EARTH-LIFE
SCIENCE
INSTITUTE

Vikas Nanda

Electron Transfer in Proteins: the Minimal Elements & How They are Connected

1
00:00:11,590 --> 00:00:09,180
[Music]

2
00:00:14,470 --> 00:00:11,600
all right so I want to thank everyone

3
00:00:17,230 --> 00:00:14,480
for the for being here

4
00:00:19,540 --> 00:00:17,240
I'm coming from the medical school at

5
00:00:20,859 --> 00:00:19,550
Rutgers University and I want to assure

6
00:00:22,300 --> 00:00:20,869
you that I actually have a sincere

7
00:00:24,550 --> 00:00:22,310
interest in the origins of life and I

8
00:00:29,050 --> 00:00:24,560
didn't just want a free trip to Japan

9
00:00:42,910 --> 00:00:29,060
and I've actually you know as part of

10
00:00:43,900 --> 00:00:42,920
the medical school as part of the

11
00:00:50,890 --> 00:00:43,910
medical school I don't know how to

12
00:00:52,150 --> 00:00:50,900
operate a computer let's see so as part

13
00:00:54,520 --> 00:00:52,160

of the medical school one of the things

14
00:00:57,100 --> 00:00:54,530
that I do is develop drugs and we're

15
00:00:59,230 --> 00:00:57,110
very interested in actually using

16
00:01:01,990 --> 00:00:59,240
structural modeling of proteins as an

17
00:01:04,359 --> 00:01:02,000
approach to designing new therapeutic

18
00:01:06,820 --> 00:01:04,369
molecules but at the same time you know

19
00:01:09,760 --> 00:01:06,830
my in my heart of hearts my fundamental

20
00:01:12,310 --> 00:01:09,770
interest is in origins of life and how

21
00:01:13,179 --> 00:01:12,320
proteins fold and how they evolve and

22
00:01:15,399 --> 00:01:13,189
one of the things that we've been

23
00:01:18,550 --> 00:01:15,409
interested in I've been studying now for

24
00:01:21,520 --> 00:01:18,560
over 15 years is the emergence of Homo

25
00:01:23,499 --> 00:01:21,530
chirality and I want to thank professor

26
00:01:25,599 --> 00:01:23,509
Whitesides for introducing this idea of

27
00:01:27,399 --> 00:01:25,609
the Leventhal it's paradox because for

28
00:01:30,309 --> 00:01:27,409
me protein folding is an important

29
00:01:33,190 --> 00:01:30,319
obstacle for for biomolecules to

30
00:01:34,300 --> 00:01:33,200
overcome and in origins and so one of

31
00:01:36,730 --> 00:01:34,310
the things that we were studying is

32
00:01:39,340 --> 00:01:36,740
using our computer simulations why our

33
00:01:41,050 --> 00:01:39,350
proteins homo chiral and what we find is

34
00:01:42,940 --> 00:01:41,060
that when sequences are homo chiral

35
00:01:45,129 --> 00:01:42,950
they're able to fold more quickly

36
00:01:46,629 --> 00:01:45,139
they're able to find their folded state

37
00:01:48,669 --> 00:01:46,639
in a much smaller phase space this is

38
00:01:50,080 --> 00:01:48,679

the phase space for example of a homo

39
00:01:51,910 --> 00:01:50,090
karl sequence whereas if your hetero

40
00:01:53,499 --> 00:01:51,920
karl you have to sample a lot more

41
00:01:55,330 --> 00:01:53,509
confirmations in order to find a unique

42
00:01:56,709 --> 00:01:55,340
state and so for us that was very

43
00:01:58,779 --> 00:01:56,719
interesting because with a very simple

44
00:02:00,520 --> 00:01:58,789
simulation we could see some very

45
00:02:02,499 --> 00:02:00,530
fundamental properties of molecules and

46
00:02:04,359 --> 00:02:02,509
then once we realized what we're sort of

47
00:02:06,760 --> 00:02:04,369
the guiding forces that were driving

48
00:02:08,529 --> 00:02:06,770
homo chirality we were then able to use

49
00:02:10,570 --> 00:02:08,539
that to sort of violate those principles

50
00:02:13,720 --> 00:02:10,580
and start to design small peptides for

51
00:02:16,240 --> 00:02:13,730
example this one this is a peptide made

52
00:02:17,830 --> 00:02:16,250
out of all L amino acids except for the

53
00:02:19,929 --> 00:02:17,840
NNC term and I have the peptide where

54
00:02:21,280 --> 00:02:19,939
there are D amino acids that cap the

55
00:02:22,030 --> 00:02:21,290
ends of the peptides and hold them in a

56
00:02:23,800 --> 00:02:22,040
fold it's

57
00:02:25,750 --> 00:02:23,810
and this peptide even though it's only a

58
00:02:27,520 --> 00:02:25,760
Timmy no assets long is very stable it's

59
00:02:29,590 --> 00:02:27,530
able to find its native state very

60
00:02:31,150 --> 00:02:29,600
quickly and very efficiently and then

61
00:02:33,160 --> 00:02:31,160
block an interaction that's key in the

62
00:02:35,260 --> 00:02:33,170
life cycle of the the influenza virus

63
00:02:38,110 --> 00:02:35,270

and we're now developing this as a

64
00:02:39,700 --> 00:02:38,120
potential therapeutic peptide so even

65
00:02:41,290 --> 00:02:39,710
though I'm at the medical school and I'm

66
00:02:43,510 --> 00:02:41,300
studying origins of life there really is

67
00:02:47,110 --> 00:02:43,520
a nice connection between these two

68
00:02:49,780 --> 00:02:47,120
sides of my life and so when I first

69
00:02:53,020 --> 00:02:49,790
came to the Medical School about 11 or

70
00:02:55,600 --> 00:02:53,030
12 years ago I was very excited to to

71
00:02:57,520 --> 00:02:55,610
meet Paul Falkowski because he got me

72
00:02:59,530 --> 00:02:57,530
engaged in another origins of life

73
00:03:01,270 --> 00:02:59,540
problem different from homo chirality

74
00:03:03,070 --> 00:03:01,280
and so this is the project I'm going to

75
00:03:06,730 --> 00:03:03,080
talk to you about today it's primarily

76
00:03:10,170 --> 00:03:06,740
the the work of a postdoc in the

77
00:03:13,120 --> 00:03:10,180
laboratory a guy a guy came to us from a

78
00:03:15,070 --> 00:03:13,130
microbial physiology lab in Israel and

79
00:03:16,450 --> 00:03:15,080
he very quickly became a structural

80
00:03:18,640 --> 00:03:16,460
biologist so I'm very proud of the work

81
00:03:20,770 --> 00:03:18,650
that he's done and this is also part of

82
00:03:23,170 --> 00:03:20,780
a larger team Douglas Pike a valuable

83
00:03:25,120 --> 00:03:23,180
graduate student in the lab postdocs Eli

84
00:03:27,210 --> 00:03:25,130
Moore and Stephan sin and in also

85
00:03:29,620 --> 00:03:27,220
collaborations with Yana Bromberg who

86
00:03:31,150 --> 00:03:29,630
you've heard about some of her work

87
00:03:35,910 --> 00:03:31,160
earlier and then support from a number

88
00:03:39,040 --> 00:03:35,920

of foundations and and associations so

89
00:03:41,650 --> 00:03:39,050
Paul when I first arted talking to him

90
00:03:44,740 --> 00:03:41,660
about this essentially was it was very

91
00:03:47,850 --> 00:03:44,750
interested in the origins of these

92
00:03:49,930 --> 00:03:47,860
complex protein nanomachines these

93
00:03:54,270 --> 00:03:49,940
oxidoreductases that were capable of

94
00:03:56,830 --> 00:03:54,280
doing very very efficiently

95
00:03:59,260 --> 00:03:56,840
sophisticated a couple electron transfer

96
00:04:00,580 --> 00:03:59,270
catalysis and these we call these

97
00:04:02,740 --> 00:04:00,590
proteins nano machines because they are

98
00:04:05,710 --> 00:04:02,750
massive macromolecular complexes and

99
00:04:08,290 --> 00:04:05,720
these are critical reactions and at some

100
00:04:09,490 --> 00:04:08,300
point you know they were they must have

101
00:04:11,710 --> 00:04:09,500
emerged because they were critical for

102
00:04:13,240 --> 00:04:11,720
for lies processes but we couldn't

103
00:04:15,130 --> 00:04:13,250
imagine something like a nitrogenase

104
00:04:17,140 --> 00:04:15,140
climbing out of the primordial soup on

105
00:04:19,599 --> 00:04:17,150
its own there had to be some kind of a

106
00:04:21,580 --> 00:04:19,609
simpler ancestor and we're really

107
00:04:23,770 --> 00:04:21,590
interested in what these ancestors look

108
00:04:25,930 --> 00:04:23,780
like and maybe as you walk back they may

109
00:04:27,850 --> 00:04:25,940
even have been small peptides in the

110
00:04:30,310 --> 00:04:27,860
earlier Keyon or late hid Ian that had

111
00:04:32,980 --> 00:04:30,320
similar function that sort of themselves

112
00:04:35,590 --> 00:04:32,990
emerge from the the primordial soup but

113
00:04:37,030 --> 00:04:35,600

as we all know the problem is we

114
00:04:39,160 --> 00:04:37,040
don't have any information about what

115
00:04:41,530 --> 00:04:39,170
these molecules look like there are no

116
00:04:44,050 --> 00:04:41,540
molecular fossils for what happened at

117
00:04:46,960 --> 00:04:44,060
these early stages all we have are the

118
00:04:48,670 --> 00:04:46,970
the extant molecules and then some

119
00:04:50,020 --> 00:04:48,680
geological information some

120
00:04:52,540 --> 00:04:50,030
mineralogical information about what the

121
00:04:54,730 --> 00:04:52,550
rocks may have look like at that time so

122
00:04:58,710 --> 00:04:54,740
the question is how can we sort of walk

123
00:05:01,990 --> 00:04:58,720
back the common ancestors of these much

124
00:05:04,210 --> 00:05:02,000
these modern extant complex proteins and

125
00:05:09,970 --> 00:05:04,220
figure out what these original molecules

126
00:05:12,190 --> 00:05:09,980
may look like so uh you know I call

127
00:05:13,390 --> 00:05:12,200
these things uh massive nano machines

128
00:05:15,220 --> 00:05:13,400
because when you look at some of these

129
00:05:17,440 --> 00:05:15,230
enzymes they really are incredibly

130
00:05:19,030 --> 00:05:17,450
complex right so these are you know

131
00:05:20,980 --> 00:05:19,040
they're they're doing critical reactions

132
00:05:24,040 --> 00:05:20,990
that you know as we've heard for many of

133
00:05:27,370 --> 00:05:24,050
the talks throughout the session they

134
00:05:29,980 --> 00:05:27,380
take advantage of this extant

135
00:05:32,860 --> 00:05:29,990
disequilibrium and on the planet and

136
00:05:34,450 --> 00:05:32,870
it's a great source of energy but you

137
00:05:37,570 --> 00:05:34,460
can't imagine you know a protein like

138
00:05:40,690 --> 00:05:37,580

this emerging spontaneously and again

139
00:05:42,580 --> 00:05:40,700
this is a manifestation of the the

140
00:05:45,190 --> 00:05:42,590
Leventhal paradox so you know think

141
00:05:46,210 --> 00:05:45,200
about something much simpler a protein

142
00:05:48,670 --> 00:05:46,220
that's maybe more medically relevant

143
00:05:51,520 --> 00:05:48,680
this is the the beta chain of insulin

144
00:05:53,980 --> 00:05:51,530
it's only about 30 amino acids right and

145
00:05:56,110 --> 00:05:53,990
so we know that forward and before for a

146
00:05:58,750 --> 00:05:56,120
protein to fold it has to go from this

147
00:06:01,500 --> 00:05:58,760
unfolded state into a single native

148
00:06:04,660 --> 00:06:01,510
state and each amino acid in this chain

149
00:06:05,980 --> 00:06:04,670
has two rotatable bonds and if you think

150
00:06:07,810 --> 00:06:05,990
about sort of the staggered eclipsed

151
00:06:09,850 --> 00:06:07,820
confirmations of bonds you could

152
00:06:11,620 --> 00:06:09,860
minimally have ten confirmations per

153
00:06:13,960 --> 00:06:11,630
amino acid so you're talking about a

154
00:06:16,570 --> 00:06:13,970
total of combinatoric Li a total of

155
00:06:18,310 --> 00:06:16,580
about 10 to the 30th power unfolded

156
00:06:20,140 --> 00:06:18,320
confirmations out of which it finds one

157
00:06:22,090 --> 00:06:20,150
folded state so even for something as

158
00:06:25,750 --> 00:06:22,100
small as the beta peptide of insulin

159
00:06:28,840 --> 00:06:25,760
this is a massive problem if you think

160
00:06:31,810 --> 00:06:28,850
about the the other sort of side of the

161
00:06:34,480 --> 00:06:31,820
xi pause paradox now this is how did

162
00:06:36,190 --> 00:06:34,490
that particular protein evolve you can

163
00:06:37,630 --> 00:06:36,200

also see that this is a massive common

164
00:06:39,700 --> 00:06:37,640
in toriel problem so even for something

165
00:06:42,130 --> 00:06:39,710
this small how do you find a specific

166
00:06:44,200 --> 00:06:42,140
sequence that has this function from the

167
00:06:46,180 --> 00:06:44,210
20 to the 30th power unique sequences

168
00:06:48,700 --> 00:06:46,190
right so even if the functional

169
00:06:49,480 --> 00:06:48,710
footprint of something that acts like a

170
00:06:51,520 --> 00:06:49,490
beta

171
00:06:53,140 --> 00:06:51,530
insulin peptide is let's say there's a

172
00:06:54,610 --> 00:06:53,150
billion sequences that could do that or

173
00:06:59,439 --> 00:06:54,620
a trillion sequences that could do that

174
00:07:02,409 --> 00:06:59,449
you still have a huge phase space to to

175
00:07:03,670 --> 00:07:02,419
to search in order to find a sequence

176
00:07:05,050 --> 00:07:03,680
that's going to have this function and

177
00:07:07,390 --> 00:07:05,060
this is something that's only thirty

178
00:07:10,300 --> 00:07:07,400
amino acids long if you look at

179
00:07:13,870 --> 00:07:10,310
something like nitrogenase at rajan YZ

180
00:07:17,350 --> 00:07:13,880
is about 2500 amino acids and so how

181
00:07:19,150 --> 00:07:17,360
does something this complex evolve when

182
00:07:21,040 --> 00:07:19,160
you have such a massive conformational

183
00:07:24,010 --> 00:07:21,050
space to search in such a massive

184
00:07:27,210 --> 00:07:24,020
sequence base and the answer you know

185
00:07:29,020 --> 00:07:27,220
this is not something that's new to

186
00:07:31,719 --> 00:07:29,030
oxidoreductases are new to this this

187
00:07:33,040 --> 00:07:31,729
particular this particular project but

188
00:07:35,529 --> 00:07:33,050

we know that proteins really are

189
00:07:36,310 --> 00:07:35,539
assembled by much smaller domains and so

190
00:07:38,379 --> 00:07:36,320
there was a couple of ways that

191
00:07:40,270 --> 00:07:38,389
nitrogenase solves this problem one is

192
00:07:41,680 --> 00:07:40,280
that there are multiple proteins that

193
00:07:44,560 --> 00:07:41,690
associate to form the macromolecular

194
00:07:45,820 --> 00:07:44,570
complex there's symmetry that a lot of

195
00:07:48,939 --> 00:07:45,830
proteins take advantage of so you can

196
00:07:52,149 --> 00:07:48,949
double or triple or or multiply your

197
00:07:54,700 --> 00:07:52,159
your complexity just by taking advantage

198
00:07:56,050 --> 00:07:54,710
of symmetric transformations and then

199
00:07:57,969 --> 00:07:56,060
the problem becomes a lot simpler now

200
00:08:00,399 --> 00:07:57,979
you think about how did these individual

201
00:08:02,620 --> 00:08:00,409
modules evolve and the phase space that

202
00:08:04,870 --> 00:08:02,630
they have to search is significantly

203
00:08:07,120 --> 00:08:04,880
smaller so I say that the solution is

204
00:08:09,010 --> 00:08:07,130
you know let's identify what these these

205
00:08:11,560 --> 00:08:09,020
individual building blocks are and they

206
00:08:13,570 --> 00:08:11,570
would be a much simpler thing to imagine

207
00:08:16,719 --> 00:08:13,580
evolving but that's not a very easy

208
00:08:19,540 --> 00:08:16,729
thing to do right we can't just sort of

209
00:08:20,860 --> 00:08:19,550
carve out pieces of nitrogen's and say

210
00:08:23,140 --> 00:08:20,870
that this piece of Auld and then this

211
00:08:25,209 --> 00:08:23,150
piece of all then these are our modules

212
00:08:26,680 --> 00:08:25,219
this identification of these ancient

213
00:08:29,140 --> 00:08:26,690

building blocks is itself a very

214
00:08:30,640 --> 00:08:29,150
difficult problem and then also once you

215
00:08:32,319 --> 00:08:30,650
identify what these domains are what

216
00:08:33,760 --> 00:08:32,329
these modules are I think another

217
00:08:35,740 --> 00:08:33,770
critical problem is figuring out how did

218
00:08:37,800 --> 00:08:35,750
they self assemble how did they they

219
00:08:44,769 --> 00:08:37,810
aggregate to make these more complex

220
00:08:46,090 --> 00:08:44,779
functional molecules so the way that we

221
00:08:47,170 --> 00:08:46,100
approached this problem was to say you

222
00:08:48,699 --> 00:08:47,180
know if you look at something like an

223
00:08:52,150 --> 00:08:48,709
oxidoreductase so this is fumarate

224
00:08:53,560 --> 00:08:52,160
reductase for example let's ignore most

225
00:08:55,329 --> 00:08:53,570
of the protein and let's really just

226
00:08:56,829 --> 00:08:55,339
look at the important part of this

227
00:08:59,350 --> 00:08:56,839
protein the one that's involved an

228
00:09:01,750 --> 00:08:59,360
electron transfer and electron transfer

229
00:09:03,310 --> 00:09:01,760
is really just mediated by this chain of

230
00:09:05,620 --> 00:09:03,320
metals this little necklace of

231
00:09:09,540 --> 00:09:05,630
that's running through the center of the

232
00:09:12,250 --> 00:09:09,550
the protein core to its active site and

233
00:09:13,870 --> 00:09:12,260
if we argue that these are really the

234
00:09:15,880 --> 00:09:13,880
the functionally important parts of the

235
00:09:17,800 --> 00:09:15,890
protein then we would imagine that the

236
00:09:19,870 --> 00:09:17,810
fundamental modules that assemble into

237
00:09:21,760 --> 00:09:19,880
these larger molecules must be centered

238
00:09:24,280 --> 00:09:21,770

around these metals so essentially what

239
00:09:25,990 --> 00:09:24,290
we did was we went into a database

240
00:09:27,940 --> 00:09:26,000
called a protein databank and for those

241
00:09:31,060 --> 00:09:27,950
of you who are not a familiar with this

242
00:09:32,950 --> 00:09:31,070
this data set the the PDB is a

243
00:09:35,170 --> 00:09:32,960
repository for the high-resolution

244
00:09:38,260 --> 00:09:35,180
structures of protein so these are

245
00:09:41,260 --> 00:09:38,270
atomic resolution structures of proteins

246
00:09:42,310 --> 00:09:41,270
not just oxidoreductases but all kinds

247
00:09:45,460 --> 00:09:42,320
of different proteins you know

248
00:09:47,500 --> 00:09:45,470
hemoglobin and and so forth and there's

249
00:09:49,480 --> 00:09:47,510
over a hundred thousand different

250
00:09:51,720 --> 00:09:49,490
proteins that have been deposited in the

251
00:09:55,990 --> 00:09:51,730
PDB XI now I think closer to about

252
00:09:58,450 --> 00:09:56,000
150,000 and of these about 10,000 or so

253
00:09:59,740 --> 00:09:58,460
have metal centers in them and what we

254
00:10:02,470 --> 00:09:59,750
did was we essentially took all of those

255
00:10:03,940 --> 00:10:02,480
proteins and we looked at what we call a

256
00:10:05,950 --> 00:10:03,950
micro environment which is essentially

257
00:10:07,300 --> 00:10:05,960
just the amino acids that are within a

258
00:10:08,950 --> 00:10:07,310
certain distance of the metal center

259
00:10:11,590 --> 00:10:08,960
which we think is the important part of

260
00:10:13,420 --> 00:10:11,600
the protein for electron transfer and we

261
00:10:15,280 --> 00:10:13,430
just we just excavated all of these out

262
00:10:17,170 --> 00:10:15,290
of these proteins so we had about 30,000

263
00:10:19,600 --> 00:10:17,180

micro environments and then we tried to

264
00:10:22,270 --> 00:10:19,610
classify these into a smaller set of

265
00:10:23,410 --> 00:10:22,280
modules based on their metal type and

266
00:10:28,600 --> 00:10:23,420
then also based on some sort of

267
00:10:30,610 --> 00:10:28,610
structural similarity and you know for

268
00:10:32,080 --> 00:10:30,620
those of you who have done comparative

269
00:10:34,600 --> 00:10:32,090
structural analysis of proteins you know

270
00:10:36,280 --> 00:10:34,610
that this is not a trivial thing to do

271
00:10:37,990 --> 00:10:36,290
particularly when you're looking at

272
00:10:40,390 --> 00:10:38,000
alignments that are of sort of

273
00:10:43,300 --> 00:10:40,400
intermediate quality so for example what

274
00:10:45,730 --> 00:10:43,310
I'm showing you here these are two two

275
00:10:48,370 --> 00:10:45,740
modules that are centered around an iron

276
00:10:49,600 --> 00:10:48,380
sulfur cluster these are about 15

277
00:10:51,820 --> 00:10:49,610
angstroms and radius we're essentially

278
00:10:54,670 --> 00:10:51,830
looking at all the amino acids that are

279
00:10:56,440 --> 00:10:54,680
within 15 and 15 angstroms of that metal

280
00:10:58,210 --> 00:10:56,450
center and we're what we want to ask

281
00:11:00,820 --> 00:10:58,220
whether they are they have similar

282
00:11:03,250 --> 00:11:00,830
protein structure holding the the metal

283
00:11:04,810 --> 00:11:03,260
in place and these are two examples of

284
00:11:06,520 --> 00:11:04,820
the types of alignments we could get and

285
00:11:08,140 --> 00:11:06,530
on this axis right here we have a

286
00:11:10,120 --> 00:11:08,150
similarity score this is essentially

287
00:11:13,540 --> 00:11:10,130
telling us how well do the atoms of

288
00:11:15,910 --> 00:11:13,550

those two environments align and you can

289
00:11:17,049 --> 00:11:15,920
see here in this alignment here the red

290
00:11:18,459 --> 00:11:17,059
and orange parts

291
00:11:19,719 --> 00:11:18,469
these are the parts that align you can

292
00:11:22,239 --> 00:11:19,729
really see only a little bit of the

293
00:11:23,969 --> 00:11:22,249
structural lines right here and then

294
00:11:26,469 --> 00:11:23,979
here's another alignment between two

295
00:11:28,569 --> 00:11:26,479
modules and you can see again only maybe

296
00:11:30,609 --> 00:11:28,579
about 20-30 percent of those two

297
00:11:33,009 --> 00:11:30,619
structures align and so they have about

298
00:11:34,509 --> 00:11:33,019
the same similarity score and so if we

299
00:11:36,579 --> 00:11:34,519
were just doing standard structural

300
00:11:37,899 --> 00:11:36,589
alignment tools we would not be able to

301
00:11:38,889 --> 00:11:37,909
say which one is a good alignment which

302
00:11:42,039 --> 00:11:38,899
one is the battle I mean they're both

303
00:11:44,469 --> 00:11:42,049
sort of on the edge of being an

304
00:11:46,239 --> 00:11:44,479
acceptable alignment but what we know is

305
00:11:48,189 --> 00:11:46,249
that in order for these these domains to

306
00:11:49,359 --> 00:11:48,199
function they must contain this metal

307
00:11:52,059 --> 00:11:49,369
centre and this metal centre is really

308
00:11:54,609 --> 00:11:52,069
the sort of the functional nucleus of

309
00:11:57,549 --> 00:11:54,619
these these modules so really those

310
00:11:58,869 --> 00:11:57,559
those metal centers must also align so

311
00:12:02,259 --> 00:11:58,879
we're using the metal centre essentially

312
00:12:04,089 --> 00:12:02,269
as a fiducial marker to tell us how good

313
00:12:06,579 --> 00:12:04,099

our alignments are so even though these

314
00:12:07,779 --> 00:12:06,589
have very similar scores here the metals

315
00:12:09,609 --> 00:12:07,789
are right on top of each other and we

316
00:12:10,779 --> 00:12:09,619
believe this alignment and here they're

317
00:12:13,029 --> 00:12:10,789
far apart from each other and we don't

318
00:12:14,409 --> 00:12:13,039
so this was a real breakthrough for us

319
00:12:16,479 --> 00:12:14,419
because it allowed us then to do this

320
00:12:18,399 --> 00:12:16,489
structure structure comparison a large

321
00:12:19,869 --> 00:12:18,409
scale and not have to go through and

322
00:12:23,399 --> 00:12:19,879
analyze each one manually and figure out

323
00:12:25,569 --> 00:12:23,409
whether we believe the alignment or not

324
00:12:28,509 --> 00:12:25,579
and so we did this for all of these

325
00:12:30,429 --> 00:12:28,519
30,000 modules and what we found is that

326
00:12:32,949 --> 00:12:30,439
there were about let's say a thousand

327
00:12:35,199 --> 00:12:32,959
different modules and we could cluster

328
00:12:37,389 --> 00:12:35,209
all of these into these these different

329
00:12:39,789 --> 00:12:37,399
classes and that number a thousand is

330
00:12:41,889 --> 00:12:39,799
not an exact number depending on what

331
00:12:43,179 --> 00:12:41,899
your threshold is for similarity you can

332
00:12:45,009 --> 00:12:43,189
make it larger you can make it smaller

333
00:12:48,339 --> 00:12:45,019
but we do find that there are a couple

334
00:12:49,899 --> 00:12:48,349
of domains that are that are a couple of

335
00:12:51,609 --> 00:12:49,909
modules that have a lot of members and

336
00:12:53,679 --> 00:12:51,619
within these we have the ferredoxin

337
00:12:55,539 --> 00:12:53,689
which is not surprising the cytochrome C

338
00:12:57,609 --> 00:12:55,549

but then also a copper binding

339
00:12:59,309 --> 00:12:57,619
plastocyanin and then a four helix

340
00:13:02,079 --> 00:12:59,319
bundle that could either bind one or two

341
00:13:03,909 --> 00:13:02,089
metal ions in the center so it looks

342
00:13:06,249 --> 00:13:03,919
like we have for example we have here a

343
00:13:12,549 --> 00:13:06,259
couple of Legos that are commonly used

344
00:13:15,369 --> 00:13:12,559
in in metalloproteins now what we're

345
00:13:16,689 --> 00:13:15,379
looking at here this is essentially you

346
00:13:18,669 --> 00:13:16,699
know the way that we're defining these

347
00:13:20,139 --> 00:13:18,679
micro environments is the metal in the

348
00:13:22,389 --> 00:13:20,149
center and then we're sort of carving

349
00:13:24,519 --> 00:13:22,399
out amino acids that are within 15

350
00:13:26,619 --> 00:13:24,529
angstroms that metal and sometimes this

351
00:13:29,349 --> 00:13:26,629
is a discontinuous piece of the protein

352
00:13:31,030 --> 00:13:29,359
may have loops that are going out to

353
00:13:33,310 --> 00:13:31,040
another domain or maybe this part of the

354
00:13:34,269 --> 00:13:33,320
proteins coming from one chain and this

355
00:13:36,009 --> 00:13:34,279
part of the proteins coming from another

356
00:13:37,569 --> 00:13:36,019
part of the chain so why would you

357
00:13:40,090 --> 00:13:37,579
believe that this is actually a relevant

358
00:13:42,400 --> 00:13:40,100
module for for evolution so one of the

359
00:13:44,920 --> 00:13:42,410
things that we noticed was that when we

360
00:13:46,269 --> 00:13:44,930
look at the size distribution of these

361
00:13:49,389 --> 00:13:46,279
modules and we put them on a log-log

362
00:13:52,300 --> 00:13:49,399
plot we see that they have a semi linear

363
00:13:55,300 --> 00:13:52,310

relationship and that is consistent with

364
00:13:57,819 --> 00:13:55,310
a model of domain evolution where you

365
00:13:59,410 --> 00:13:57,829
have duplication of these modules so you

366
00:14:00,879 --> 00:13:59,420
can imagine that you have old modules

367
00:14:03,970 --> 00:14:00,889
that have been around for a long time

368
00:14:05,710 --> 00:14:03,980
and they Vedad a long time to duplicate

369
00:14:07,990 --> 00:14:05,720
within genomes and so there's a lot of

370
00:14:09,999 --> 00:14:08,000
copies of those and then you have at the

371
00:14:11,860 --> 00:14:10,009
same time innovation you have new

372
00:14:13,900 --> 00:14:11,870
domains that are being invented and

373
00:14:15,939 --> 00:14:13,910
those exist at you know at a much

374
00:14:18,180 --> 00:14:15,949
smaller fraction and so when you have

375
00:14:20,829 --> 00:14:18,190
this sort of a process of domain

376
00:14:23,160 --> 00:14:20,839
innovation and then duplication you get

377
00:14:27,069 --> 00:14:23,170
this sort of linear relationship of

378
00:14:29,410 --> 00:14:27,079
module size versus frequency on a

379
00:14:31,210 --> 00:14:29,420
log-log plot so this made this gave us

380
00:14:33,970 --> 00:14:31,220
some confidence that even though we are

381
00:14:35,980 --> 00:14:33,980
sort of creating these shaved pieces of

382
00:14:37,949 --> 00:14:35,990
proteins that this was a functionally

383
00:14:40,629 --> 00:14:37,959
relevant and evolutionarily selectable

384
00:14:44,790 --> 00:14:40,639
domain that we could then think about in

385
00:14:47,920 --> 00:14:44,800
terms of its its functional consequences

386
00:14:49,480 --> 00:14:47,930
so now the question is some of these

387
00:14:53,590 --> 00:14:49,490
domains for example like cytochrome C

388
00:14:54,850 --> 00:14:53,600

contain hundreds or close to a thousand

389
00:14:57,280 --> 00:14:54,860
different micro environments extracted

390
00:14:59,949 --> 00:14:57,290
from the PDB am I saying that all of

391
00:15:02,230 --> 00:14:59,959
these modules have a common origin

392
00:15:05,590 --> 00:15:02,240
they're all came from a from an ERV

393
00:15:06,610 --> 00:15:05,600
cytochrome C type domain and now we're

394
00:15:08,920 --> 00:15:06,620
seeing them occurring in all these

395
00:15:10,900 --> 00:15:08,930
different proteins well that's that's

396
00:15:12,819 --> 00:15:10,910
pretty hard to believe so for example

397
00:15:15,490 --> 00:15:12,829
here we're looking at just one of these

398
00:15:17,559 --> 00:15:15,500
sets of modules so each of these dots

399
00:15:20,199 --> 00:15:17,569
represents one microenvironment from a

400
00:15:22,300 --> 00:15:20,209
specific protein and an edge represents

401
00:15:23,679 --> 00:15:22,310
two micro environments that are that

402
00:15:25,990 --> 00:15:23,689
have an acceptable alignment to each

403
00:15:28,540 --> 00:15:26,000
other and so for example to get from

404
00:15:30,699 --> 00:15:28,550
this micro environment right here from

405
00:15:32,319 --> 00:15:30,709
one protein to this micro environment

406
00:15:34,179 --> 00:15:32,329
right here from another protein we would

407
00:15:35,499 --> 00:15:34,189
have to go through minimally 12

408
00:15:37,059 --> 00:15:35,509
different intermediates to get from

409
00:15:39,220 --> 00:15:37,069
there to there and we're not just

410
00:15:41,679 --> 00:15:39,230
staying within prokaryotes who may be

411
00:15:43,870 --> 00:15:41,689
going in between to a eukaryotic module

412
00:15:46,410 --> 00:15:43,880
and then back to a prokaryotic module so

413
00:15:48,519 --> 00:15:46,420

clearly these are not

414
00:15:51,040 --> 00:15:48,529
convincing evolutionary trajectories

415
00:15:52,150 --> 00:15:51,050
based on structural similarity alone so

416
00:15:53,639 --> 00:15:52,160
another way of thinking about this I

417
00:15:56,500 --> 00:15:53,649
think about are we discriminating

418
00:15:57,850 --> 00:15:56,510
homology versus analogy and really what

419
00:16:00,430 --> 00:15:57,860
I'm saying is that for example within a

420
00:16:02,769 --> 00:16:00,440
particular module like the ferredoxin or

421
00:16:04,449 --> 00:16:02,779
the cytochrome C with similarity alone

422
00:16:06,759 --> 00:16:04,459
we don't know if we were looking at all

423
00:16:08,710 --> 00:16:06,769
bird wings or whether there's bat wings

424
00:16:11,290 --> 00:16:08,720
and butterfly wings mixed in to this

425
00:16:13,030 --> 00:16:11,300
data set so that's an important thing to

426
00:16:15,819 --> 00:16:13,040
keep in mind that structural similarity

427
00:16:19,949 --> 00:16:15,829
itself is not sufficient to to prove

428
00:16:26,769 --> 00:16:24,759
okay so we have a thousand or so modules

429
00:16:29,380 --> 00:16:26,779
that are being used to build these more

430
00:16:31,269 --> 00:16:29,390
complex proteins now you know what's

431
00:16:33,490 --> 00:16:31,279
really interesting we want to understand

432
00:16:34,540 --> 00:16:33,500
the emergence of complexity within these

433
00:16:38,829 --> 00:16:34,550
systems if we want to say are there any

434
00:16:40,480 --> 00:16:38,839
rules that can guide how these modules

435
00:16:42,430 --> 00:16:40,490
are connected together can we figure out

436
00:16:45,130 --> 00:16:42,440
the rules for how these these legos are

437
00:16:46,960 --> 00:16:45,140
assembled and for this we take advantage

438
00:16:48,880 --> 00:16:46,970

of the fact that we are really

439
00:16:50,050 --> 00:16:48,890
interested in electron transfer

440
00:16:53,230 --> 00:16:50,060
you know we're interested in the ability

441
00:16:56,250 --> 00:16:53,240
of proteins to to move electrons from

442
00:17:00,010 --> 00:16:56,260
one side of the protein to another from

443
00:17:01,150 --> 00:17:00,020
from a an active side to a to a

444
00:17:04,179 --> 00:17:01,160
different part of the protein and

445
00:17:06,250 --> 00:17:04,189
they're what we can take advantage of is

446
00:17:08,829 --> 00:17:06,260
that since we have the high resolution

447
00:17:10,449 --> 00:17:08,839
structures for all of these proteins we

448
00:17:12,909 --> 00:17:10,459
know the distance between each of the

449
00:17:15,850 --> 00:17:12,919
metal cofactors and there was a very

450
00:17:19,590 --> 00:17:15,860
important and influential study from Les

451
00:17:23,350 --> 00:17:19,600
Sutton's lab in the early 2000s where

452
00:17:25,449 --> 00:17:23,360
they looked at a set of oxidoreductases

453
00:17:27,970 --> 00:17:25,459
and they looked at the distances between

454
00:17:30,280 --> 00:17:27,980
pairs of metal cofactors that were

455
00:17:32,590 --> 00:17:30,290
involved in electron transfer and what

456
00:17:34,690 --> 00:17:32,600
you can see here is on this plot here

457
00:17:37,180 --> 00:17:34,700
you have the distance on this axis

458
00:17:38,770 --> 00:17:37,190
between two metal sites within a within

459
00:17:40,930 --> 00:17:38,780
a protein in an electron transport chain

460
00:17:44,159 --> 00:17:40,940
and then on this axis right here the log

461
00:17:46,780 --> 00:17:44,169
of the electron transfer rate right and

462
00:17:49,750 --> 00:17:46,790
for all electron transport chains within

463
00:17:51,580 --> 00:17:49,760

proteins the metal cofactors are at most

464
00:17:53,080 --> 00:17:51,590
found with by fourteen angstroms away

465
00:17:54,070 --> 00:17:53,090
from each other and at fourteen

466
00:17:56,680 --> 00:17:54,080
angstroms you're now thinking about

467
00:17:59,050 --> 00:17:56,690
electron transfer rates on the on the

468
00:18:00,730 --> 00:17:59,060
scale of microseconds any further than

469
00:18:02,800 --> 00:18:00,740
that then these these the electron

470
00:18:06,010 --> 00:18:02,810
transfer rates become too slow to really

471
00:18:07,750 --> 00:18:06,020
be biologically relevant and so what we

472
00:18:10,120 --> 00:18:07,760
said was that well if we were interested

473
00:18:13,450 --> 00:18:10,130
in electron transport chains we can then

474
00:18:15,550 --> 00:18:13,460
just look for modules where the distance

475
00:18:17,740 --> 00:18:15,560
between cofactors falls within this this

476
00:18:18,940 --> 00:18:17,750
distance cutoff so really what we're

477
00:18:21,100 --> 00:18:18,950
doing now is we're going through the

478
00:18:22,390 --> 00:18:21,110
same data set of proteins and now

479
00:18:24,670 --> 00:18:22,400
instead of collect connecting

480
00:18:26,770 --> 00:18:24,680
microenvironments based on structural

481
00:18:28,540 --> 00:18:26,780
similarity we're connecting them based

482
00:18:30,370 --> 00:18:28,550
on their spatial adjacency within a

483
00:18:32,350 --> 00:18:30,380
protein so we can say that for example

484
00:18:34,300 --> 00:18:32,360
this type of ferredoxin domain or this

485
00:18:35,800 --> 00:18:34,310
type of iron-sulfur domain is often

486
00:18:37,690 --> 00:18:35,810
found next to a molybdenum site

487
00:18:40,510 --> 00:18:37,700
this type of heme domain for example is

488
00:18:42,760 --> 00:18:40,520

often found next to a an iron sulfur

489
00:18:44,950 --> 00:18:42,770
site and we can build this map of

490
00:18:50,020 --> 00:18:44,960
spatial connectivity within

491
00:18:53,650 --> 00:18:50,030
oxidoreductases so we did this and this

492
00:18:55,950 --> 00:18:53,660
is what we got so there are a lot of

493
00:18:58,360 --> 00:18:55,960
interesting things about this network

494
00:19:01,780 --> 00:18:58,370
what I'm showing you here each of these

495
00:19:04,230 --> 00:19:01,790
nodes now is not a specific protein site

496
00:19:06,160 --> 00:19:04,240
it's a module so it's the collection of

497
00:19:08,740 --> 00:19:06,170
micro environments that all have

498
00:19:10,870 --> 00:19:08,750
structural similarity the size of the

499
00:19:12,550 --> 00:19:10,880
node represents the number of

500
00:19:14,650 --> 00:19:12,560
connections it makes with other types of

501
00:19:16,960 --> 00:19:14,660
modules so these are connections these

502
00:19:18,940 --> 00:19:16,970
edges are two other modules that are

503
00:19:22,420 --> 00:19:18,950
beyond our threshold for structural

504
00:19:24,910 --> 00:19:22,430
similarity and then the edges themselves

505
00:19:27,070 --> 00:19:24,920
represent a the thickness of the edges

506
00:19:29,740 --> 00:19:27,080
represents the number of instances of a

507
00:19:31,000 --> 00:19:29,750
particular connection that we see and we

508
00:19:33,580 --> 00:19:31,010
can see here that those same four

509
00:19:37,030 --> 00:19:33,590
modules that were highly represented in

510
00:19:39,280 --> 00:19:37,040
the the data set there also are they

511
00:19:42,220 --> 00:19:39,290
make a large number of connections with

512
00:19:44,470 --> 00:19:42,230
other types of modules in the in this

513
00:19:48,640 --> 00:19:44,480

spatial adjacency Network within the

514
00:19:52,690 --> 00:19:48,650
span and so what rules can we get for

515
00:19:55,300 --> 00:19:52,700
the assembly of electron transport

516
00:19:57,400 --> 00:19:55,310
chains from looking at this well one of

517
00:20:01,120 --> 00:19:57,410
the things that we noticed for about 30%

518
00:20:03,580 --> 00:20:01,130
of module module connections we have

519
00:20:05,710 --> 00:20:03,590
instead of connecting one type of module

520
00:20:07,240 --> 00:20:05,720
to another we had these loops and so

521
00:20:10,270 --> 00:20:07,250
what a loop here represents essentially

522
00:20:10,690 --> 00:20:10,280
is in a ferredoxin type module connected

523
00:20:12,850 --> 00:20:10,700
to another

524
00:20:15,250 --> 00:20:12,860
ferredoxin Taekwon jewel or a cytochrome

525
00:20:16,660 --> 00:20:15,260
C connected to another cytochrome C or a

526
00:20:19,150 --> 00:20:16,670
rubra dachshund connected to another

527
00:20:21,840 --> 00:20:19,160
rubra dachshund and so you know again

528
00:20:23,530 --> 00:20:21,850
this is not something that is new to

529
00:20:25,090 --> 00:20:23,540
oxidoreductases this is something that

530
00:20:27,970 --> 00:20:25,100
you classically see in a lot of

531
00:20:29,740 --> 00:20:27,980
different multi-domain proteins is that

532
00:20:31,240 --> 00:20:29,750
the way that you make complexity or you

533
00:20:33,520 --> 00:20:31,250
make larger proteins from smaller

534
00:20:36,070 --> 00:20:33,530
domains is through duplication and

535
00:20:39,250 --> 00:20:36,080
diversification so there are some very

536
00:20:40,600 --> 00:20:39,260
clear examples of this and oxido

537
00:20:43,000 --> 00:20:40,610
reductase you have these seen for

538
00:20:45,340 --> 00:20:43,010

example these multi heme proteins and

539
00:20:47,320 --> 00:20:45,350
geo bacter for example that allow you

540
00:20:50,110 --> 00:20:47,330
that allow electron transfer from

541
00:20:53,260 --> 00:20:50,120
mineral substrates into the into the

542
00:20:55,600 --> 00:20:53,270
cell or here's a an iron sulphur wire

543
00:20:57,250 --> 00:20:55,610
that's made out of multiple ferredoxin

544
00:20:59,820 --> 00:20:57,260
x' that are connected together we see

545
00:21:02,200 --> 00:20:59,830
similar things for plastocyanin x' for

546
00:21:03,580 --> 00:21:02,210
ferritin and so forth but i think what's

547
00:21:05,710 --> 00:21:03,590
really interesting here is that it's not

548
00:21:07,240 --> 00:21:05,720
just sort of these very clear examples

549
00:21:10,140 --> 00:21:07,250
where you have these multi cofactor

550
00:21:12,310 --> 00:21:10,150
chains but nearly every module has

551
00:21:14,530 --> 00:21:12,320
examples of these sort of duplications

552
00:21:16,540 --> 00:21:14,540
so clearly you know an important rule

553
00:21:18,220 --> 00:21:16,550
for how do you build complexity is to

554
00:21:21,210 --> 00:21:18,230
just copy something and connect it to a

555
00:21:23,400 --> 00:21:21,220
domain of the same kind so that that is

556
00:21:25,870 --> 00:21:23,410
that's one rule that came out of this

557
00:21:27,910 --> 00:21:25,880
but one of the other things that we

558
00:21:30,100 --> 00:21:27,920
found very interesting and it jumps out

559
00:21:32,500 --> 00:21:30,110
at you if you color the nodes by the

560
00:21:35,260 --> 00:21:32,510
types of cofactors that they bind is

561
00:21:37,960 --> 00:21:35,270
that all of the COFA all of the

562
00:21:39,630 --> 00:21:37,970
cofactors of the same type are connected

563
00:21:43,810 --> 00:21:39,640

to each other so for example here yellow

564
00:21:46,180 --> 00:21:43,820
represents iron sulfur cofactors and

565
00:21:48,280 --> 00:21:46,190
this is not just for iron poor sulfur

566
00:21:49,810 --> 00:21:48,290
this is to our and - sulfur or something

567
00:21:51,730 --> 00:21:49,820
like rubber dachshund where you have a

568
00:21:53,860 --> 00:21:51,740
single iron and poor cystines

569
00:21:56,500 --> 00:21:53,870
coordinating it so all of these are

570
00:21:57,880 --> 00:21:56,510
connected to each other and a connection

571
00:21:59,650 --> 00:21:57,890
here remember does not mean structural

572
00:22:02,830 --> 00:21:59,660
similarity so we're not saying that a

573
00:22:04,780 --> 00:22:02,840
risky type iron sulfur cluster site

574
00:22:06,820 --> 00:22:04,790
looks a lot like a four iron four sulfur

575
00:22:08,200 --> 00:22:06,830
from bacterial ferredoxin they're

576
00:22:09,730 --> 00:22:08,210
structurally very distinct from each

577
00:22:11,230 --> 00:22:09,740
other but what we're saying is that

578
00:22:11,620 --> 00:22:11,240
they're often found connected to each

579
00:22:13,840 --> 00:22:11,630
other

580
00:22:17,620 --> 00:22:13,850
the same thing is true for these four

581
00:22:19,690 --> 00:22:17,630
helix bundle type single iron sites that

582
00:22:23,170 --> 00:22:19,700
are connected to a lot of other mono

583
00:22:24,490 --> 00:22:23,180
metal binding sites same thing are true

584
00:22:27,850 --> 00:22:24,500
for heme binding sites same thing

585
00:22:29,170 --> 00:22:27,860
true for copper binding sites so this is

586
00:22:31,570 --> 00:22:29,180
actually very interesting why are we

587
00:22:34,930 --> 00:22:31,580
getting this metal segregation within

588
00:22:36,280 --> 00:22:34,940

this this this graph and so there's

589
00:22:39,460 --> 00:22:36,290
there's a couple of explanations for

590
00:22:40,420 --> 00:22:39,470
this so one would be that what we're

591
00:22:42,090 --> 00:22:40,430
seeing here because these are

592
00:22:45,190 --> 00:22:42,100
essentially we're arguing that these are

593
00:22:47,500 --> 00:22:45,200
electron transport pathways so one

594
00:22:51,430 --> 00:22:47,510
argument would be that all iron sulfur

595
00:22:53,290 --> 00:22:51,440
sites have similar redox potentials and

596
00:22:54,970 --> 00:22:53,300
so the fact that you have all of these

597
00:22:56,280 --> 00:22:54,980
these iron software sites connected to

598
00:22:58,420 --> 00:22:56,290
each other is essentially just a

599
00:23:00,100 --> 00:22:58,430
thermodynamic phenomenon that this

600
00:23:04,300 --> 00:23:00,110
allows that you don't have any high high

601
00:23:06,280 --> 00:23:04,310
barriers from transfer from one from one

602
00:23:07,930 --> 00:23:06,290
site to the next but we know from

603
00:23:10,630 --> 00:23:07,940
protein engineering studies that you can

604
00:23:13,450 --> 00:23:10,640
have the same metal site and just make

605
00:23:15,370 --> 00:23:13,460
single amino acid changes around in the

606
00:23:18,310 --> 00:23:15,380
second shell around metal site and you

607
00:23:20,410 --> 00:23:18,320
can move the redox potential by over a

608
00:23:22,480 --> 00:23:20,420
volt so you can for example with iron

609
00:23:25,060 --> 00:23:22,490
sulfur sites or with heme sites have a

610
00:23:27,550 --> 00:23:25,070
huge tuning potential without changing

611
00:23:29,470 --> 00:23:27,560
the metal type so that is it's a

612
00:23:30,760 --> 00:23:29,480
possible explanation but it's not

613
00:23:33,490 --> 00:23:30,770

necessarily a very convincing one

614
00:23:35,710 --> 00:23:33,500
another one to think about is protein

615
00:23:37,540 --> 00:23:35,720
biosynthesis so if you're making a

616
00:23:41,170 --> 00:23:37,550
protein that contains multiple cofactors

617
00:23:43,000 --> 00:23:41,180
it might be easier in terms of assembly

618
00:23:45,030 --> 00:23:43,010
to have all the Medeco factors be the

619
00:23:47,320 --> 00:23:45,040
same and then you can provide multiple

620
00:23:49,860 --> 00:23:47,330
iron sulfur clusters to a single protein

621
00:23:51,730 --> 00:23:49,870
or multiple teams to a single protein

622
00:23:53,980 --> 00:23:51,740
but we know that there are many examples

623
00:23:56,410 --> 00:23:53,990
of oxidoreductases that have multiple

624
00:23:58,570 --> 00:23:56,420
different cofactor types in them so

625
00:23:59,920 --> 00:23:58,580
those are two explanations but another

626
00:24:01,950 --> 00:23:59,930
which we think is particularly

627
00:24:05,380 --> 00:24:01,960
tantalizing one that we're now exploring

628
00:24:07,720 --> 00:24:05,390
experimentally within the lab is perhaps

629
00:24:10,840 --> 00:24:07,730
what this is suggesting is that these

630
00:24:13,450 --> 00:24:10,850
evolutionary some physical connections

631
00:24:15,220 --> 00:24:13,460
between modules represent duplication

632
00:24:17,500 --> 00:24:15,230
and then significant diversification

633
00:24:20,830 --> 00:24:17,510
that you have you know the iron sulfur

634
00:24:22,810 --> 00:24:20,840
site constraining the the first shell

635
00:24:25,570 --> 00:24:22,820
ligands but then beyond that you get

636
00:24:27,640 --> 00:24:25,580
significant diversification of the

637
00:24:29,500 --> 00:24:27,650
second shell and a microenvironment on

638
00:24:32,440 --> 00:24:29,510

the protein so in other words that these

639
00:24:33,820 --> 00:24:32,450
these these connections in space may

640
00:24:37,960 --> 00:24:33,830
actually represent evolutionary

641
00:24:38,350 --> 00:24:37,970
connections and this is you know I would

642
00:24:39,460 --> 00:24:38,360
say that

643
00:24:41,530 --> 00:24:39,470
this is something that we still haven't

644
00:24:42,940 --> 00:24:41,540
proven but it's the way I like to think

645
00:24:44,980 --> 00:24:42,950
about this is you know we have this

646
00:24:46,930 --> 00:24:44,990
expression in English that if you're

647
00:24:48,340 --> 00:24:46,940
comparing apples and oranges you're

648
00:24:49,990 --> 00:24:48,350
talking about two very different things

649
00:24:50,980 --> 00:24:50,000
right they're both fruits but they're

650
00:24:53,500 --> 00:24:50,990
very different fruits from each other

651
00:24:55,570 --> 00:24:53,510
one is citrus the other is not but what

652
00:24:57,910 --> 00:24:55,580
if you were to walk out and you find a

653
00:24:58,660 --> 00:24:57,920
tree that had both apples and oranges on

654
00:25:00,400 --> 00:24:58,670
the same tree

655
00:25:01,900 --> 00:25:00,410
now you'd search say well maybe this

656
00:25:03,820 --> 00:25:01,910
expert that expression doesn't make so

657
00:25:05,800 --> 00:25:03,830
much sense maybe apples and oranges are

658
00:25:07,630 --> 00:25:05,810
a lot similar more similar than we

659
00:25:11,110 --> 00:25:07,640
thought and what we're thinking we might

660
00:25:12,250 --> 00:25:11,120
be seeing here and that span is what we

661
00:25:14,500 --> 00:25:12,260
originally thought to be apples and

662
00:25:16,930 --> 00:25:14,510
oranges occurring on the same tree so if

663
00:25:20,440 --> 00:25:16,940

that's the case then what we have now is

664
00:25:22,120 --> 00:25:20,450
a tool for discriminating analogy

665
00:25:25,980 --> 00:25:22,130
etymology so if we go back for example

666
00:25:28,510 --> 00:25:25,990
to the the heme binding cytochrome Seema

667
00:25:30,940 --> 00:25:28,520
module so this is a module that had

668
00:25:33,010 --> 00:25:30,950
about a thousand different members and

669
00:25:36,940 --> 00:25:33,020
let's take this single module now and we

670
00:25:38,610 --> 00:25:36,950
cluster it using a loo vein clustering

671
00:25:41,500 --> 00:25:38,620
method it's just a one way of sort of

672
00:25:44,890 --> 00:25:41,510
classifying sub sub graphs within a

673
00:25:46,510 --> 00:25:44,900
larger graph let's say we classify this

674
00:25:49,510 --> 00:25:46,520
into like eight smaller segments and

675
00:25:51,550 --> 00:25:49,520
this is these connections here are based

676
00:25:54,160 --> 00:25:51,560
on structural similarity and then we

677
00:25:56,230 --> 00:25:54,170
take that and we now generate a span for

678
00:25:58,570 --> 00:25:56,240
that so now we say within those sub

679
00:26:00,280 --> 00:25:58,580
graphs which ones are found spatially

680
00:26:02,230 --> 00:26:00,290
next to each other within the same same

681
00:26:04,800 --> 00:26:02,240
protein and what we find is actually

682
00:26:07,510 --> 00:26:04,810
within this larger module there are

683
00:26:09,520 --> 00:26:07,520
subclasses so that one in two are often

684
00:26:10,690 --> 00:26:09,530
found connected to each other but we

685
00:26:13,540 --> 00:26:10,700
never see connections between one and

686
00:26:15,490 --> 00:26:13,550
two in any of the other cytochrome C

687
00:26:16,980 --> 00:26:15,500
type modules between three and eight so

688
00:26:19,660 --> 00:26:16,990

maybe there actually are two

689
00:26:22,180 --> 00:26:19,670
evolutionarily different cytochrome C

690
00:26:23,860 --> 00:26:22,190
type modules that all they share really

691
00:26:25,960 --> 00:26:23,870
is just this chemical similarity that

692
00:26:28,780 --> 00:26:25,970
they bind means but they have otherwise

693
00:26:30,160 --> 00:26:28,790
independent evolutionary origins so in

694
00:26:32,710 --> 00:26:30,170
other words what we're looking at here

695
00:26:34,600 --> 00:26:32,720
is convergent evolution of one two class

696
00:26:42,570 --> 00:26:34,610
and the three through a class but then

697
00:26:46,330 --> 00:26:45,040
so then that suggests if we look at the

698
00:26:48,940 --> 00:26:46,340
go back and look at the network with

699
00:26:51,340 --> 00:26:48,950
this mind set that perhaps we have four

700
00:26:52,040 --> 00:26:51,350
or five or six fundamental modules maybe

701
00:26:53,600 --> 00:26:52,050
the ferry dock

702
00:26:55,490 --> 00:26:53,610
and the ferredoxin obvious is great

703
00:26:57,950 --> 00:26:55,500
analyzing when this may have been one of

704
00:27:00,640 --> 00:26:57,960
the first iron-sulfur module then

705
00:27:03,200 --> 00:27:00,650
diversified into a number of other

706
00:27:05,780 --> 00:27:03,210
module types the same thing with maybe

707
00:27:10,400 --> 00:27:05,790
one of these cytochrome c type modules

708
00:27:15,230 --> 00:27:10,410
the four helix bundle as a source and

709
00:27:17,320 --> 00:27:15,240
then the plastocyanin for copper so

710
00:27:21,110 --> 00:27:17,330
we're very excited about exploring now

711
00:27:23,240 --> 00:27:21,120
whether or not we can use these as sort

712
00:27:25,390 --> 00:27:23,250
of archetypes for understanding the

713
00:27:31,760 --> 00:27:25,400

evolution of these original

714
00:27:32,960 --> 00:27:31,770
oxidoreductase modules so what I'd like

715
00:27:35,660 --> 00:27:32,970
to say at this point is that you know we

716
00:27:38,660 --> 00:27:35,670
can take something like a large complex

717
00:27:40,970 --> 00:27:38,670
oxido reductase and decompose it into

718
00:27:42,200 --> 00:27:40,980
smaller modules and we believe that

719
00:27:44,540 --> 00:27:42,210
these modules are behaving like

720
00:27:45,830 --> 00:27:44,550
evolutionarily selectable domains

721
00:27:47,990 --> 00:27:45,840
they're functionally discrete they're

722
00:27:50,360 --> 00:27:48,000
selectable and we believe that this

723
00:27:53,270 --> 00:27:50,370
complexity likely evolved from the

724
00:27:57,290 --> 00:27:53,280
assembly of these smaller modules into

725
00:27:59,120 --> 00:27:57,300
these much larger complexes and very

726
00:28:01,400 --> 00:27:59,130
simply through domain duplication to

727
00:28:03,170 --> 00:28:01,410
build wires but then also maybe through

728
00:28:05,870 --> 00:28:03,180
diversification to develop these more

729
00:28:07,580 --> 00:28:05,880
functionally specialized branches of

730
00:28:10,400 --> 00:28:07,590
these electron transport pathways and

731
00:28:12,680 --> 00:28:10,410
perhaps we can start exploring these

732
00:28:15,230 --> 00:28:12,690
these spatial adjacency relationships as

733
00:28:17,390 --> 00:28:15,240
a construction to look much more deeply

734
00:28:18,950 --> 00:28:17,400
into the fossil history of proteins

735
00:28:21,200 --> 00:28:18,960
where we're not depending on the

736
00:28:23,030 --> 00:28:21,210
vagaries of structural alignments which

737
00:28:26,150 --> 00:28:23,040
are themselves more sensitive to deep

738
00:28:27,560 --> 00:28:26,160

time than sequence alignments for for

739
00:28:29,840 --> 00:28:27,570
making these relationships but also

740
00:28:33,340 --> 00:28:29,850
using this as another way to establish

741
00:28:38,600 --> 00:28:33,350
connections between structural domains

742
00:28:40,550 --> 00:28:38,610
and as with our our work on Homo

743
00:28:43,340 --> 00:28:40,560
chirality and our ability to relate that

744
00:28:45,080 --> 00:28:43,350
to the design of therapeutic peptides we

745
00:28:47,570 --> 00:28:45,090
also see in this case evolution in

746
00:28:49,340 --> 00:28:47,580
design as two sides of the same coin so

747
00:28:50,900 --> 00:28:49,350
while we are studying the evolutionary

748
00:28:52,940 --> 00:28:50,910
relationships between these modules

749
00:28:54,740 --> 00:28:52,950
we're also now thinking about ways of

750
00:28:57,680 --> 00:28:54,750
hooking these things up together to

751
00:29:00,170 --> 00:28:57,690
start to make nanoscale devices for for

752
00:29:01,460 --> 00:29:00,180
by electronics and Eric you may look at

753
00:29:03,110 --> 00:29:01,470
this and you may see a bifurcating

754
00:29:04,970 --> 00:29:03,120
pathway right here we're actually very

755
00:29:05,990 --> 00:29:04,980
interested in going back to the span and

756
00:29:07,100 --> 00:29:06,000
think about

757
00:29:08,480 --> 00:29:07,110
in addition to just pairwise

758
00:29:11,450 --> 00:29:08,490
interactions maybe multi-body

759
00:29:13,670 --> 00:29:11,460
interactions between electron between

760
00:29:17,030 --> 00:29:13,680
these metal sites is to look at the

761
00:29:23,750 --> 00:29:17,040
emergence of more complex topologies and

762
00:29:26,690 --> 00:29:23,760
electron transport and so perhaps

763
00:29:28,910 --> 00:29:26,700

instead of having a single ancestor that

764
00:29:31,070 --> 00:29:28,920
has led to you know the emergence of all

765
00:29:33,260 --> 00:29:31,080
these different oxidoreductases

766
00:29:36,380 --> 00:29:33,270
maybe we had several luca's or maybe you

767
00:29:38,000 --> 00:29:36,390
want to call it Lukas that themselves

768
00:29:41,990 --> 00:29:38,010
assembled in different ways to make

769
00:29:43,070 --> 00:29:42,000
these modern extant nanomachines and so

770
00:29:45,500 --> 00:29:43,080
what we're doing now is we're starting

771
00:29:47,480 --> 00:29:45,510
to ask how can we walk from these very

772
00:29:49,670 --> 00:29:47,490
simple domains which themselves are

773
00:29:51,440 --> 00:29:49,680
fairly functionally naive to perhaps

774
00:29:53,060 --> 00:29:51,450
complexes where you have two or three

775
00:29:54,350 --> 00:29:53,070
domains that can do more interesting

776
00:29:56,900 --> 00:29:54,360
things we want to move in this direction

777
00:29:58,400 --> 00:29:56,910
to words the complex machines and see

778
00:30:00,620 --> 00:29:58,410
what are the minimal assemblies that

779
00:30:03,650 --> 00:30:00,630
give us the the functional the catalytic

780
00:30:04,820 --> 00:30:03,660
properties that we want but also what's

781
00:30:07,460 --> 00:30:04,830
interesting now is we already have

782
00:30:08,780 --> 00:30:07,470
fairly simple archetypes of what these

783
00:30:11,000 --> 00:30:08,790
original modules may have looked like

784
00:30:12,830 --> 00:30:11,010
can we start to walk them back even

785
00:30:14,930 --> 00:30:12,840
further can we go backwards in evolution

786
00:30:16,490 --> 00:30:14,940
paths before the common ancestor and

787
00:30:19,100 --> 00:30:16,500
start thinking about what these

788
00:30:21,470 --> 00:30:19,110

prebiotic peptide and/or peptides may

789
00:30:22,790 --> 00:30:21,480
have looked like and of course this is

790
00:30:24,620 --> 00:30:22,800
something that we heard a little bit

791
00:30:26,690 --> 00:30:24,630
about yesterday thinking about how

792
00:30:27,890 --> 00:30:26,700
peptides may have interacted with

793
00:30:31,400 --> 00:30:27,900
minerals themselves that are already

794
00:30:33,380 --> 00:30:31,410
capable of redox catalysis how baptized

795
00:30:36,020 --> 00:30:33,390
may have aided this and I'll just show

796
00:30:38,590 --> 00:30:36,030
one slide that shows one of our forays

797
00:30:40,820 --> 00:30:38,600
into this area and we've been looking at

798
00:30:43,280 --> 00:30:40,830
bacterial ferredoxin for a long time

799
00:30:45,800 --> 00:30:43,290
it's a protein that binds to for R and

800
00:30:48,890 --> 00:30:45,810
for sulfur clusters it's about 60 amino

801
00:30:52,280 --> 00:30:48,900
acids it itself is clearly a domain

802
00:30:54,500 --> 00:30:52,290
duplication of - 20 to 30 amino acid

803
00:30:56,390 --> 00:30:54,510
domains and what we've done is by

804
00:30:58,370 --> 00:30:56,400
looking at the the reach of the business

805
00:31:00,710 --> 00:30:58,380
end of this molecule that is responsible

806
00:31:03,410 --> 00:31:00,720
for binding the Orang sulfur cluster

807
00:31:05,660 --> 00:31:03,420
we've been able to reduce this 60 amino

808
00:31:07,670 --> 00:31:05,670
acid protein about fivefold

809
00:31:10,190 --> 00:31:07,680
to this small cyclic peptide which is

810
00:31:11,720 --> 00:31:10,200
only 12 amino acids and this 12 amino

811
00:31:14,150 --> 00:31:11,730
acid peptide is able to stay believed

812
00:31:16,370 --> 00:31:14,160
bind a 4-iron for self per cluster this

813
00:31:17,260 --> 00:31:16,380

is an EPR spectrum showing that it has

814
00:31:22,510 --> 00:31:17,270
the

815
00:31:23,800 --> 00:31:22,520
salt for protein but will be find

816
00:31:26,620 --> 00:31:23,810
particularly exciting about this

817
00:31:29,440 --> 00:31:26,630
particular design is that if you notice

818
00:31:31,600 --> 00:31:29,450
here if you look at the the topology of

819
00:31:33,550 --> 00:31:31,610
this protein all of the backbone amides

820
00:31:35,680 --> 00:31:33,560
which are in blue are pointing in

821
00:31:39,370 --> 00:31:35,690
towards the the iron sulfur cluster and

822
00:31:41,260 --> 00:31:39,380
this is something that a structural

823
00:31:43,660 --> 00:31:41,270
biologists Miller and white would call a

824
00:31:45,040 --> 00:31:43,670
cationic nest so essentially all these

825
00:31:47,380 --> 00:31:45,050
back ammonium eyes are creating a nice

826
00:31:49,330 --> 00:31:47,390
stable binding site for an iron sulfur

827
00:31:50,890 --> 00:31:49,340
cluster and so what happens now is that

828
00:31:53,140 --> 00:31:50,900
because you have the stable binding site

829
00:31:55,120 --> 00:31:53,150
we're able to take this complex and

830
00:31:57,250 --> 00:31:55,130
oxidize and reduce it thousands of times

831
00:31:59,530 --> 00:31:57,260
and it doesn't fall apart so this thing

832
00:32:02,380 --> 00:31:59,540
has a redox potential close to that a

833
00:32:05,560 --> 00:32:02,390
ferredoxin but it's very stable and I've

834
00:32:08,500 --> 00:32:05,570
designed several iron sulfur proteins in

835
00:32:09,640 --> 00:32:08,510
my career as a protein designer and the

836
00:32:11,530 --> 00:32:09,650
best that we've been able to do before

837
00:32:13,510 --> 00:32:11,540
this was about 16 cycles before the

838
00:32:14,830 --> 00:32:13,520

thing falls apart and in fact the most

839
00:32:16,810 --> 00:32:14,840
recent design before this fell apart

840
00:32:18,970 --> 00:32:16,820
after one cycle so something that has

841
00:32:20,230 --> 00:32:18,980
this extent of stability is is

842
00:32:23,500 --> 00:32:20,240
unprecedented so we're very excited

843
00:32:25,900 --> 00:32:23,510
about this and so designs like this that

844
00:32:28,000 --> 00:32:25,910
are inspired by these these small

845
00:32:30,400 --> 00:32:28,010
domains may be a way for us to

846
00:32:33,370 --> 00:32:30,410
extrapolate back to what prebiotic

847
00:32:36,310 --> 00:32:33,380
peptides may look like so I'll end there

848
00:32:37,810 --> 00:32:36,320
and thanks again for the invitation and

849
00:32:46,300 --> 00:32:37,820
the chance to speak and I welcome any

850
00:32:47,590 --> 00:32:46,310
questions wonderful talk thank you and

851
00:32:51,160 --> 00:32:47,600
we'll take the first question from

852
00:32:52,080 --> 00:32:51,170
George that was really fabulous I made

853
00:32:54,700 --> 00:32:52,090
it

854
00:32:56,830 --> 00:32:54,710
screaming at classy a couple of quick

855
00:32:58,840 --> 00:32:56,840
questions in these electron transfer

856
00:33:00,670 --> 00:32:58,850
chains are you seeing hopping are you

857
00:33:03,280 --> 00:33:00,680
seeing drift are you seeing tunneling

858
00:33:05,920 --> 00:33:03,290
what is the mechanism so we're agnostic

859
00:33:07,210 --> 00:33:05,930
to the mechanism we don't know if we're

860
00:33:09,130 --> 00:33:07,220
seeing tunneling or if we're seeing

861
00:33:14,200 --> 00:33:09,140
hopping right I mean that you're looking

862
00:33:17,050 --> 00:33:14,210
essentially at connections between metal

863
00:33:22,060 --> 00:33:17,060

clusters that are within the within the

864
00:33:24,220 --> 00:33:22,070
context of the protein matrix so if

865
00:33:27,160 --> 00:33:24,230
you're hairy grey then you would you

866
00:33:28,990 --> 00:33:27,170
would be looking for essentially hopping

867
00:33:30,400 --> 00:33:29,000
intermediates between these these these

868
00:33:32,110 --> 00:33:30,410
metal clusters so aromatics

869
00:33:33,100 --> 00:33:32,120
example and so one of the things that

870
00:33:34,930 --> 00:33:33,110
were interested in looking at now that

871
00:33:36,970 --> 00:33:34,940
we sort of identified what these

872
00:33:39,520 --> 00:33:36,980
electron transport pathways are see

873
00:33:40,840 --> 00:33:39,530
whether we see amino acids between them

874
00:33:43,000 --> 00:33:40,850
that may be affecting the conductivity

875
00:33:45,760 --> 00:33:43,010
the beta of the environment that would

876
00:33:47,560 --> 00:33:45,770
help us delineate the mechanism okay and

877
00:33:49,930 --> 00:33:47,570
then one other questions and many I

878
00:33:51,490 --> 00:33:49,940
could ask if you look at the genomic

879
00:33:54,070 --> 00:33:51,500
structure of these systems that have

880
00:33:56,590 --> 00:33:54,080
these hypothesized multiple domain

881
00:33:59,470 --> 00:33:56,600
replications are they contiguous are

882
00:34:01,690 --> 00:33:59,480
they intron exon mixtures how do you

883
00:34:03,400 --> 00:34:01,700
actually get these piled together in the

884
00:34:06,490 --> 00:34:03,410
genome in such a fashion that you end up

885
00:34:08,650 --> 00:34:06,500
with the collection of contiguous amino

886
00:34:10,060 --> 00:34:08,660
acids as you see so we simply haven't

887
00:34:11,620 --> 00:34:10,070
done that but I think that's that's an

888
00:34:13,750 --> 00:34:11,630

important next step is to start thinking

889
00:34:16,419 --> 00:34:13,760
because sequence evolution doesn't

890
00:34:18,970 --> 00:34:16,429
happen in structure you don't blong

891
00:34:21,400 --> 00:34:18,980
units together it has to happen at the

892
00:34:22,930 --> 00:34:21,410
level of sequence and previously we

893
00:34:25,090 --> 00:34:22,940
tried to look at this problem using

894
00:34:27,280 --> 00:34:25,100
sequence analysis alone trying to

895
00:34:28,810 --> 00:34:27,290
extrapolate from one type of metal

896
00:34:32,020 --> 00:34:28,820
binding site to another through sequence

897
00:34:33,940 --> 00:34:32,030
intermediates and I think that combining

898
00:34:35,110 --> 00:34:33,950
that analysis with the structural

899
00:34:35,500 --> 00:34:35,120
analysis would be a way to get your

900
00:34:41,800 --> 00:34:35,510
question

901
00:34:43,480 --> 00:34:41,810
definitely hey so great talk for the you

902
00:34:45,330 --> 00:34:43,490
know acid sequences that are involved in

903
00:34:48,190 --> 00:34:45,340
these highly stable kind of small

904
00:34:49,750 --> 00:34:48,200
systems that stabilize these yes yeah

905
00:34:52,000 --> 00:34:49,760
yeah is there any relationship between

906
00:34:53,650 --> 00:34:52,010
the amino acids present in those and the

907
00:34:55,810 --> 00:34:53,660
biosynthetic pathways by which those

908
00:34:58,330 --> 00:34:55,820
amino acids are generated meaning are

909
00:34:59,950 --> 00:34:58,340
they kind of initially early amino acids

910
00:35:06,640 --> 00:34:59,960
or these something's that came along

911
00:35:08,770 --> 00:35:06,650
quite a bit later so I don't know so I

912
00:35:10,840 --> 00:35:08,780
don't I don't know if there is if we can

913
00:35:13,480 --> 00:35:10,850

only make these things with simple amino

914
00:35:15,640 --> 00:35:13,490
acids what I'll say is that the sequence

915
00:35:17,340 --> 00:35:15,650
pattern that we're using here is very

916
00:35:21,130 --> 00:35:17,350
similar to the one that was proposed by

917
00:35:23,530 --> 00:35:21,140
day - and Vanek which is that that small

918
00:35:25,030 --> 00:35:23,540
four amino acid repeat and critical to

919
00:35:27,580 --> 00:35:25,040
that is having a cysteine obviously

920
00:35:30,520 --> 00:35:27,590
binding the the cluster within the

921
00:35:37,170 --> 00:35:30,530
flanking amino acids amino acids like

922
00:35:40,500 --> 00:35:37,180
lysine and glycine work very well yeah

923
00:35:42,410 --> 00:35:40,510
whoever has the Chumash the magic cube

924
00:35:45,289 --> 00:35:42,420
yeah

925
00:35:49,370 --> 00:35:45,299
hopefully quick stunning like everybody

926
00:35:53,329 --> 00:35:49,380
else says it's a question about the way

927
00:35:56,030 --> 00:35:53,339
you use this this paradigm for

928
00:35:58,190 --> 00:35:56,040
evolutionary interpretation it looks

929
00:36:00,020 --> 00:35:58,200
like this natural modularization gives

930
00:36:03,849 --> 00:36:00,030
you a kind of typology of functional

931
00:36:06,500 --> 00:36:03,859
States and if I think about an

932
00:36:09,500 --> 00:36:06,510
evolutionary model I want a model of

933
00:36:11,510 --> 00:36:09,510
states and transitions if I think about

934
00:36:14,270 --> 00:36:11,520
what people often do in trying to

935
00:36:16,490 --> 00:36:14,280
recover old protein folds or old protein

936
00:36:17,990 --> 00:36:16,500
fold fragments they look at the

937
00:36:20,660 --> 00:36:18,000
recruitment of a thing that's

938
00:36:23,990 --> 00:36:20,670

effectively a unit that can move that

939
00:36:27,740 --> 00:36:24,000
can mutate that can do whatever is it

940
00:36:30,829 --> 00:36:27,750
possible to think about looking at the

941
00:36:34,309 --> 00:36:30,839
the repurposing of existing structures

942
00:36:37,430 --> 00:36:34,319
with minimal changes in the way fold

943
00:36:39,710 --> 00:36:37,440
reconstruction people often do and then

944
00:36:42,349 --> 00:36:39,720
looking at these as sort of attractors

945
00:36:45,190 --> 00:36:42,359
to the viable states that tell you where

946
00:36:48,170 --> 00:36:45,200
a spandrel can be anchored to make a

947
00:36:50,329 --> 00:36:48,180
kind of comprehensive evolutionary

948
00:36:52,549 --> 00:36:50,339
reconstruction that is both what you do

949
00:36:58,660 --> 00:36:52,559
and also makes contact with the strong

950
00:37:03,079 --> 00:37:01,370
we need to do some groundwork here right

951
00:37:06,020 --> 00:37:03,089
so what we're looking at right now these

952
00:37:08,120 --> 00:37:06,030
are structural modules whether they are

953
00:37:11,809 --> 00:37:08,130
functional modules or not are not

954
00:37:13,940 --> 00:37:11,819
necessarily discrete pieces of sequence

955
00:37:16,000 --> 00:37:13,950
and we need to get to that state in

956
00:37:18,109 --> 00:37:16,010
order to do things like ancestral

957
00:37:21,620 --> 00:37:18,119
reconstruction methods to see if we can

958
00:37:23,120 --> 00:37:21,630
figure out what our what are the what

959
00:37:26,150 --> 00:37:23,130
are the intermediates between two types

960
00:37:32,059 --> 00:37:26,160
of two types of modules on the span

961
00:37:35,059 --> 00:37:32,069
would be and I'm I'm very much inspired

962
00:37:36,230 --> 00:37:35,069
by the work of Brian and/or bond at the

963
00:37:37,970 --> 00:37:36,240

University of Maryland I don't know if

964
00:37:41,660 --> 00:37:37,980
you've seen some of this work where they

965
00:37:45,200 --> 00:37:41,670
essentially go from three helix protein

966
00:37:46,849 --> 00:37:45,210
that binds to BSA to a one helix three

967
00:37:48,890 --> 00:37:46,859
beta sheet protein that binds to an

968
00:37:50,390 --> 00:37:48,900
immunoglobulin and what they're able to

969
00:37:52,849 --> 00:37:50,400
do is through a series of single amino

970
00:37:55,220 --> 00:37:52,859
acid mutations keeping the binding sites

971
00:37:55,880 --> 00:37:55,230
for both of those domains intact walky

972
00:37:58,400 --> 00:37:55,890
from a protein

973
00:37:59,750 --> 00:37:58,410
has one structure to important that is

974
00:38:01,609 --> 00:37:59,760
another structure and at the very center

975
00:38:03,799 --> 00:38:01,619
of that with a single amino acid

976
00:38:05,569 --> 00:38:03,809
mutation you can go to one structure or

977
00:38:07,250 --> 00:38:05,579
you can go to the other right and I

978
00:38:10,250 --> 00:38:07,260
think that what we're is what we're

979
00:38:12,589 --> 00:38:10,260
saying here with spatial connection

980
00:38:15,140 --> 00:38:12,599
being an evolutionary connection that is

981
00:38:17,000 --> 00:38:15,150
I think at this point still a hypothesis

982
00:38:18,500 --> 00:38:17,010
and the only way for us to really see

983
00:38:20,509 --> 00:38:18,510
whether that's plausible is to go into

984
00:38:21,650 --> 00:38:20,519
the laboratory and try and design some

985
00:38:23,660 --> 00:38:21,660
of these pathways and see whether

986
00:38:25,849 --> 00:38:23,670
they're plausible so the the way that I

987
00:38:27,140 --> 00:38:25,859
approach that as a protein engineer will

988
00:38:29,329 --> 00:38:27,150

be to try and actually engineer some of

989
00:38:31,099 --> 00:38:29,339
these transition fossils and see whether

990
00:38:37,579 --> 00:38:31,109
we can make them behave the way that we

991
00:38:40,339 --> 00:38:37,589
would expect them to oh it was a really

992
00:38:42,859 --> 00:38:40,349
intriguing talk thank you so much so in

993
00:38:46,039 --> 00:38:42,869
terms of like the transition from this

994
00:38:48,769 --> 00:38:46,049
prebiotic to the biotic function one of

995
00:38:51,380 --> 00:38:48,779
the the key question is just is the

996
00:38:53,870 --> 00:38:51,390
actual maintenance and the evolution of

997
00:38:57,079 --> 00:38:53,880
the functionality of this polypeptide

998
00:39:00,109 --> 00:38:57,089
that Co associated with the metal and I

999
00:39:03,470 --> 00:39:00,119
was wondering so this type of mall this

1000
00:39:06,499 --> 00:39:03,480
type of minimal module could form in

1001
00:39:08,990 --> 00:39:06,509
prebiotic era however we all know that

1002
00:39:10,940 --> 00:39:09,000
the protein can be only being replicated

1003
00:39:12,740 --> 00:39:10,950
through this genetic coding and that's

1004
00:39:15,740 --> 00:39:12,750
always been a problematic but I was

1005
00:39:19,700 --> 00:39:15,750
wondering it seems like this minimal

1006
00:39:21,740 --> 00:39:19,710
module have almost minimal sequence

1007
00:39:23,390 --> 00:39:21,750
specificity meaning that doesn't

1008
00:39:27,289 --> 00:39:23,400
necessarily need to be this specific

1009
00:39:30,289 --> 00:39:27,299
sequence in an inner primary mode so do

1010
00:39:33,410 --> 00:39:30,299
you have you ever looked into like this

1011
00:39:37,039 --> 00:39:33,420
the phase space this the functional

1012
00:39:38,690 --> 00:39:37,049
landscape of this type of module and if

1013
00:39:42,799 --> 00:39:38,700

that landscape is big enough to

1014
00:39:45,049 --> 00:39:42,809
basically cover a wide range of

1015
00:39:47,180 --> 00:39:45,059
different combination of amino acids

1016
00:39:50,180 --> 00:39:47,190
that can actually do this redox cycle

1017
00:39:53,599 --> 00:39:50,190
then do you think that will leverage the

1018
00:39:57,470 --> 00:39:53,609
the era catastrophe that was thought to

1019
00:39:59,870 --> 00:39:57,480
be necessary for this genetic code so so

1020
00:40:01,579 --> 00:39:59,880
what you're asking if I understand and

1021
00:40:03,859 --> 00:40:01,589
clarify me is that did we come across

1022
00:40:06,859 --> 00:40:03,869
the 112 amino acid sequence that works

1023
00:40:10,099 --> 00:40:06,869
or is this it just just sort of one a

1024
00:40:12,569 --> 00:40:10,109
very evolvable sequence and the

1025
00:40:14,910 --> 00:40:12,579
the answer is that we've only tried two

1026
00:40:16,410 --> 00:40:14,920
or three sequences right and we have one

1027
00:40:20,640 --> 00:40:16,420
that doesn't work and we have two that

1028
00:40:22,710 --> 00:40:20,650
do but those are they are not a very

1029
00:40:23,760 --> 00:40:22,720
good if we were at to answer the

1030
00:40:26,130 --> 00:40:23,770
question that you're asking

1031
00:40:27,960 --> 00:40:26,140
we wouldn't design the sequences we

1032
00:40:30,450 --> 00:40:27,970
would build libraries you know and see

1033
00:40:31,500 --> 00:40:30,460
what is the success rate for that and I

1034
00:40:34,109 --> 00:40:31,510
think that's a very good thing to try

1035
00:40:36,569 --> 00:40:34,119
definitely especially like given the is

1036
00:40:38,540 --> 00:40:36,579
it's the backbone it seems to be the key

1037
00:40:41,490 --> 00:40:38,550
which doesn't require the sidechain

1038
00:40:43,530 --> 00:40:41,500

might this could it kind of imply that

1039
00:40:45,180 --> 00:40:43,540
that's right I mean the all of the

1040
00:40:46,410 --> 00:40:45,190
interactions here are either system.the

1041
00:40:48,120 --> 00:40:46,420
cysteines have to be there the four

1042
00:40:51,329 --> 00:40:48,130
systems and then everything else is

1043
00:40:52,800 --> 00:40:51,339
backbone so in theory there should be

1044
00:40:59,690 --> 00:40:52,810
highly design and will highly evolved

1045
00:41:04,920 --> 00:41:01,740
can't talk to you until you get the cube

1046
00:41:08,280 --> 00:41:04,930
Thanks yeah thanks for the great talk

1047
00:41:10,589 --> 00:41:08,290
I have two questions ones may be easy

1048
00:41:14,190 --> 00:41:10,599
and the other one's a little bit harder

1049
00:41:15,960 --> 00:41:14,200
I think in 2008 Leslie not in this group

1050
00:41:17,370 --> 00:41:15,970
another piece of work out of that group

1051
00:41:18,150 --> 00:41:17,380
they created these things that they were

1052
00:41:20,910 --> 00:41:18,160
calling them maquettes

1053
00:41:22,980 --> 00:41:20,920
and I think they had a 16 amino acids

1054
00:41:24,930 --> 00:41:22,990
and I was just wondering what's the

1055
00:41:25,829 --> 00:41:24,940
commonality or difference between what

1056
00:41:30,950 --> 00:41:25,839
you're showing and what they were

1057
00:41:34,319 --> 00:41:30,960
showing so so that's that sequence is a

1058
00:41:35,670 --> 00:41:34,329
very similar to this if you look closely

1059
00:41:37,980 --> 00:41:35,680
at this you can essentially see that

1060
00:41:40,260 --> 00:41:37,990
there's a cysteine two amino acids in

1061
00:41:42,180 --> 00:41:40,270
another cysteine the difference there is

1062
00:41:44,250 --> 00:41:42,190
that they have three amino acids between

1063
00:41:46,680 --> 00:41:44,260

and their glycine so they're very

1064
00:41:48,720 --> 00:41:46,690
flexible right so there's no there's not

1065
00:41:51,960 --> 00:41:48,730
necessarily a specific conformation for

1066
00:41:53,370 --> 00:41:51,970
those and what I would argue the reason

1067
00:41:55,670 --> 00:41:53,380
why I think this peptide is so well

1068
00:41:58,020 --> 00:41:55,680
behaved in terms of both yield

1069
00:42:01,079 --> 00:41:58,030
specificity for iron sulfur binding and

1070
00:42:03,690 --> 00:42:01,089
then also redox stability is that it

1071
00:42:05,490 --> 00:42:03,700
adopts we've essentially addressed the

1072
00:42:07,530 --> 00:42:05,500
Leventhal paradox for this it has a

1073
00:42:10,290 --> 00:42:07,540
unique conformation in the a post age

1074
00:42:11,940 --> 00:42:10,300
that we believe is in tactic is pre

1075
00:42:16,260 --> 00:42:11,950
organized to bind the iron sulfur

1076
00:42:18,059 --> 00:42:16,270
cluster and so in that sense it's better

1077
00:42:19,230 --> 00:42:18,069
behaved but in a lot of ways it's

1078
00:42:21,660 --> 00:42:19,240
similar right essentially what they did

1079
00:42:23,430 --> 00:42:21,670
also was to go into ferredoxin see what

1080
00:42:24,809 --> 00:42:23,440
was the spacing between cysteines

1081
00:42:27,870 --> 00:42:24,819
and then just make a model peptide on

1082
00:42:29,280 --> 00:42:27,880
maquette that had that same spacing and

1083
00:42:31,230 --> 00:42:29,290
essentially all we've done here is

1084
00:42:32,970 --> 00:42:31,240
rather than look at the sequence we've

1085
00:42:34,079 --> 00:42:32,980
looked at the structure tried to figure

1086
00:42:36,300 --> 00:42:34,089
out what are the key elements of the

1087
00:42:38,400 --> 00:42:36,310
structure that give you that that iron

1088
00:42:40,530 --> 00:42:38,410

sulfur binding can make make a maquette

1089
00:42:41,970 --> 00:42:40,540
that that mimics that okay and then the

1090
00:42:46,440 --> 00:42:41,980
second question that I thought that was

1091
00:42:49,800 --> 00:42:46,450
the easy question yeah so maybe a little

1092
00:42:52,380 --> 00:42:49,810
bit speculative but for these four iron

1093
00:42:54,270 --> 00:42:52,390
first of all four cubes the possibility

1094
00:42:56,309 --> 00:42:54,280
of site differentiation of the cluster

1095
00:42:58,020 --> 00:42:56,319
is really important for like a

1096
00:43:00,290 --> 00:42:58,030
connotates activity like Joseph talked

1097
00:43:03,660 --> 00:43:00,300
about but also the entire radical Sam

1098
00:43:06,329 --> 00:43:03,670
protein family has a similar mode of

1099
00:43:08,640 --> 00:43:06,339
binding for in that case this is the s

1100
00:43:10,500 --> 00:43:08,650
adenosylmethionine so the site

1101
00:43:13,050 --> 00:43:10,510
differentiation having this open iron

1102
00:43:14,700 --> 00:43:13,060
coordination seems important do you

1103
00:43:17,190 --> 00:43:14,710
think it will be possible to make an

1104
00:43:20,430 --> 00:43:17,200
open coordination I are definitely gonna

1105
00:43:22,530 --> 00:43:20,440
try right I mean we're definitely move

1106
00:43:24,930 --> 00:43:22,540
one of the ligands or e replace them

1107
00:43:28,530 --> 00:43:24,940
with the midazolam combinations of

1108
00:43:30,740 --> 00:43:28,540
things so if you know if for example

1109
00:43:33,720 --> 00:43:30,750
having a three iron four sulfur site is

1110
00:43:36,089 --> 00:43:33,730
advantageous or whether we can stabilize

1111
00:43:38,220 --> 00:43:36,099
that that fourth metal with hydroxyl

1112
00:43:40,559 --> 00:43:38,230
that condemn you replace with an active

1113
00:43:51,290 --> 00:43:40,569

site ligand and that would be very

1114
00:43:53,400 --> 00:43:51,300
exciting I think so - yes with a P loop

1115
00:44:03,510 --> 00:43:53,410
well I'm sorry I don't understand what

1116
00:44:05,220 --> 00:44:03,520
wouldn't matter with me yes that's right

1117
00:44:06,390 --> 00:44:05,230
so another have had the glycines in

1118
00:44:08,010 --> 00:44:06,400
there said you could access that

1119
00:44:10,620 --> 00:44:08,020
left-handed conformation right rather

1120
00:44:12,540 --> 00:44:10,630
than using non natural amino acids yes

1121
00:44:13,950 --> 00:44:12,550
so you could say that's right from the

1122
00:44:15,660 --> 00:44:13,960
very beginning you've probably got a P

1123
00:44:17,880 --> 00:44:15,670
loop right at the beginning very early

1124
00:44:20,430 --> 00:44:17,890
in life and all you've got is P loops

1125
00:44:21,870 --> 00:44:20,440
everywhere pubes yeah so and this the

1126
00:44:23,160 --> 00:44:21,880
shares a lot of similarity to P loops I

1127
00:44:25,140 --> 00:44:23,170
mean essentially in a P loop all you

1128
00:44:26,900 --> 00:44:25,150
have is a row of a my it's pointing at

1129
00:44:28,650 --> 00:44:26,910
your your nucleotide so

1130
00:44:30,660 --> 00:44:28,660
electrostatically this behaves a lot

1131
00:44:32,760 --> 00:44:30,670
like the P loop I think what's unique

1132
00:44:35,220 --> 00:44:32,770
about this typology that also presents

1133
00:44:36,020 --> 00:44:35,230
the primary shell ligands in a sidechain

1134
00:44:37,820 --> 00:44:36,030
confirmations

1135
00:44:39,590 --> 00:44:37,830
they can hold a metal core element yes

1136
00:44:41,690 --> 00:44:39,600
so just one question about that then

1137
00:44:44,930 --> 00:44:41,700
well by the way are there glycines in

1138
00:44:46,430 --> 00:44:44,940

this in the 12 amino acid ones no

1139
00:44:48,680 --> 00:44:46,440
there's no glycines in business so it's

1140
00:44:50,090 --> 00:44:48,690
also at alum D amino acids and they're

1141
00:44:52,430 --> 00:44:50,100
already amino acids there Ellen do you

1142
00:44:55,030 --> 00:44:52,440
know ask could you imagine a loop like

1143
00:44:58,160 --> 00:44:55,040
that without cysteine

1144
00:44:59,450 --> 00:44:58,170
could you imagine oh absolutely yeah you

1145
00:45:00,710 --> 00:44:59,460
could but then it and all the

1146
00:45:01,610 --> 00:45:00,720
interactions would be electrostatic and

1147
00:45:04,310 --> 00:45:01,620
so maybe then that could bind a

1148
00:45:07,070 --> 00:45:04,320
phosphate much like much like these

1149
00:45:09,110 --> 00:45:07,080
cationic nests do or up or other anions

1150
00:45:14,930 --> 00:45:09,120
I asked because it looks like cystines

1151
00:45:19,850 --> 00:45:14,940
quite hard to make early on in life so

1152
00:45:22,460 --> 00:45:19,860
Mike what what were the file-based amino

1153
00:45:28,520 --> 00:45:22,470
acids before cysteine there must have

1154
00:45:29,720 --> 00:45:28,530
been a file so what were they that we're

1155
00:45:31,970 --> 00:45:29,730
I mean we can make something

1156
00:45:34,120 --> 00:45:31,980
structurally anything we want we don't

1157
00:45:36,410 --> 00:45:34,130
have to use the natural alphabet here

1158
00:45:45,200 --> 00:45:36,420
what would you what would you suggest

1159
00:45:47,030 --> 00:45:45,210
there's a possibility okay I mean we

1160
00:45:49,730 --> 00:45:47,040
just need to make we need to make some

1161
00:45:51,890 --> 00:45:49,740
bonds between amino acids to make a ring

1162
00:45:55,040 --> 00:45:51,900
or a linear structure or maquette if you

1163
00:45:57,560 --> 00:45:55,050

will but but the point is we understand

1164
00:46:00,020 --> 00:45:57,570
of course that imidazoles and and and

1165
00:46:03,980 --> 00:46:00,030
dials and the cystines were are not

1166
00:46:05,450 --> 00:46:03,990
found in contradict meteorites so what

1167
00:46:11,750 --> 00:46:05,460
would you but there were hydrogen

1168
00:46:13,490 --> 00:46:11,760
sulfide so so what do you give us what

1169
00:46:22,970 --> 00:46:13,500
do you get to play with

1170
00:46:27,010 --> 00:46:22,980
just-just-just glycine my favorites are

1171
00:46:30,140 --> 00:46:27,020
alanine especially and asparagine and

1172
00:46:31,730 --> 00:46:30,150
aubergine okay and probably aspartate

1173
00:46:34,100 --> 00:46:31,740
okay and I think I'm kind of stuck with

1174
00:46:36,200 --> 00:46:34,110
those three okay so so we like us part

1175
00:46:37,760 --> 00:46:36,210
eight maybe not for iron sulfur but for

1176
00:46:41,150 --> 00:46:37,770
binding other types of metal clusters

1177
00:46:45,710 --> 00:46:41,160
definitely like manganese oxides for

1178
00:46:48,950 --> 00:46:45,720
example anyone have a last question for

1179
00:46:50,120 --> 00:46:48,960
the speaker in that case

1180
00:46:54,620 --> 00:46:50,130
thank you

1181
00:47:14,130 --> 00:46:54,630
[Applause]