 01:29:42,960
we approach the problem of figuring out

2300
01:29:46,310 --> 01:29:44,480
what the

2301
01:29:48,550 --> 01:29:46,320
previous structure

2302
01:29:51,030 --> 01:29:48,560
and enzymes and

2303
01:29:52,550 --> 01:29:51,040
proteins of yeast would have been if we

2304
01:29:55,030 --> 01:29:52,560
didn't know that e coli existed we mean

2305
01:29:56,629 --> 01:29:55,040
like the ancestral sequences

2306
01:29:57,830 --> 01:29:56,639
like the precursor

2307
01:29:59,590 --> 01:29:57,840
oh i see

2308
01:30:01,830 --> 01:29:59,600
i mean we could do we haven't done it

2309
01:30:03,030 --> 01:30:01,840
yet but a cool experiment to do would be

2310
01:30:05,430 --> 01:30:03,040
to do the sort of ancestral

2311
01:30:07,350 --> 01:30:05,440
reconstruction and ask you know how does

2312
01:30:08,950 --> 01:30:07,360
the property change for

2313
01:30:10,709 --> 01:30:08,960

proteins that are perceived to be more

2314
01:30:12,390 --> 01:30:10,719
ancient but we haven't done that yet

2315
01:30:14,149 --> 01:30:12,400
okay interesting and may i get your

2316
01:30:16,790 --> 01:30:14,159
contact information after the question

2317
01:30:20,629 --> 01:30:16,800
maybe offline just so some of you okay

2318
01:30:26,390 --> 01:30:23,590
josh ariola uc san diego i had a quick

2319
01:30:28,950 --> 01:30:26,400
question for valerio

2320
01:30:30,629 --> 01:30:28,960
um i was wondering if you were able to

2321
01:30:33,350 --> 01:30:30,639
observe any

2322
01:30:36,310 --> 01:30:33,360
protective effect on the rna by the

2323
01:30:37,510 --> 01:30:36,320
peptide or the protein

2324
01:30:39,669 --> 01:30:37,520
um

2325
01:30:41,270 --> 01:30:39,679
protective like effect you mean like

2326
01:30:43,430 --> 01:30:41,280
yeah yeah

2327
01:30:45,750 --> 01:30:43,440
if you had like

2328
01:30:47,669 --> 01:30:45,760
high magnesium and high ph i was

2329
01:30:50,629 --> 01:30:47,679
wondering if you could see

2330
01:30:53,510 --> 01:30:50,639
less rna cleavage like self cleavage

2331
01:30:55,110 --> 01:30:53,520
when you have the peptide present

2332
01:30:57,030 --> 01:30:55,120
no we didn't perform this kind of

2333
01:31:00,310 --> 01:30:57,040
experiment and we perform like a

2334
01:31:02,470 --> 01:31:00,320
hydrolysis by erenesis and proteases

2335
01:31:04,950 --> 01:31:02,480
and that one yeah we perform it so like

2336
01:31:07,910 --> 01:31:04,960
removing uh actually there's this uh by

2337
01:31:11,110 --> 01:31:07,920
ernest so we had like uh escalator adta

2338
01:31:13,990 --> 01:31:11,120

in the in the media and uh we saw that

2339
01:31:17,430 --> 01:31:14,000
when we added edta the complex get

2340
01:31:19,030 --> 01:31:17,440
degradated when instead like the dta

2341
01:31:21,590 --> 01:31:19,040
it's removed from the media so there is

2342
01:31:23,750 --> 01:31:21,600
magnesium the the complex is stable and

2343
01:31:24,470 --> 01:31:23,760
the aeronasis is not able to degradate

2344
01:31:35,830 --> 01:31:24,480
so

2345
01:31:37,990 --> 01:31:35,840
protect the the the the binding from the

2346
01:31:40,790 --> 01:31:38,000
the cleavage by the protein rnase in

2347
01:31:42,390 --> 01:31:40,800
presence of magnesium or not so

2348
01:31:44,629 --> 01:31:42,400
but yeah it's a good experiment like to

2349
01:31:47,110 --> 01:31:44,639
to try also with the higher

2350
01:31:50,229 --> 01:31:47,120
concentration and titration yeah cool

2351
01:31:54,790 --> 01:31:53,030
hi um i'm self son from university of

2352
01:31:55,830 --> 01:31:54,800
arizona and i have a question to steven

2353
01:31:58,310 --> 01:31:55,840
freed

2354
01:32:00,790 --> 01:31:58,320
uh i know this is a long shot but i was

2355
01:32:02,149 --> 01:32:00,800
wondering if there is a software or

2356
01:32:04,149 --> 01:32:02,159
something

2357
01:32:06,149 --> 01:32:04,159
that allows you to calculate

2358
01:32:07,990 --> 01:32:06,159
refoldability as a matrix from the

2359
01:32:10,229 --> 01:32:08,000
sequence just like you calculate this

2360
01:32:11,830 --> 01:32:10,239
order propensity i think it's an amazing

2361
01:32:13,030 --> 01:32:11,840
goal that we would love to be able to do

2362
01:32:15,510 --> 01:32:13,040
and i think that

2363
01:32:17,830 --> 01:32:15,520

the the stage that we're operating at is

2364
01:32:19,669 --> 01:32:17,840
to try to collect features like

2365
01:32:21,510 --> 01:32:19,679
biophysical structural that we can

2366
01:32:23,270 --> 01:32:21,520
associate with it and then

2367
01:32:24,950 --> 01:32:23,280
i think that ultimately as we sort of

2368
01:32:27,030 --> 01:32:24,960
get more and more features and map more

2369
01:32:29,030 --> 01:32:27,040
protiums it shouldn't be too crazy to

2370
01:32:31,350 --> 01:32:29,040
involve some machine learning algorithm

2371
01:32:33,030 --> 01:32:31,360
to assimilate it all together but

2372
01:32:34,550 --> 01:32:33,040
for that i'll maybe ask for your help

2373
01:32:36,550 --> 01:32:34,560
because to me machine learning still

2374
01:32:41,030 --> 01:32:36,560
mystifies me

2375
01:32:44,470 --> 01:32:42,950
so i have two questions one fairly

2376
01:32:46,629 --> 01:32:44,480
specific and more general and the

2377
01:32:48,149 --> 01:32:46,639
specific one is is definitely for i

2378
01:32:50,229 --> 01:32:48,159
guess just for stephen

2379
01:32:51,910 --> 01:32:50,239
and the general one mostly i think

2380
01:32:53,990 --> 01:32:51,920
applies to your talk but could apply to

2381
01:32:55,350 --> 01:32:54,000
others so please chime in if it does so

2382
01:32:56,790 --> 01:32:55,360
the first one is

2383
01:32:58,470 --> 01:32:56,800
um

2384
01:33:01,430 --> 01:32:58,480
for the ribosome you said for the

2385
01:33:03,830 --> 01:33:01,440
ribosome refolding uh

2386
01:33:05,750 --> 01:33:03,840
whatever the results there

2387
01:33:07,910 --> 01:33:05,760
do i understand that the assay for that

2388
01:33:10,070 --> 01:33:07,920

was simply you had an extract and you

2389
01:33:11,510 --> 01:33:10,080
you heat it up to to unfold it and

2390
01:33:14,229 --> 01:33:11,520
re-fold and then you were using that

2391
01:33:15,990 --> 01:33:14,239
protea uh protease assay that you use

2392
01:33:18,629 --> 01:33:16,000
for the other proteins it's all

2393
01:33:20,550 --> 01:33:18,639
was that all the same assay yeah so the

2394
01:33:23,110 --> 01:33:20,560
basic structure of the assays you take

2395
01:33:25,510 --> 01:33:23,120
an entire extract

2396
01:33:27,990 --> 01:33:25,520
add solid guanidinium chloride to it to

2397
01:33:30,390 --> 01:33:28,000
unfold everything in it and then dilute

2398
01:33:32,390 --> 01:33:30,400
it out in order to

2399
01:33:33,270 --> 01:33:32,400
refold things and then you compare that

2400
01:33:37,750 --> 01:33:33,280
to

2401
01:33:40,070 --> 01:33:37,760
the original unfolding but where they're

2402
01:33:42,629 --> 01:33:40,080
otherwise compositionally identical it's

2403
01:33:44,790 --> 01:33:42,639
just simply had different histories and

2404
01:33:47,830 --> 01:33:44,800
then the confirmation of the proteins is

2405
01:33:49,430 --> 01:33:47,840
then probed with the protease so when we

2406
01:33:51,990 --> 01:33:49,440
say that the large subunit seems to be

2407
01:33:54,870 --> 01:33:52,000
refoldable what we really mean is that

2408
01:33:56,950 --> 01:33:54,880
amongst the 36

2409
01:33:59,430 --> 01:33:56,960
large ribosomal proteins for which we

2410
01:34:01,910 --> 01:33:59,440
have data we can't tell any difference

2411
01:34:04,149 --> 01:34:01,920
in the proteolysis profile before versus

2412
01:34:06,149 --> 01:34:04,159
after but it seems to be quite different

2413
01:34:07,910 --> 01:34:06,159

for the small subunit

2414
01:34:10,709 --> 01:34:07,920
and the second more general question is

2415
01:34:12,550 --> 01:34:10,719
that uh for for your assay and for any

2416
01:34:14,870 --> 01:34:12,560
any for most of these other talks as

2417
01:34:16,229 --> 01:34:14,880
well of course membrane proteins are

2418
01:34:17,669 --> 01:34:16,239
very important to biology now or

2419
01:34:19,430 --> 01:34:17,679
probably very important from very early

2420
01:34:21,990 --> 01:34:19,440
on perhaps some simple membrane proteins

2421
01:34:23,910 --> 01:34:22,000
but they kind of represent a

2422
01:34:25,590 --> 01:34:23,920
particularly difficult challenge i think

2423
01:34:27,750 --> 01:34:25,600
for some of these so like in your

2424
01:34:29,350 --> 01:34:27,760
refolding assay presumably you're not

2425
01:34:30,950 --> 01:34:29,360
yeah in a position to look at anything

2426
01:34:33,030 --> 01:34:30,960
but the soluble proteins and in the last

2427
01:34:35,350 --> 01:34:33,040
talk one of the screens was for

2428
01:34:37,350 --> 01:34:35,360
solubility and i think

2429
01:34:38,870 --> 01:34:37,360
maybe sort of implied that

2430
01:34:40,470 --> 01:34:38,880
that it's important to have that

2431
01:34:41,590 --> 01:34:40,480
solubility but

2432
01:34:42,870 --> 01:34:41,600
in fact there are probably a lot of

2433
01:34:44,470 --> 01:34:42,880
early proteins it's very important that

2434
01:34:46,229 --> 01:34:44,480
they not have that property that they

2435
01:34:48,629 --> 01:34:46,239
they punch into a membrane so that's

2436
01:34:50,709 --> 01:34:48,639
that's the more general question i think

2437
01:34:52,550 --> 01:34:50,719
certainly anyone who thinks that they

2438
01:34:54,470 --> 01:34:52,560

might have something relevant please

2439
01:34:55,830 --> 01:34:54,480
chime in but

2440
01:34:59,030 --> 01:34:55,840
slava do you want to comment were you

2441
01:35:02,950 --> 01:35:01,270
i can just say at least briefly for ours

2442
01:35:04,629 --> 01:35:02,960
so yeah you're absolutely right our

2443
01:35:06,709 --> 01:35:04,639
assay has a blind spot to membrane

2444
01:35:08,550 --> 01:35:06,719
proteins because we essentially lice

2445
01:35:10,229 --> 01:35:08,560
without detergent and then they all come

2446
01:35:13,189 --> 01:35:10,239
out and then we do all of our refolding

2447
01:35:14,629 --> 01:35:13,199
on the clarified extract so in essence

2448
01:35:17,350 --> 01:35:14,639
we would love to know more about it but

2449
01:35:21,510 --> 01:35:17,360
we can't say much about it

2450
01:35:24,950 --> 01:35:22,790
hello i'm andrew wheeler from the

2451
01:35:26,950 --> 01:35:24,960
university of arizona i have a question

2452
01:35:28,550 --> 01:35:26,960
for steven freed so

2453
01:35:31,270 --> 01:35:28,560
when you're looking at these domains

2454
01:35:33,030 --> 01:35:31,280
that have different abilities to refold

2455
01:35:35,750 --> 01:35:33,040
uh you mentioned acidity and the

2456
01:35:38,310 --> 01:35:35,760
complexity of these domains but um

2457
01:35:39,990 --> 01:35:38,320
did you also look at any other sort of

2458
01:35:41,590 --> 01:35:40,000
features of the sequence for considering

2459
01:35:42,790 --> 01:35:41,600
what might be driving those differences

2460
01:35:44,870 --> 01:35:42,800
and how well they can reflect maybe

2461
01:35:46,229 --> 01:35:44,880
repeat that with your math tip down

2462
01:35:50,229 --> 01:35:46,239
sorry

2463
01:35:51,430 --> 01:35:50,239

yeah so uh when you're looking at these

2464
01:35:54,470 --> 01:35:51,440
different domains with different

2465
01:35:56,390 --> 01:35:54,480
abilities to refold you can talked about

2466
01:35:58,149 --> 01:35:56,400
acidity and the complexity of them but

2467
01:35:59,910 --> 01:35:58,159
have you considered any other features

2468
01:36:01,830 --> 01:35:59,920
of these sequences that might be driving

2469
01:36:05,109 --> 01:36:01,840
their ability to refold

2470
01:36:07,830 --> 01:36:05,119
so at a very gross level the sequence

2471
01:36:10,709 --> 01:36:07,840
will be reflected in those sort of

2472
01:36:13,430 --> 01:36:10,719
ecod fold groups just because

2473
01:36:15,590 --> 01:36:13,440
as sort of claudia spoke very elegantly

2474
01:36:18,629 --> 01:36:15,600
about the we can sort of use hidden

2475
01:36:20,470 --> 01:36:18,639
markov models in order to group proteins

2476
01:36:22,550 --> 01:36:20,480
together to these sort of lineages that

2477
01:36:24,709 --> 01:36:22,560
will of course have some sequence

2478
01:36:26,790 --> 01:36:24,719
conservation so in that sense when we

2479
01:36:28,790 --> 01:36:26,800
say that ob folds

2480
01:36:30,709 --> 01:36:28,800
always refold we are saying something

2481
01:36:32,709 --> 01:36:30,719
about that you know sort of neighborhood

2482
01:36:34,709 --> 01:36:32,719
of sequence compositions that have that

2483
01:36:36,790 --> 01:36:34,719
property but in terms of like whether or

2484
01:36:38,629 --> 01:36:36,800
not like kind of like a bag of letters

2485
01:36:40,870 --> 01:36:38,639
type of you know analysis of other

2486
01:36:42,070 --> 01:36:40,880
certain amino acids that correlates and

2487
01:36:44,149 --> 01:36:42,080
we haven't done that that would be a

2488
01:36:47,109 --> 01:36:44,159

good thing to do

2489
01:36:51,109 --> 01:36:50,310
hi i'm jason greenwald from etheric uh

2490
01:36:52,470 --> 01:36:51,119
so

2491
01:36:54,390 --> 01:36:52,480
i have to first preface this with saying

2492
01:36:55,430 --> 01:36:54,400
i'm a bit biased because i'm

2493
01:36:56,470 --> 01:36:55,440
very

2494
01:36:58,870 --> 01:36:56,480
much

2495
01:37:00,950 --> 01:36:58,880
in favor of not in any like

2496
01:37:02,390 --> 01:37:00,960
so i really believe it's true but it's

2497
01:37:04,390 --> 01:37:02,400
what i study is

2498
01:37:06,950 --> 01:37:04,400
amyloid peptide aggregation in the

2499
01:37:09,189 --> 01:37:06,960
origin of life and so i have this

2500
01:37:11,430 --> 01:37:09,199
thought that perhaps early

2501
01:37:14,310 --> 01:37:11,440
proteins came out of amyloid structures

2502
01:37:16,470 --> 01:37:14,320
and took um i think it was joanna who

2503
01:37:18,229 --> 01:37:16,480
was making comments about um

2504
01:37:21,830 --> 01:37:18,239
stretches of hydrophobic

2505
01:37:23,669 --> 01:37:21,840
residues being selected against or i'm

2506
01:37:25,510 --> 01:37:23,679
not even sure i remember the detail now

2507
01:37:26,310 --> 01:37:25,520
but i just want to point out that there

2508
01:37:28,229 --> 01:37:26,320
are

2509
01:37:30,950 --> 01:37:28,239
studies and one i remember uh from

2510
01:37:32,390 --> 01:37:30,960
conflict saying that organism complexity

2511
01:37:34,870 --> 01:37:32,400
anti-correlates with the beta

2512
01:37:36,709 --> 01:37:34,880
aggregation propensity of the proteome

2513
01:37:39,750 --> 01:37:36,719

so that being that the more simple

2514
01:37:41,109 --> 01:37:39,760
organisms in principle the older ones uh

2515
01:37:42,950 --> 01:37:41,119
perhaps have more

2516
01:37:45,350 --> 01:37:42,960
propensity in their proteins to be to

2517
01:37:47,990 --> 01:37:45,360
have beta aggregation that sort of fits

2518
01:37:49,590 --> 01:37:48,000
with my not my theory but a theory of

2519
01:37:52,229 --> 01:37:49,600
early proteins coming from

2520
01:37:54,470 --> 01:37:52,239
beta-structured aggregates but also it

2521
01:37:56,550 --> 01:37:54,480
relates to your work stefan i think

2522
01:37:58,550 --> 01:37:56,560
they're saying that if refoldability

2523
01:38:00,950 --> 01:37:58,560
which i think is a super cool idea as a

2524
01:38:03,750 --> 01:38:00,960
potential um

2525
01:38:04,550 --> 01:38:03,760
marker of how old a peptide is a protein

2526
01:38:05,510 --> 01:38:04,560
is

2527
01:38:06,950 --> 01:38:05,520
um

2528
01:38:08,310 --> 01:38:06,960
it might also be

2529
01:38:10,550 --> 01:38:08,320
that there's

2530
01:38:11,669 --> 01:38:10,560
uh some part of the refoldability that

2531
01:38:14,470 --> 01:38:11,679
doesn't necessarily show up in your

2532
01:38:15,990 --> 01:38:14,480
assay because it's aggregation

2533
01:38:18,470 --> 01:38:16,000
that's so sorry that was more like

2534
01:38:20,550 --> 01:38:18,480
blabbing than a question um so i'll make

2535
01:38:22,229 --> 01:38:20,560
one question then and someone else can

2536
01:38:25,750 --> 01:38:22,239
talk if they want to

2537
01:38:29,910 --> 01:38:28,550
yeah you talked about uh that's super

2538
01:38:32,709 --> 01:38:29,920

cool talk by the way i really like that

2539
01:38:34,950 --> 01:38:32,719
kind of work uh where you're replacing

2540
01:38:36,950 --> 01:38:34,960
trying to make primitive looking uh

2541
01:38:38,310 --> 01:38:36,960
protein see um

2542
01:38:40,870 --> 01:38:38,320
but you showed that the fold was

2543
01:38:41,990 --> 01:38:40,880
different i think by the cd right so you

2544
01:38:44,149 --> 01:38:42,000
had something that wasn't folded but

2545
01:38:45,590 --> 01:38:44,159
then you try to model it folded is did i

2546
01:38:47,350 --> 01:38:45,600
miss something or is that sort of a

2547
01:38:48,870 --> 01:38:47,360
little bit out of sync

2548
01:38:50,149 --> 01:38:48,880
with what you expect or do you think it

2549
01:38:52,310 --> 01:38:50,159
may have

2550
01:38:54,550 --> 01:38:52,320
retained some of the same fold

2551
01:38:56,950 --> 01:38:54,560
no actually we were it was kinda we kind

2552
01:38:59,270 --> 01:38:56,960
of expected that the the

2553
01:39:01,510 --> 01:38:59,280
mostly the the

2554
01:39:03,510 --> 01:39:01,520
the variance got like almost 30 percent

2555
01:39:05,350 --> 01:39:03,520
of difference and we can expect it that

2556
01:39:07,189 --> 01:39:05,360
he was supposed to lose the function the

2557
01:39:08,390 --> 01:39:07,199
the structure so

2558
01:39:10,629 --> 01:39:08,400
and uh

2559
01:39:13,189 --> 01:39:10,639
yeah so

2560
01:39:15,109 --> 01:39:13,199
it was kind of possible i can also like

2561
01:39:16,950 --> 01:39:15,119
uh

2562
01:39:19,669 --> 01:39:16,960
it's typical of like this probiotic

2563
01:39:22,310 --> 01:39:19,679

protein of like uh like smoke this

2564
01:39:24,550 --> 01:39:22,320
early alphabet i was also slava show

2565
01:39:26,790 --> 01:39:24,560
before they tend to be more disordered

2566
01:39:28,709 --> 01:39:26,800
so it kind of fit with our

2567
01:39:31,830 --> 01:39:28,719
like with our theory that it's not

2568
01:39:33,270 --> 01:39:31,840
needed no surprise to us like

2569
01:39:36,149 --> 01:39:33,280
thanks i know everyone wants to go home

2570
01:39:37,990 --> 01:39:36,159
but um just one quick question for cloud

2571
01:39:39,270 --> 01:39:38,000
sorry i get the names right here

2572
01:39:41,270 --> 01:39:39,280
claudia

2573
01:39:43,910 --> 01:39:41,280
right you talked about the order i i

2574
01:39:45,990 --> 01:39:43,920
like that concept too of uh how pep

2575
01:39:47,590 --> 01:39:46,000
proteins are evolving their structures

2576
01:39:50,310 --> 01:39:47,600
but can you tell the direction it's

2577
01:39:52,870 --> 01:39:50,320
going i i um geometry kind of lost me

2578
01:39:55,510 --> 01:39:52,880
there can you say it's going from sh3 to

2579
01:39:58,709 --> 01:39:55,520
cradle not reverse

2580
01:40:01,189 --> 01:39:58,719
yeah with these patterns we can have

2581
01:40:03,189 --> 01:40:01,199
some idea of the

2582
01:40:05,910 --> 01:40:03,199
which one is the ancestral fold and

2583
01:40:08,070 --> 01:40:05,920
which one is the daughter fault but not

2584
01:40:09,270 --> 01:40:08,080
all the time so in circular permutation

2585
01:40:15,830 --> 01:40:09,280
it's hard

2586
01:40:19,510 --> 01:40:17,830
i think we're a little bit overdue so

2587
01:40:23,260 --> 01:40:19,520
please join me in thanking all of our

2588
01:40:28,149 --> 01:40:25,830

[Applause]

2589
01:40:31,669 --> 01:40:28,159
and hopefully there's a chance to chat