Supplemental Data - Figures

Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification - Moesa et al.
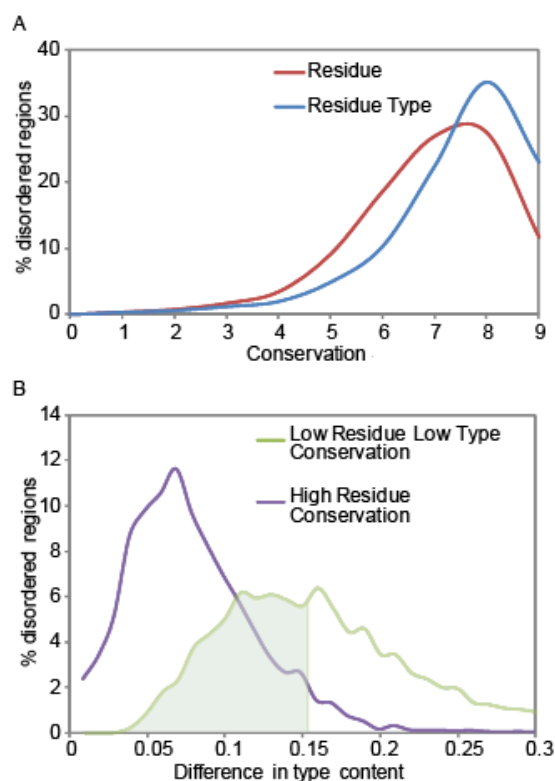

Figure S1. A) Distribution of the residue and residue type (positive, negative, polar, hydrophobic, special) conservation scores of human intrinsically disordered regions predicted by IUPred. B) Distribution of the difference in residue type content (denoted by Euclidean distance) between orthologous proteins for intrinsically disordered regions predicted by IUPred with high residue conservation (purple) and low residue, low residue type conservation (green). A smaller difference in type content indicates greater conservation of chemical composition.
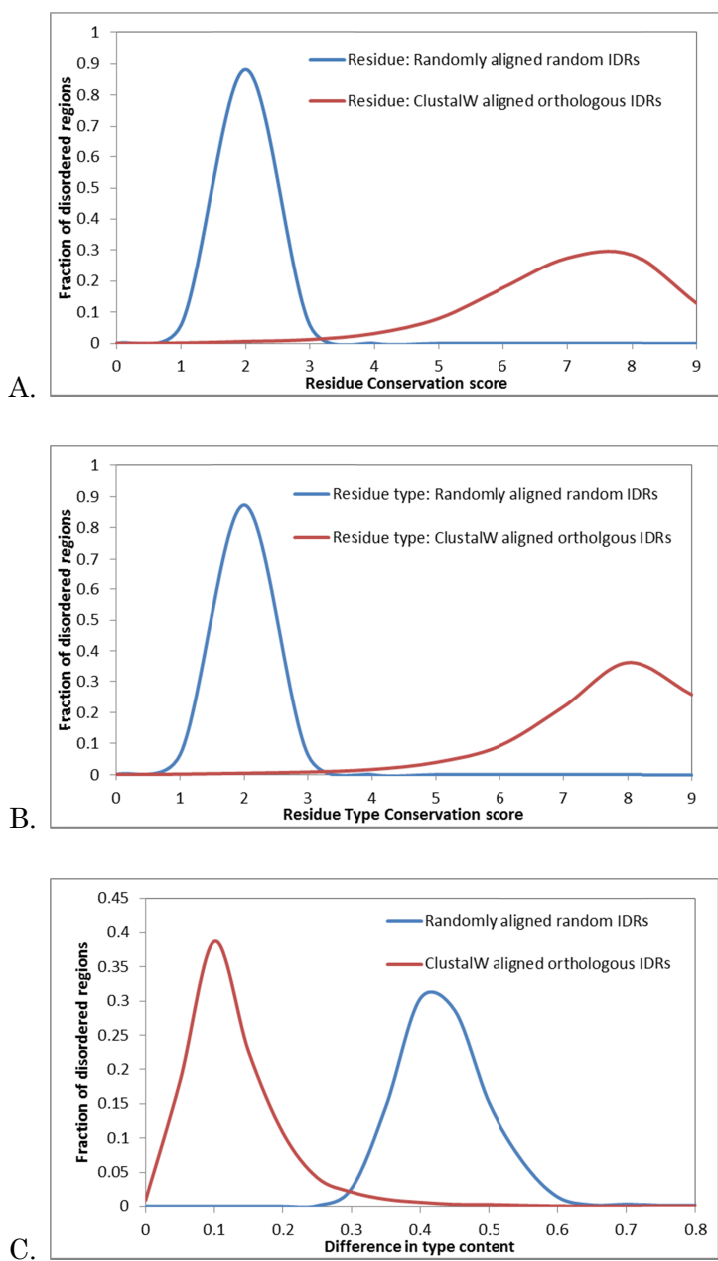
A.

B.

C.

Figure S2. A) Residue conservation score, B) residue type conservation score and C) difference in type content in randomly aligned IDRs from human and mouse proteins (blue) and orthologous IDRs aligned by CLUSTALW (red). Randomly selected human and mouse IDRs were aligned by aligning their N and C-terminii and randomly inserting gaps into either of the sequences. The distributions show that the alignments of IDRs made by CLUSTALW are significantly better than random ($p \ll 0.01$, Wilcoxon rank sum test).
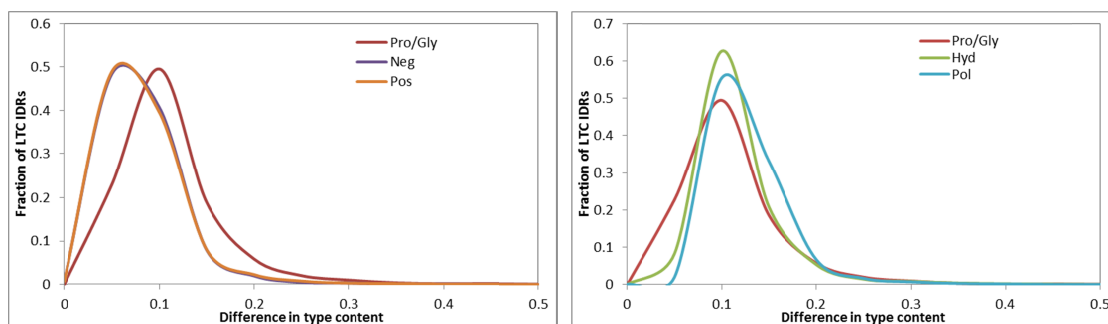
A.

B.

Figure S3. Difference in residue type content between orthologous IDRs for Pro+Gly compared to A) charged residues, and B) polar and hydrophobic residues. Difference in residue type content decreased with increasing conservation of chemical composition in IDRs. The figures signify that the composition of Pro and Gly residues is poorly conserved compared to charged residues, but better conserved than polar and hydrophobic residues in LTC IDRs (p << 0.01, Wilcoxon rank sum test).
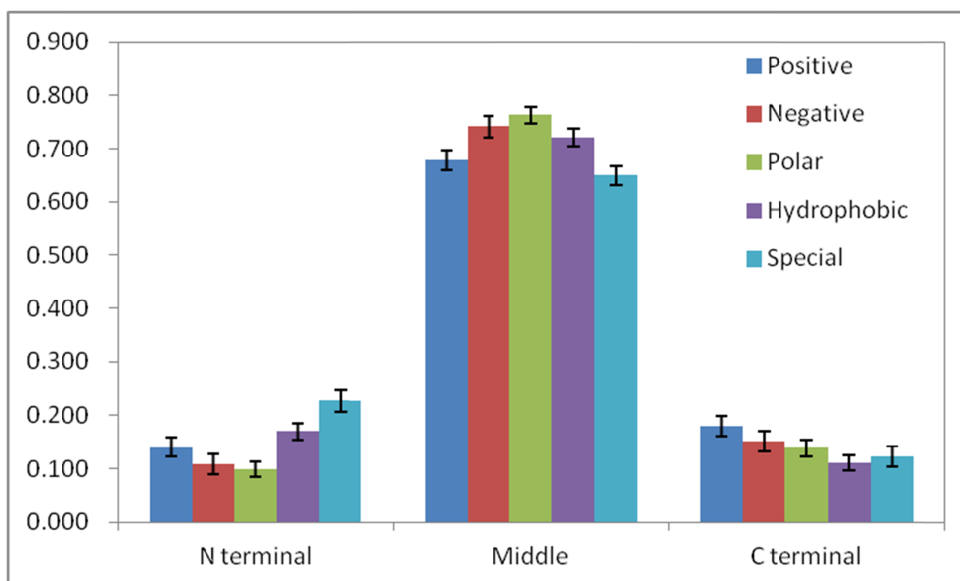
Figure S4. 95% Confidence intervals for the proportion of IDRs in each cluster at a specific location shown in Figure 4B.

High Residue conservation (HR)

Low Residue High Type content conservation (HTC)

Low Residue Low Type content conservation (LTC)

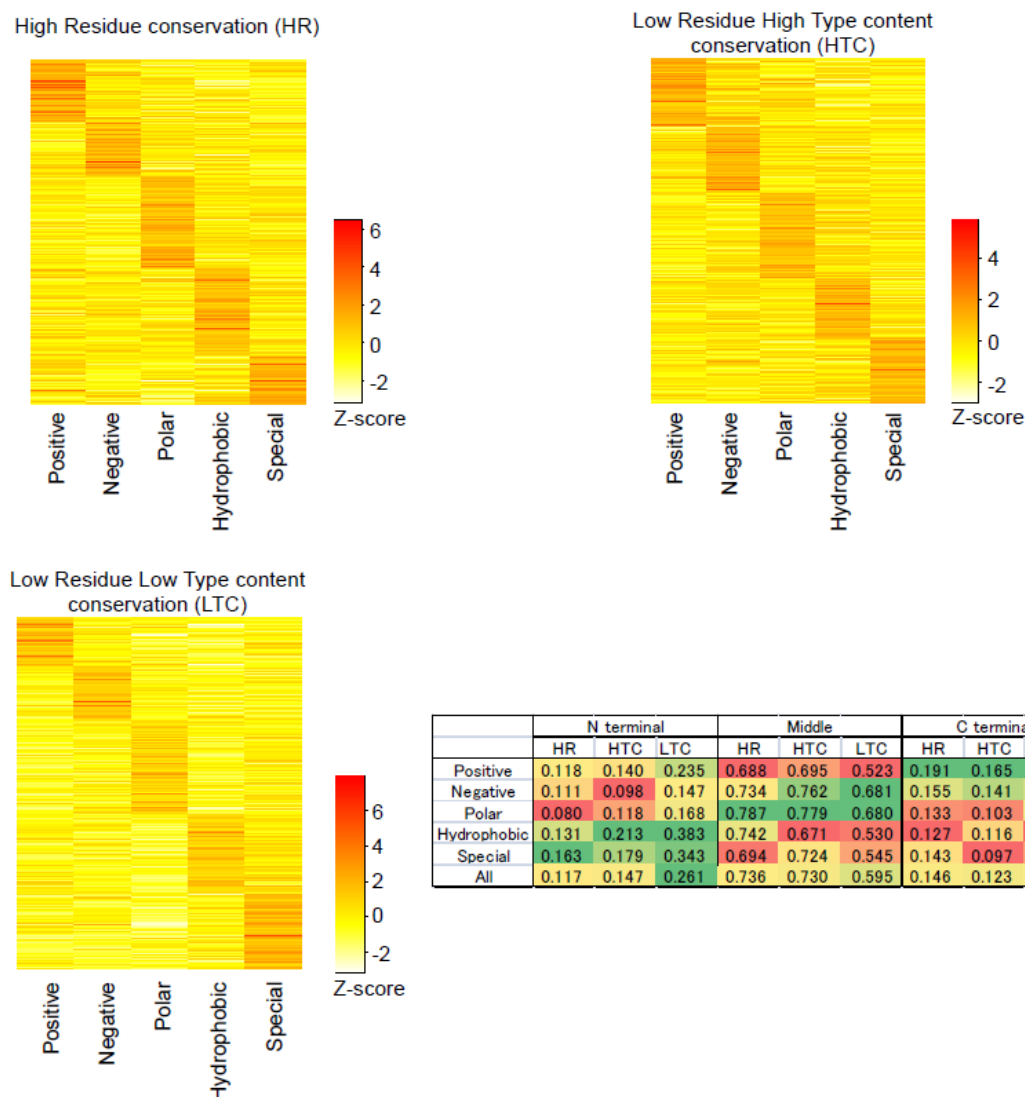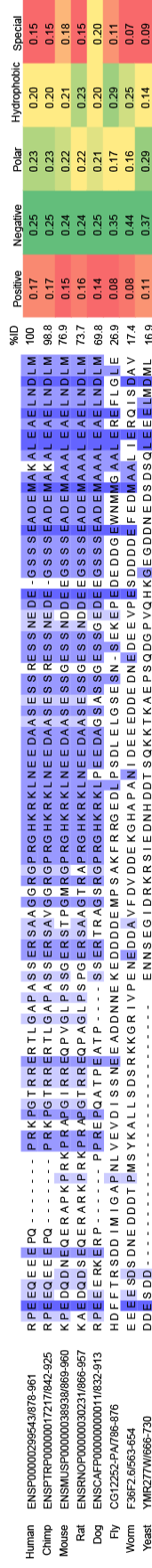|	N terminal			Middle			C terminal		
	HR	HTC	LTC	HR	HTC	LTC	HR	HTC	LTC
Positive	0.118	0.140	0.235	0.688	0.695	0.523	0.191	0.165	0.243
Negative	0.111	0.098	0.147	0.734	0.762	0.681	0.155	0.141	0.171
Polar	0.080	0.118	0.168	0.787	0.779	0.680	0.133	0.103	0.152
Hydrophobic	0.131	0.213	0.383	0.742	0.671	0.530	0.127	0.116	0.088
Special	0.163	0.179	0.343	0.694	0.724	0.545	0.143	0.097	0.112
All	0.117	0.147	0.261	0.736	0.730	0.595	0.146	0.123	0.144

Figure S5. HR, HTC and LTC IDR clusters based on chemical composition. Each row in the heatmap corresponds to an IDR and each value indicates the z-score for the fraction of five types of residues in one IDR. Location preferences of HR, HTC and LTC IDRs in 5 types of clusters are shown in the adjacent tables. Columns are colored by fraction of IDRs in each cluster at a specific location. Green indicating the highest propensity at a specific location while red indicates a low propensity with yellow denoting an average value. IDRs separated by conservation show distinct location preference based on their chemical composition.

A.

		Sequence
Human	ENSP00000320295/393-431	AFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAE
Chimp	ENSPTRP00000014499/393-431	AFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAE
Mouse	ENSMUSP00000071134/393-431	AFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAE
Rat	ENSRNOP00000023452/393-431	AFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAE
Dog	ENSCAFP00000029328/393-431	AFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQDATAE
Yeast	YFL037W/393-431	AFLHWYTSEGMDELEFSEAESNMNDLVSEYQQYQEATVE

		%ID	Positive	Negative	Polar	Hydrophobic	Special
Human	ENSP00000299543/878-961	100	0.17	0.25	0.23	0.20	0.15
Chimp	ENSPTRP00000017217/842-925	98.8	0.17	0.25	0.23	0.20	0.15
Mouse	ENSMUSP00000038938/869-960	76.9	0.15	0.24	0.22	0.21	0.18
Rat	ENSRNOP00000030231/866-957	73.7	0.16	0.24	0.22	0.23	0.15
Dog	ENSCAFP00000000001/832-913	69.8	0.14	0.25	0.21	0.20	0.20
Fly	CG12252-PA/786-876	26.9	0.08	0.35	0.17	0.29	0.11
Worm	F36F2.6/563-654	17.4	0.08	0.44	0.16	0.25	0.07
Yeast	YMR277W/666-730	16.9	0.11	0.37	0.29	0.14	0.09

B.

Figure S6. A) Alignment and conservation of a highly conserved (HR) IDR in the human Tubulin beta-4 chain (ENSP00000320295, DisProt id: DP00114). This region is highly acidic and may be used to bind cations. B) Alignment of the C-terminal intrinsically disordered region in the TFIIF-associating CTD phosphatase, CTDP1 (ENSP00000029954, DisProt id: DP00177). This IDR forms an alpha helix when bound to the C-terminal region of Transcription factor IIF, RAP74, which it binds using a few conserved hydrophobic and negatively charged residues. The IDR shows relatively high conservation among mammals but is poorly conserved among lower eukaryotes. However, the high prevalence of negatively charged residues and the low abundance of positively charged and special residues is preserved throughout all orthologous proteins despite their poor conservation. Sequences are colored by levels of identity with the human IDR, also noted as %ID after the sequence. The fractions of residues are colored according to their abundance with red denoting a low propensity and green denoting a high propensity.