Vladimir N Uversky (Volodya)

My first contact with Volodya was in1999 I was asked to review a paper submitted to PNAS by Volodya; my own work strongly supported the new findings in Volodya's paper, so I wrote a very strong favorable review. However. another reviewer rejected this paper. Eventually this work was published: Gillespie et al., Proteins: Structure, Function, Bioinformatics 41: 415-427 (2000). This paper now has ~ 2,300 Google Scholar citations. At that time, my group and collaborators were improving our predictors of protein disorder; Voldya's paper gave us a huge collection of manually curated IDPs.

My first in person meeting with Voldya occurred in San Francisco in 2002; I was in San Francisco to attend the Biophysical Society Meeting and invited Volodya to join me for lunch as he lived not too far away. I was so impressed by the number of great ideas that just pored out.

Vladimir N Uversky (Volodya)

One year later, in 2003, I joined the Department of Biochemistry and Molecular Biology at Indiana University School of Medicine. My primary goal was to develop the Center for Computational Biology and Bioinformatics. My first thought was to invite Volodya to join us in Indiana; eventually, in 2004, Volodya came to Indiana. Since then, our collaboration has resulted in 130 co-authored publications. He left Indiana in 2010 to take a position at the University of South Florida, in Tampa, Florida, but we have continued to publish together.

While in Indiana, Volodya was instrumental in the development of the IDP Subgroup of the Biophysical Society which was launched in 2007 and also in the development of the IDP Gordon Research Conference, which held its first meeting in 2010. These two meetings have played major roles in the development of the IDP scientific community.

Vladimir N Uversky (Volodya)

I distinctly remember a telephone call I received from University of North Carolina Chemistry Professor Gary Pielak in 2006. Gary asked me whether Volodya had been sick in the past year? I said, no, but why do you ask? He said that his group's journal club tracks Volodya's publications and, because he did not have a new paper for an entire month, they thought that he might have been ill.

Indeed, Volodya works harder, more efficiently, and with more good ideas than anyone I know. His Google Scholar profile shows that he has more than 640 publications with at least 10 citations, an h-index over 140, and will soon surpass 94,000 citations.

The work I will describe in my seminar would not have occured without Volodya's substantial help. Volodya. I can't thank you enough.

Evolution of IDPs, IDRs, and IDDs, and the Role of Disorder in the Evolution of the Genetic Code

A. Keith Dunker
Department of Biochemistry and Molecular Biology
Indiana University School of Medicine
(kedunker@iupui.edu)

IDP Webinar Honoring Vladimir Uversky

April 13, 2022

IDPs: FIVE EVOLUTIONARY BEHAVIORS

- Conserved-sequence IDPs
- Variable-sequence IDPs
- Nonconserved IDPs
- INDELS = IDPs
- Fusions

Brown CJ et al., J Mol Evol 55: 100-110 (2002) ● ●

Chen JW et al., J Proteome Res 5: 579-587 (2006) ●

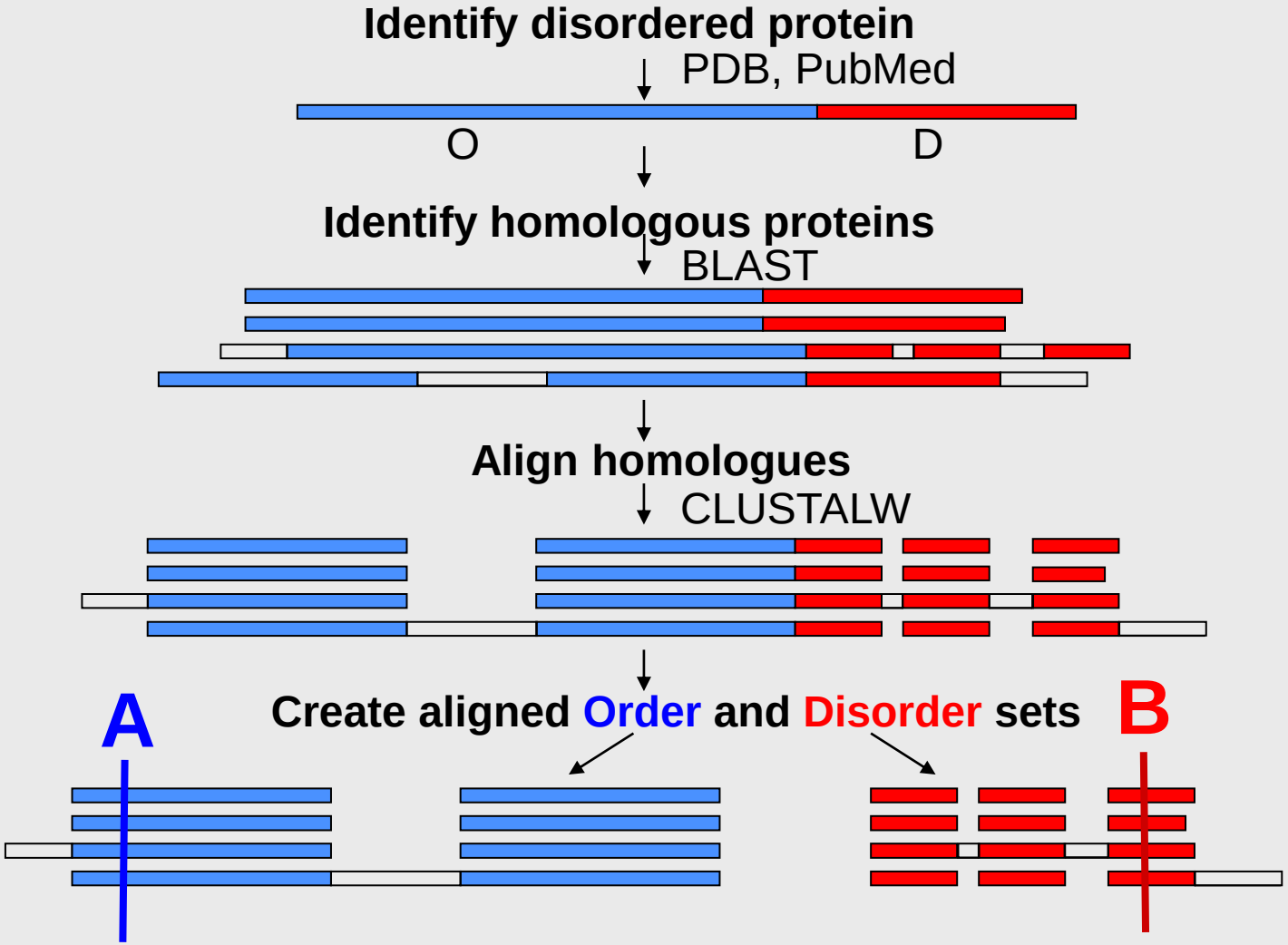
Bellay J et al., Genome Biology 12: R14 (2011) ● ● ●

Fukuchi S et al., JMB 355: 845-857 (2006) ●

Brown CJ et al., COSB 21: 441-446 (2011) ●

Hegyi H et al., PLoS comp biol 5: e1000552 (2009) ●

Molecular Evolution Analysis of Order and Disorder



For all A, B in O, D regions:

1. Mutation rates?

2. Substitution types?

Brown C, et al., *J Mol Evol* 55: 104-110 (2002)

#2 Evolution of Intrinsically Disordered Proteins

Number	Rates	Comments
14 Families	D >> O	As expected
4 Families	D > O	As expected
3 Families	D ~ O	Unexpected
2 Families	O > D	A total surprise!

Brown CJ et al., *J. Mol. Evol.* 55: 104-110 (2002)

Molecular Evolution Analysis of Order and Disorder

- **Mutation rates: often higher in disordered regions as compared to structured protein regions**

  Brown C, et al., *J. Mol. Evol.* 55: 104-110 (2002)

- **Evolutionary changes: mostly conservative for structured proteins, often not conservative for disordered proteins**

  I ←→ L; E←→ D; R ←→ K; common in O, also in D

  G ←→ W; G ←→ E; G ←→ R: rare in O, common in D!

  Radivojac P, et al., *Pac Symp Biocomput* 2002: 589-600

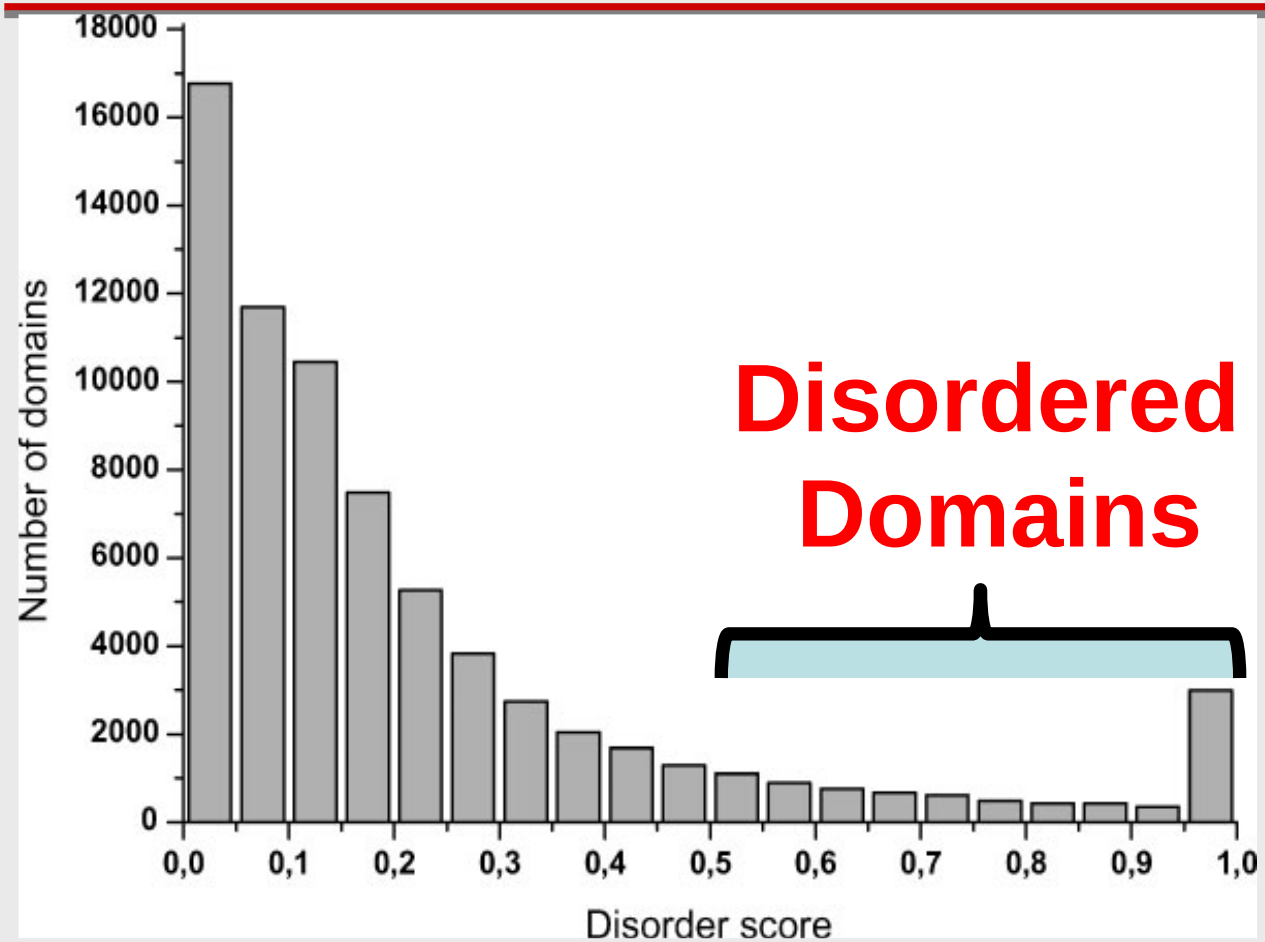
  Midic U, et al., 15th ACM SIG/KDD Int'l Conf. (2009)

Supports existence of disordered protein inside cells!

Structured vs Disordered Domains

Structured Domains have Conserved:	Disordered Domains have Conserved:
●Sequence,	●Sequence,
●Structure,	●Disorder,
●Function.	●Function.

Source of Disordered Domains: Pfam Database



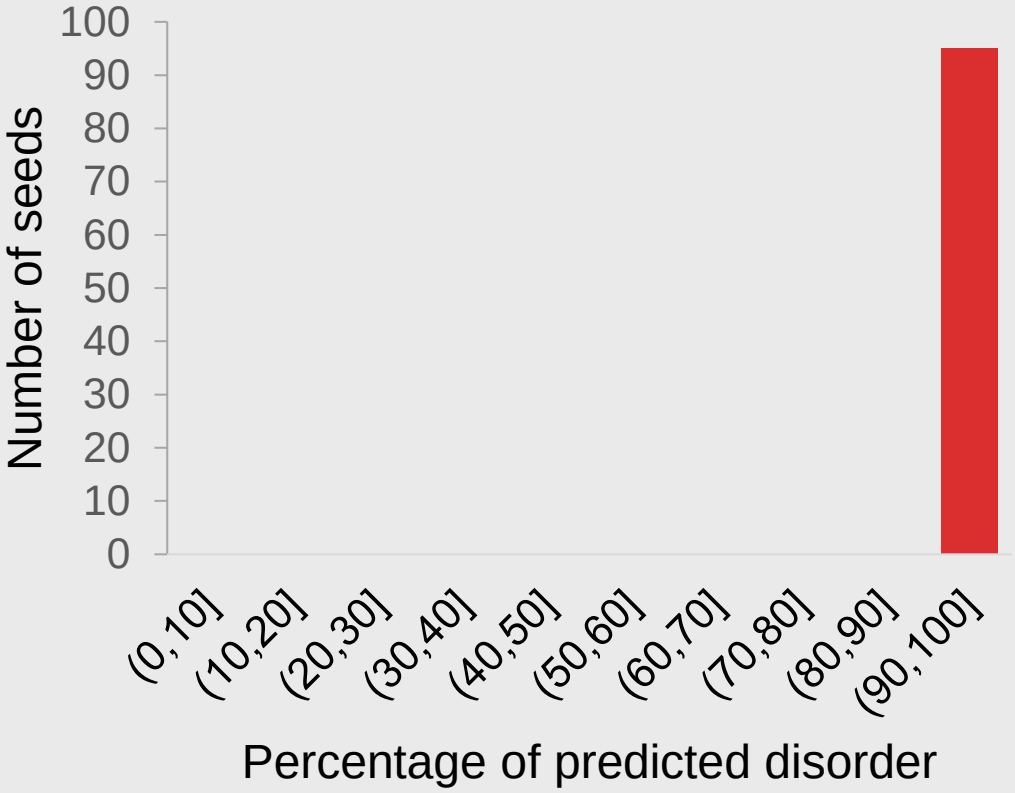
Tompa P et al., *Bioessays* 31: 328-335 (2009)

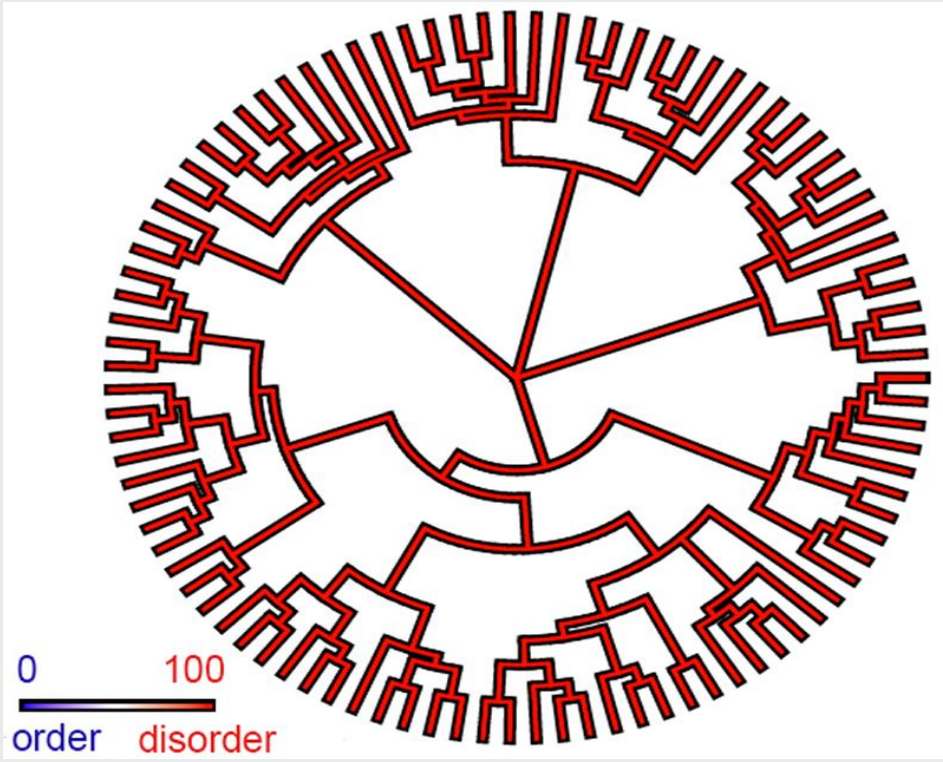
Williams RW et al., *IDPs* 1:e25724 (2013)

Results: 618 predicted to be 100% disordered

Zhou J et al., Prot Sci 28: 1652-1663 (2019)

Keratin, high sulfur B2 protein
Keratin_B2_2(PF13885)





Zhou J et al., Prot Sci 28: 1652-1663 (2019)

IDPs: FIVE EVOLUTIONARY BEHAVIORS

- Conserved-sequence IDPs
- →→ ***Variable-sequence IDPs*** ←←
- Nonconserved IDPs
- INDELS = IDPs
- Fusions

Brown CJ et al., J Mol Evol 55: 100-110 (2002) ● ●

Chen JW et al., J Proteome Res 5: 579-587 (2006) ●

Bellay J et al., Genome Biology 12: R14 (2011) ● ● ●

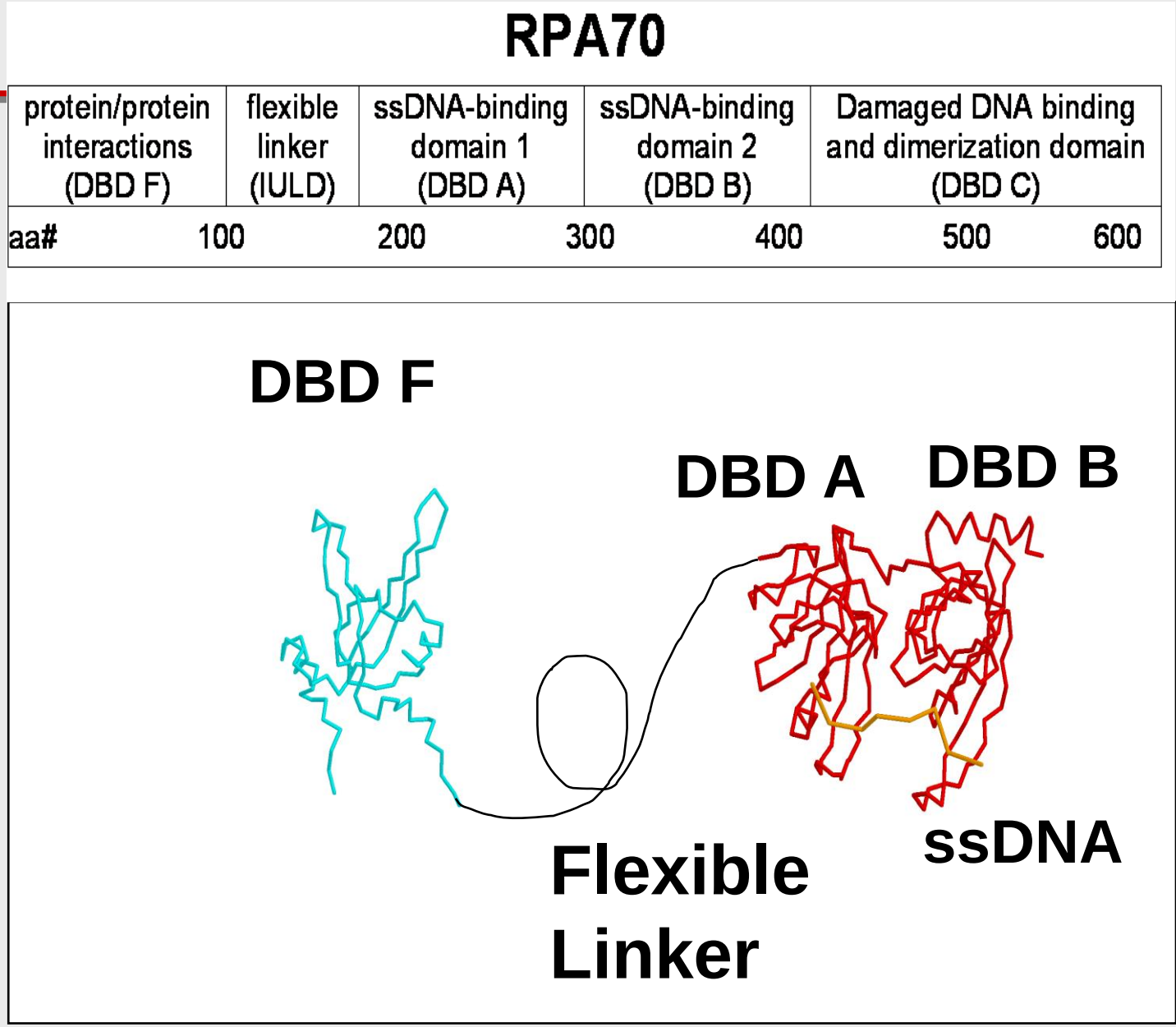
Fukuchi S et al., JMB 355: 845-857 (2006) ●

Brown CJ et al., COSB 21: 441-446 (2011) ●

Hegyi H et al., PLoS comp biol 5: e1000552 (2009) ●

Replication Protein A 70 (RPA70)

RPA70 contains a long disordered linker domain



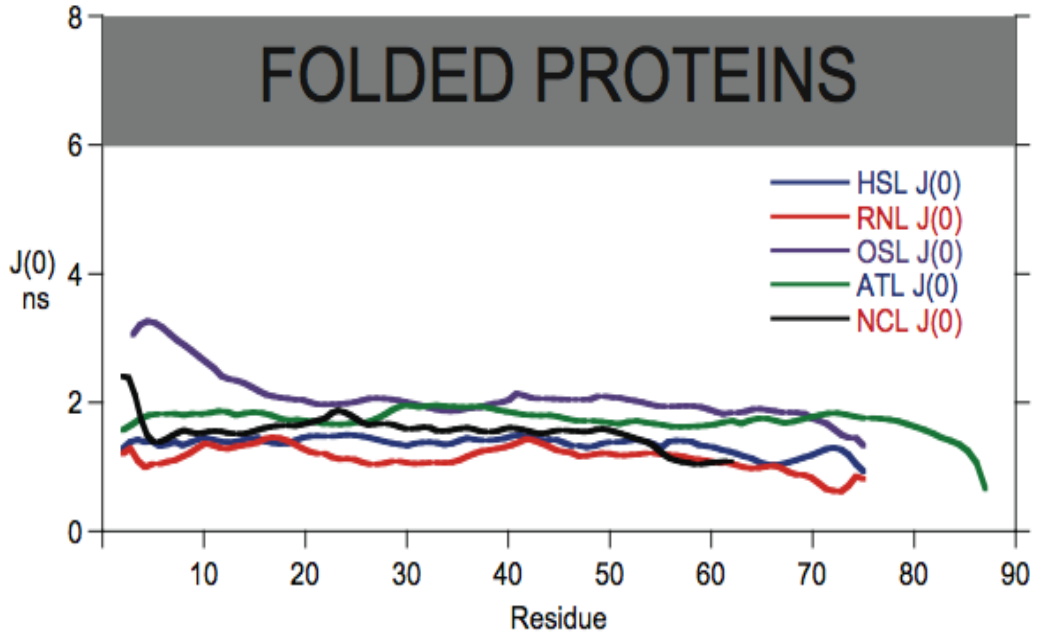
Daughdrill GW et al., J Mol Evol, 65: 277-288 (2007)

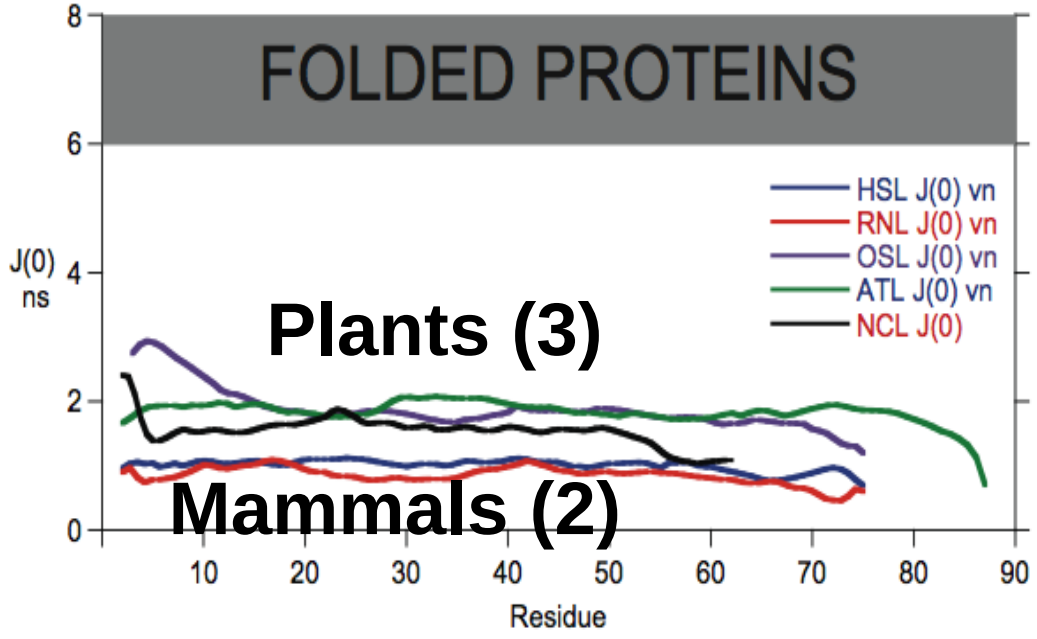
RPA70 Flexibility

Backbone dynamics of the RPA70 linkers are conserved in plants and mammals

J(0) – reorientation time

Smaller number, quicker reorientation, more flexible





Daughdrill, G.W., et al., J Mol Evol 65:277-88 (2007)

RPA70 Linker Conclusions

- **Linker length and flexibility is relatively well conserved despite lack of sequence conservation**
- **Mammals (human and rat) similar to each other and slightly more flexible than plants (rice and arabidopsis), which are also similar to each other**
- **Additional work shows that a likely reason for increased flexibility in mammalian linker is an increased glycine content compared to the plants**

IDPs: FIVE EVOLUTIONARY BEHAVIORS

- Conserved-sequence IDPs
- Variable-sequence IDPs
- →→ ***Nonconserved IDPs*** ←←
- INDELS = IDPs
- Fusions

Brown CJ et al., J Mol Evol 55: 100-110 (2002) ● ●

Chen JW et al., J Proteome Res 5: 579-587 (2006) ●

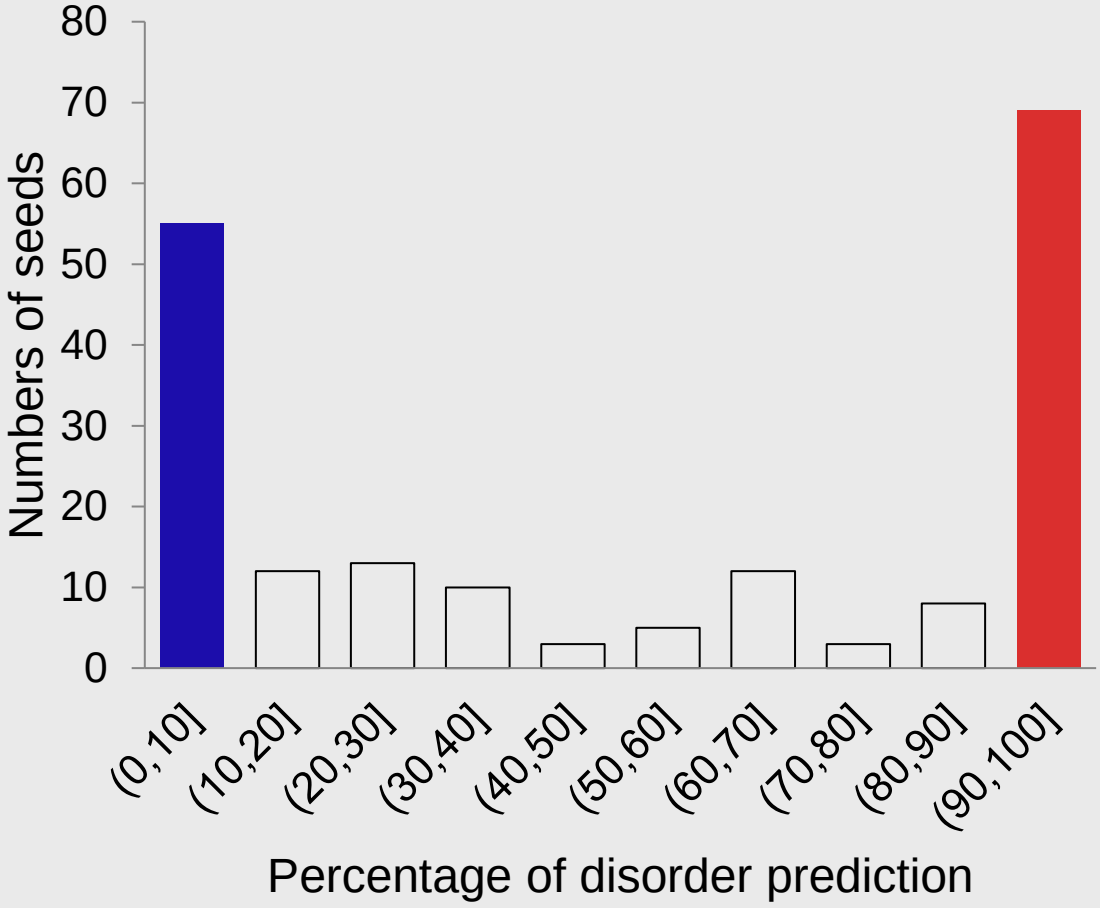
Bellay J et al., Genome Biology 12: R14 (2011) ● ● ●

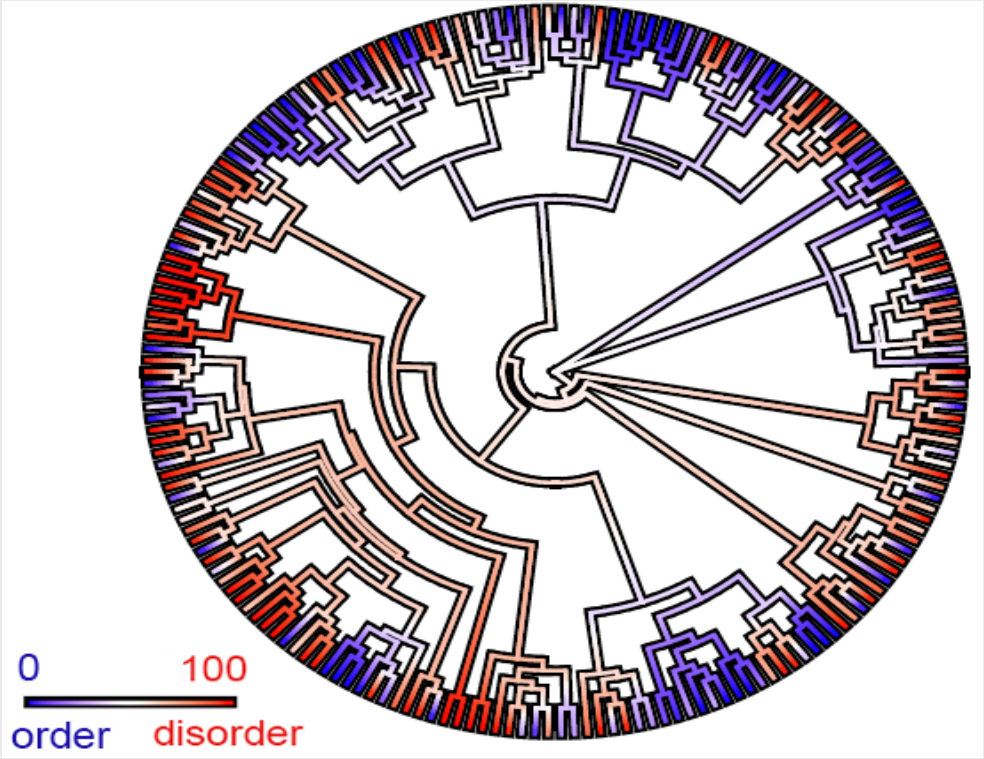
Fukuchi S et al., JMB 355: 845-857 (2006) ●

Brown CJ et al., COSB 21: 441-446 (2011) ●

Hegyi H et al., PLoS comp biol 5: e1000552 (2009) ●

Zinc Finger C2H2 (PF00096)





Zhou J et al., Prot Sci 28: 1652-1663 (2019)

Structured vs Disordered Domains

Structured Domains:

Structure is three to ten times more strongly conserved than sequence.

Illergard K et al., *Proteins* 2009; 77:499-508 (2009)

Disordered Domains:

For some disordered domains, sequence is more strongly conserved than lack of structure.

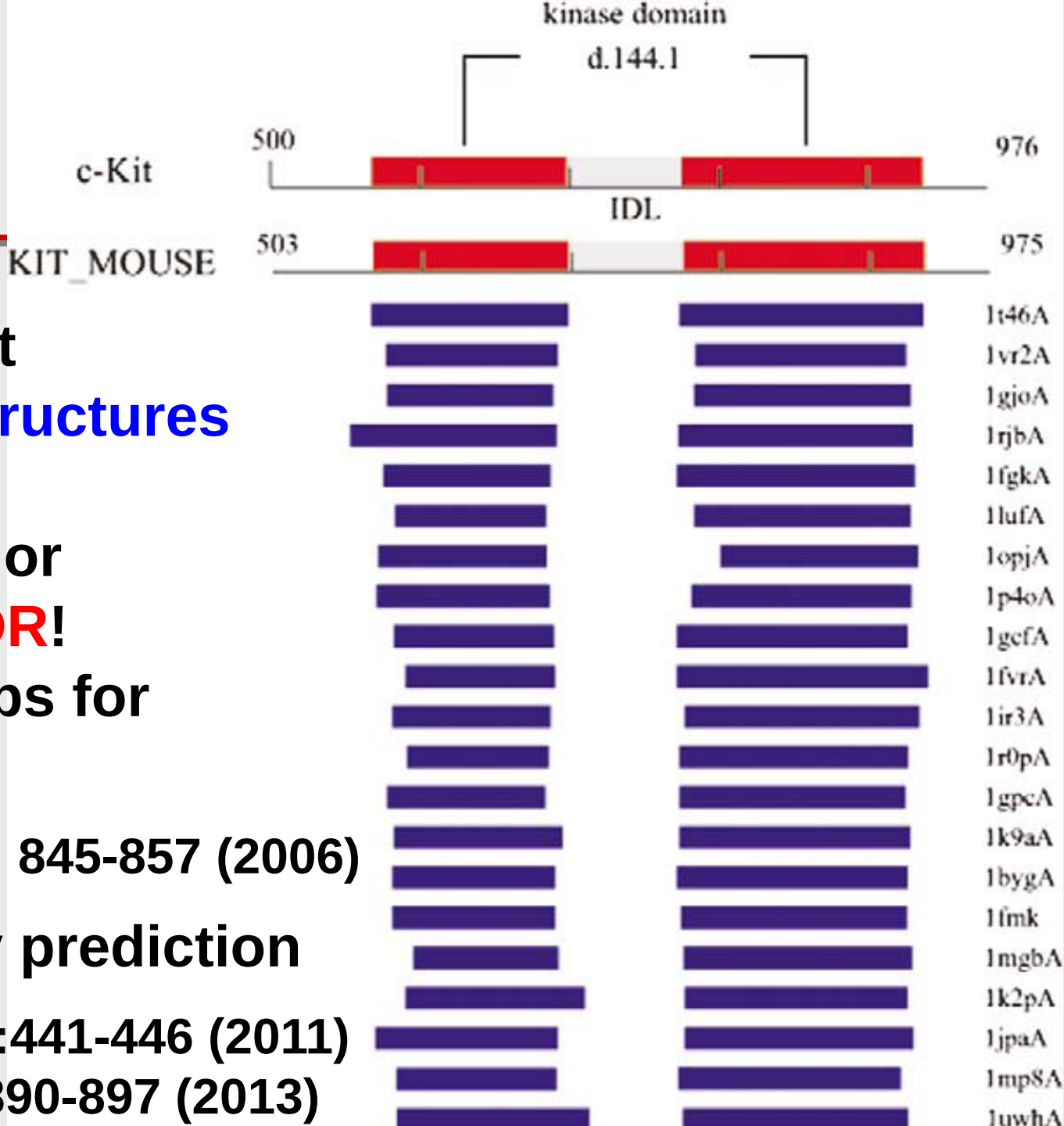
INDELs ARE Often IDRs

- Identify INDELs by sequence alignment
- Seek pair of PDB structures ± INDELS
- Is insert structured or an IDR? → It's an IDR!
- Repeat above 3 steps for additional proteins!

Fukuchi S et al., JMB 355: 845-857 (2006)

- INDELs are IDRs by prediction

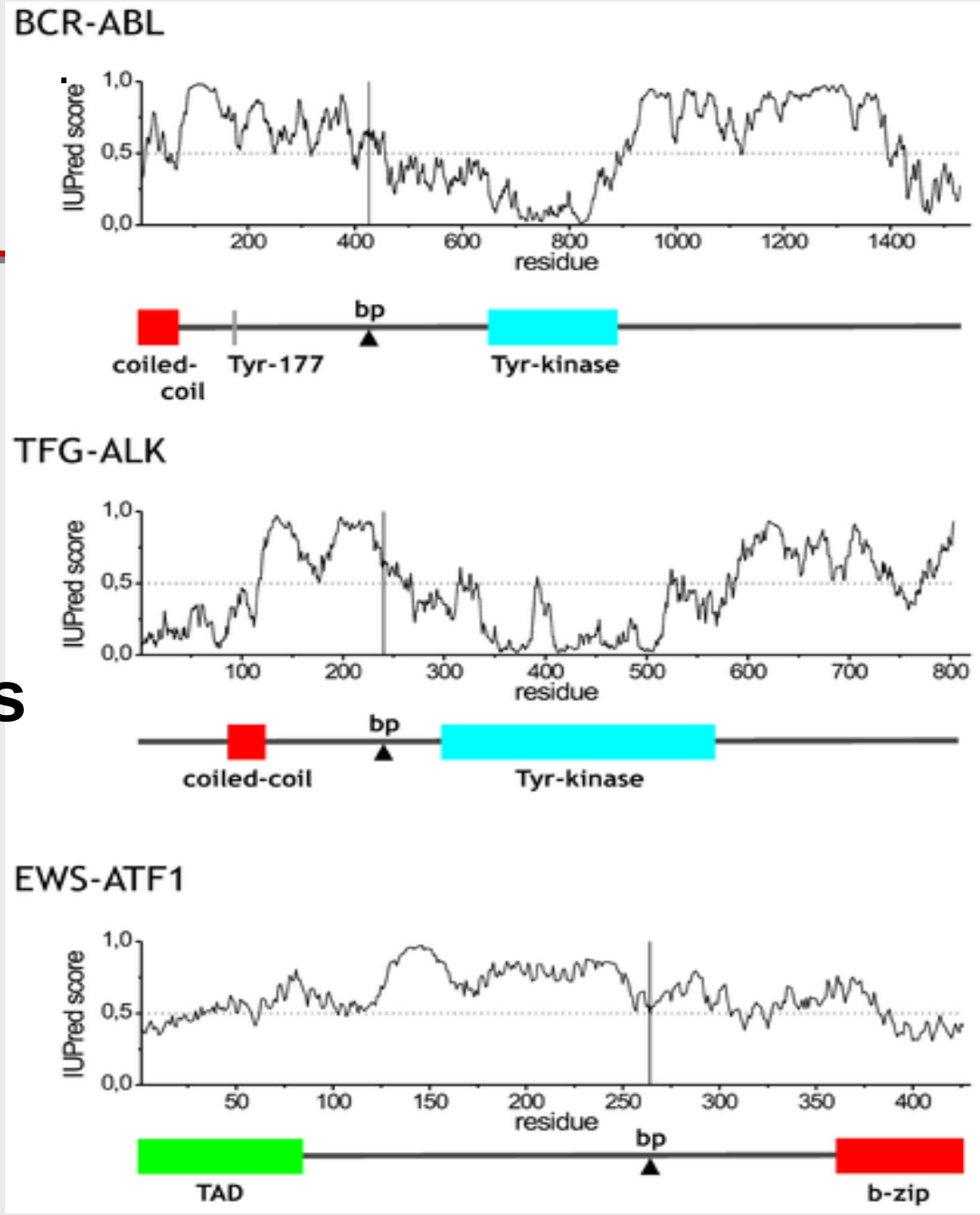
Brown CJ et al., COSB 21:441-446 (2011)
Light S et al., BBA 1834: 890-897 (2013)



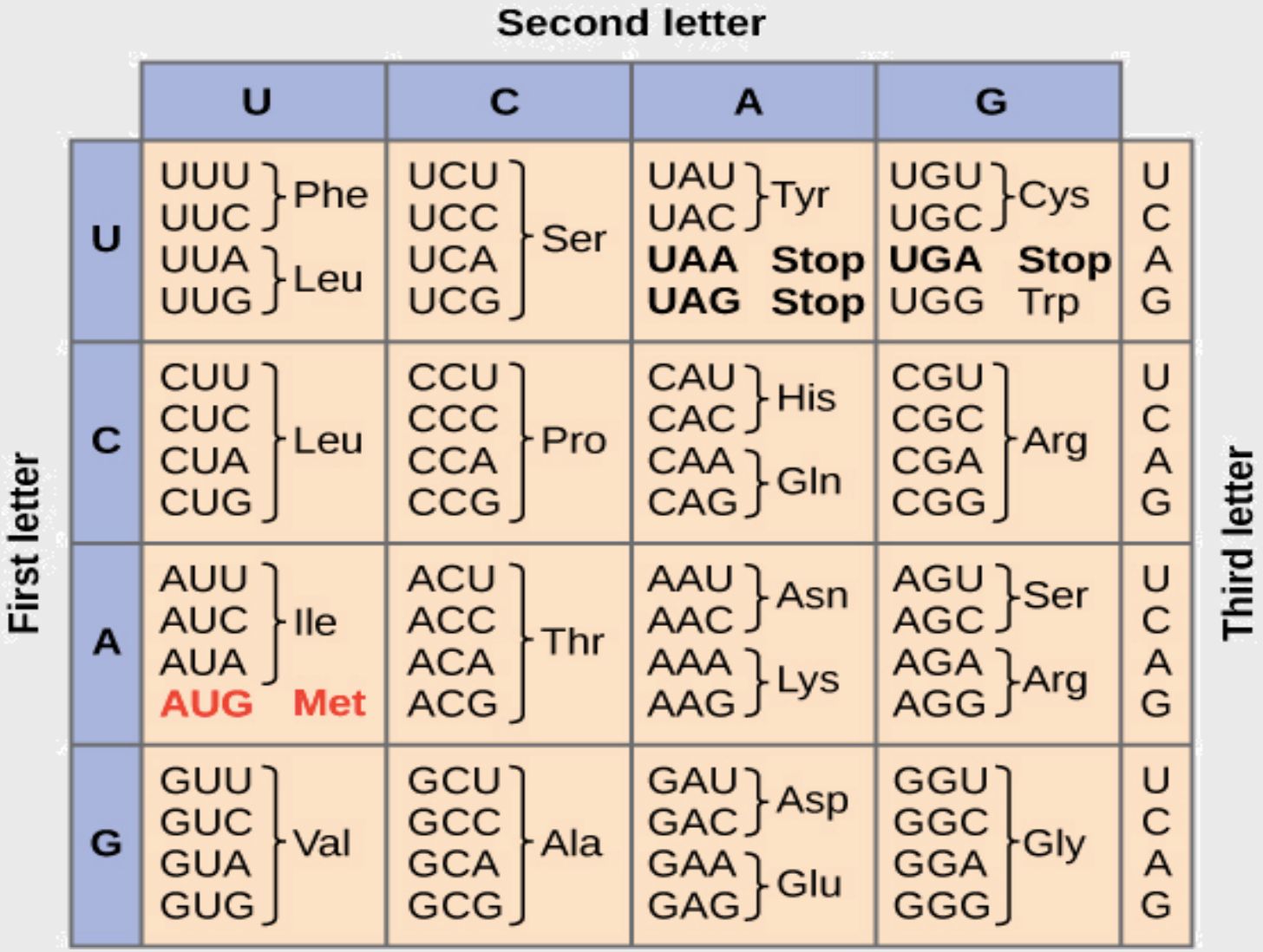
Protein Fusions

- Three oncogene proteins that result from fusions
- Note that fusions (bp) join two proteins at sites where both proteins are IDRs

Hegyi H et al., PLoS comp biol 5: e1000552 (2009)



Evolution of the Genetic Code



Note: Many AA determined by first two letters.

From this & other features, Crick, Woese, among others speculated that the genetic code evolved.

Evolution of the Genetic Code

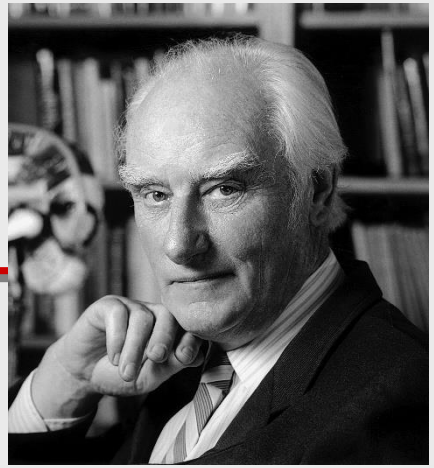
Carl Woese

Woese C.R. On the Evolution of the Genetic Code. *Proc. Nat'l Acad Sci USA* 54: 71-75 (1965)

Focused on relationship between codon position-specific error rates in E. coli and structure of genetic code. Highest error rates at third position.

No mention of selective advantage of late amino acids.

Evolution of the Genetic Code

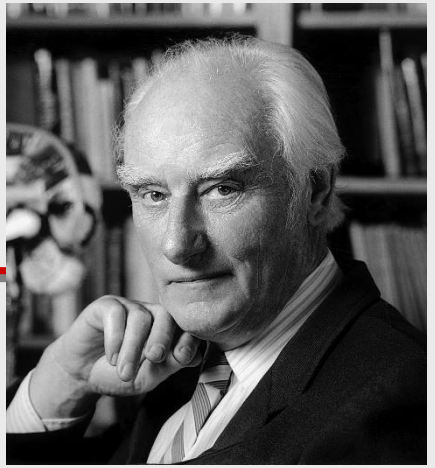


Francis Crick

Crick, F.C.H. Codon-Anti Codon Pair: The Wobble Hypothesis. *J. Mol. Biol.* 19:548-555 (1966)

Only about 35 tRNAs, yet 64 codons. Many tRNAs recognized by more than one codon. Non-Watson-Crick base pairing at third codon position due to movement or wobble at third codon position. Thus third position typically not important for which amino acid is coded.

Evolution of the Genetic Code



Francis Crick

Crick, F.C.H. The Origin of the Genetic Code. *J. Mol. Biol.* 38:367-379 (1968)

"In subsequent steps additional ammo acids were substituted *when they were able to confer a selective advantage*, until eventually the code became frozen in its present form."

No mention of possible selective advantage of the amino acids that were added later.

Evolution of the Genetic Code

My views based on work of Crick, Woese, others:

- **Original codons = two-letter for information, any third letter for stability (wobble)**
- **Thus, at most 4 x 4 = 16 amino acids initially.**
- **Other features → ~12+ early amino acids**

In addition to the Code itself, many other factors

- **Some AAs made from others**
- **Some synthesis pathways have many more steps than other pathways, etc.**
- **> 100 papers on genetic code evolution**

Evolution of the Genetic Code

Trifonov E.N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261:139–151 (2000)

Edward Trifonov

What is the selective advantage for adding the late amino acids?

Increasing sequence complexity

Too vague. Could not find a compelling hypothesis for the selective advantages of the amino acids added late to the genetic code.

Evolution of the Genetic Code

Edward Trifonov

Ernesto Di Mauro

Earlier → Later

Chronology: G A V D P S E L T R N K Q I C H F M Y W

ID Rank: P E S K Q H D R G A T C N V L M I Y F W

Disordered → Structured

Di Mauro, E, Dunker, A.K., & Trifonov, E. *Cellular Origin, Life in Extreme Habitats, and Astrobiology*. 22: 415-435 (2012)

Implications

- **If the genetic code evolved, and**
- **If early proteins lacked aromatic residues,**

then the selective advantage resulting from the added amino acids is due to

increased structural stability

Stable structure → Enzymes!

Strategy

1. Identify ancient proteins using criterion of universal occurrence (Several researchers have done this)
2. Identify ancient enzyme with fewest aromatics. This turned out to be Dephospho-CoA Kinase.
3. Using synthetic DNA and genetic engineering, clone express, and purify WT enzymes and leucine mutants with all aromatics replaced by leucines. Called DPCK-WT and DPCK-LH
4. Characterize and compare DPCK-WT and DPCK-LH.

Selected DPCK Proteins

Klara
Hlouchova
Charles University
Prague

DPCK: Dephospho CoA-Kinase

Functional Group: COG 0237 (Clusters of Orthologous Groups)

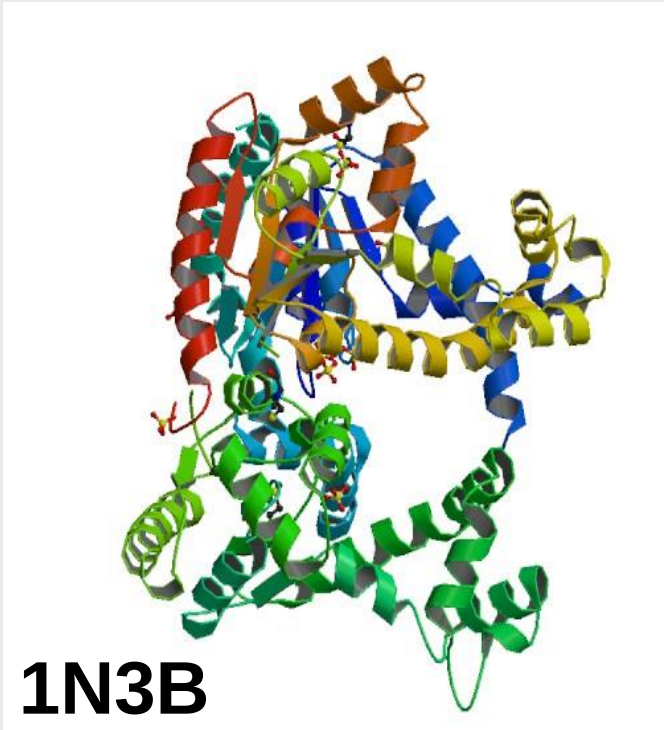
Eight known PDB structures :

***Aquifex aeolicus* : 2IF2**

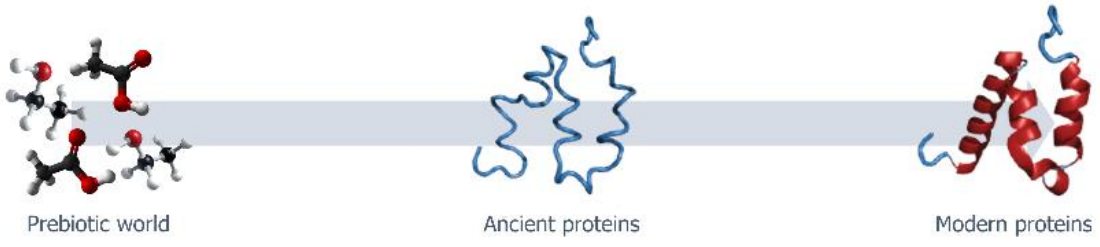
***Escherichia coli.* : 1N3B, 1T3H, 1VHL, 1VHT, 1VIY**

***Haemophilus influenzae* : 1JJV**

***Thermotoga maritima* : 2GRJ**



1N3B



A

Earlier → Later

(i)	G	A	V	D	P	S	E	L	T	R	N	K	Q	I	C	H	F	M	Y	W
(ii)	G	A	D	E	V	S	I	L	P	T	K	F	R	H	N	Q	C	Y	M	W
(iii)	P	E	S	K	Q	H	D	R	G	A	T	C	N	V	L	M	I	Y	F	W

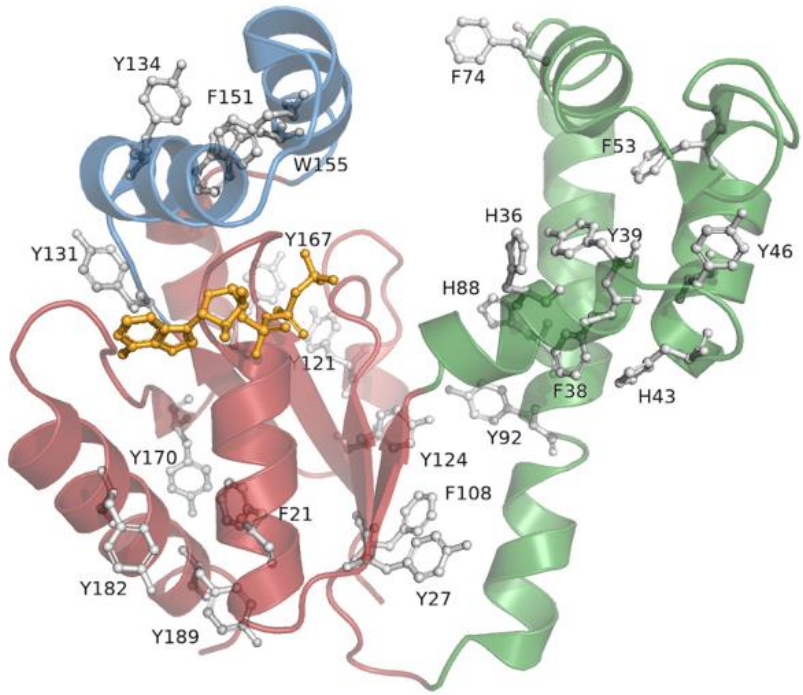
Disordered → Ordered

B

Position	21	27	36	38	39	46	53	74	88	92	108
DPCK-WT	F	Y	H	F	Y	Y	F	F	H	Y	F
DPCK-LH	L	L	H	L	L	L	L	L	H	L	L
DPCK-M	L	P	R	V	V	L	V	V	S	R	V

Position	121	124	131	134	151	155	167	170	182	189
DPCK-WT	Y	Y	Y	Y	F	W	Y	Y	Y	Y
DPCK-LH	L	L	L	L	L	L	L	L	L	L
DPCK-M	D	V	D	E	A	I	L	V	R	L

C



Enzyme	Substrate	K_m (μM)	V_{max} (μM/min)	k_{cat} (1/min)	K_{cat}/K_m (1/(μM*min))
DPCK-WT	dCoA (with 200 μM ATP)	24.3 ± 1.7	1.57 ± 0.17	49.0 ± 5.2	2.02
	ATP (with 200 μM dCoA)	12.7 ± 0.3	1.41 ± 0.05	43.8 ± 1.5	3.45
	ATP (without dCoA)	n.d.	n.d.	n.d.	n.d.
DPCK-LH	dCoA (with 200 μM ATP)	>200	n.d.	n.d.	n.d.
	ATP (with 200 μM dCoA)	65.9 ± 4.5	0.30 ± 0.01	1.42 ± 0.03	0.0215
	ATP (without dCoA)	40.2 ± 3.8	0.11 ± 0.013	0.52 ± 0.03	0.0130
DPCK-M	dCoA (with 200 μM ATP)	n.d.	n.d.	n.d.	n.d.
	ATP (with 200 μM dCoA)	32.3 ± 3.4	0.088 ± 0.003	0.23 ± 0.01	0.0071
	ATP (without dCoA)	29.2 ± 2.5	0.085 ± 0.002	0.22 ± 0.01	0.0075

n.d. - not detectable

Structural Characterization of DPCK-LH

DPCK-LH: • Unfolds in low urea • Trypsin sensitive • Binds ANS • NMR spectrum poorly dispersed • Greater hydrodynamic radius than DPCK-WT **From these data, DPCK-LH is likely molten globular; This work suggests early proteins without aromatics were likely IDPs!!**

Also, DPCK-LH is first enzyme that we know of without aromatic residues for stabilization of the hydrophobic core.

Structural Changes Upon ATP Binding

DPCK-LH:

Upon ATP addition, the hydrodynamic radius of DPCK-LH decreases by about 20%, roughly 10 times greater than the decrease observed for DPCK-WT.

The binding energy used to fold the enzyme takes away from energy available for catalysis; thus much reduced catalytic rates and efficiency for the DPCK-LH mutant.

Intrinsically Disordered Proteins

THANK YOU!!!

(kedunker@iupui.edu)

(klara.hlouchova@natur.cuni.cz)

Funding: NIH, NSF, INGEN,
IUPUI Signature Centers Initiative