

## Corrigendum

# Robustness by intrinsically disordered C-termini and translational readthrough

April Snofrid Kleppe and Erich Bornberg-Bauer\*

Institute of Biodiversity and Evolution, University of Münster, Hüfferstr. 1, 48151 Münster, Germany

*Nucleic Acids Research*, 2019, 46(19), 10184–10194, <https://doi.org/10.1093/nar/gky778>

Post publication, we discovered a programming error in the script concerned with calculation of intrinsic disorder. The error was in a programming script, that was supposed to extract the ratio of disordered residues, by using Iupred, of the individual genes. Iupred is a published software program, that infers how disordered a residue (amino acid) is of a given protein, yielded into a textfile. From the textfile, one can extract the ratio of disordered residues, which is the number of how many amino acids are disordered of the given protein. However, the ratio of disordered residues for each gene, were added to a list and the mean value of the list was reported for the sequence in question. This was done instead of reporting the individual value of the sequence. We have added the lines of code where this mishap occurs, further down. Essentially, this error was committed by confusing ‘ratio of disordered residues’ with ‘mean disorder’, during programming. This loop was ongoing, and the mean value was affected by the initial values added to the list.

The lines of incorrect code in question:

```
disorder_list.append(seq_disorder_fraction[0])  
#append ratio of disordered residues to a list  
disorder_mean = np.mean(disorder_list)  
#disorder mean was taken from previous list  
out_list.append((seq_id, disorder_mean, anchor_mean))  
# out_list was written to a text file
```

Lines of corrected code in bold:

```
disorder_list.append(seq_disorder_fraction[0])  
#append ratio of disordered residues to a list  
disorder_mean = np.mean(disorder_list)  
#disorder mean was taken from previous list  
out_list.append((seq_id, seq_disorder_fraction[0], anchor_mean))  
# out_list was written to a text file
```

---

\*To whom correspondence should be addressed. Erich Bornberg-Bauer. Tel: +49 251 83 21630; Fax: +49 251 83 24668; Email: [ebb@uni-muenster.de](mailto:ebb@uni-muenster.de)

Several weeks after the submission and acceptance of the published paper, the same pipeline was applied for new data sets. Upon applying the same pipeline to new data sets, extreme anomalies were observed. Upon investigating the raw files, generated by Iupred, we took single samplings of individual genes and compared to the data given by the script. The compared data differed to such an extent that it indicated that something was incorrect with the script. We did not discover this programming flaw previously, because when we inferred our data with experimentally verified disorder (as reported in the initial submission), these leaky proteins were indeed disordered at the C-termini. It was an unfortunate co-incidence that those who were experimentally verified also showed disorder in our data set. Otherwise we might have caught the programming flaw earlier.

After the finding of this programming flaw, we re-examined the sets regarding disorder. We find that when comparing the disordered C-termini between the sets, with the correct Iupred values, there are significantly more disordered C-termini in the leaky set. The difference is less dramatic than we initially reported, but there is still a significant difference between the sets. We believe the difference we find is of a significant relevance as disordered C-termini can explain the seemingly high tolerance, and thereby prevalence, of translational readthrough.

The following corrections were made to the published article:

## ABSTRACT

**Old:** Our main finding is that proteins undergoing TR are highly expressed and have intrinsically disordered C-termini.

**New:** Our main finding is that proteins undergoing TR are highly expressed and have a higher proportion of intrinsically disordered C-termini.

## RESULTS AND DISCUSSION

### Protein characteristics

#### *Error prone proteins have highly disordered C-termini*

**Old:** However, we did find the non-leaky set to be significantly different from the leaky and semi-leaky set with respect to disorder distribution (Figure 2 and Supplementary Table S6): leaky and semi-leaky proteins have an overall wider distribution and lower ratio of disordered residues than non-leaky proteins.

**New:** We did not find the distributions of disordered residues to differ between the sets (see Figure 2).

**Old:** Moreover, we analysed the last 30 amino acids of the peptide chains separately and found a significant difference between sets. The majority of genes belonging to the non-leaky set have a ratio of disordered residues  $<0.5$ , whereas the vast majority of genes belonging to the leaky and semi-leaky set have a ratio of disordered residues  $>0.5$  (Figure 3). We found five of our proteins to be curated in the DisProt database (67) (Supplementary Table S1).

**New:** Moreover, we analysed the last 30 amino acids of the peptide chains separately. All sets have a high frequency of disordered C-termini (see Figure 3 and Supplemental Figure S6.), but the leaky set has a significantly higher proportion than the non-leaky set (Mann Whitney one-sided rank test,  $p$ value 0.03). We found five of our proteins to be curated in the DisProt database (67) (see Table S1).

## FIGURE 2 CAPTION

**Old:** Ratio of disordered residues of full protein sequences. The Y-axis displays density of sequences and the X-axis display ratio of disordered residues. The colours display what set the proteins belong too. (A) The leaky set is intermediate between the non-leaky and semi-leaky sets. The leaky set (red) is mostly overlapping with the semi-leaky set (blue), but also overlapping with the non-leaky set (green). (B) The leaky and semi-leaky sets are clustered as one (purple), whereas non-leaky is maintained unaltered (green).

**New:** Ratio of disordered residues of full protein sequences. The Y-axis displays density of sequences and the X-axis display ratio of disordered residues. The colours display what set the proteins belong too. A: The leaky set (red) is mostly overlapping with the semi-leaky set (blue), but also overlapping with the non-leaky set (green). B: The leaky and semi-leaky sets are clustered as one (purple), whereas non-leaky is maintained unaltered (green). The leaky and semi-leaky sets have a significantly higher proportion of disordered residues (Mann Whitney test,  $P$ -value 0.005,  $U$ -value 966).

## FIGURE 3 CAPTION

**Old:** Ratio of disordered residues of last 30 amino acids of protein sequences in sets. The Y-axis displays density of sequences and the X-axis displays ratio of disordered residues. Leaky and semi-leaky sets are clustered as one (purple), whereas non-leaky is maintained unaltered (green).

**New:** Ratio of disordered residues of last 30 amino acids of protein sequences in sets. The Y-axis displays density of sequences and the X-axis displays ratio of disordered residues. Leaky and semi-leaky sets are clustered as one (purple), whereas non-leaky is maintained unaltered (green). Many proteins of both error prone and non-leaky set have intrinsically disordered C-termini, but the C-termini of error-prone proteins are more disordered.

## FIGURE S6

Supplementary Figure S6 has been replaced.

## TABLE S6

The following *P*-values have been corrected:

Sets: Variable	Z-value	New <i>P</i> -value	Old <i>P</i> -value
Leaky vs Semi-leaky: disorder CDS	1.065	0.287	0.078
Leaky vs Non-leaky: disorder CDS	-0.751	0.453	0.0
Semi-leaky vs Non-leaky: disorder CDS	-1.225	0.221	0.0
Leaky vs Semi-leaky: disorder C-termini	1.055	0.291	0.847
Leaky vs Non-leaky: disorder C-termini	1.893	0.058	0.0
Semi-leaky vs Non-leaky: disorder C-termini	-0.192	0.848	0.0

## TABLE S7

The following R-squared and *P*-value have been corrected:

Variables	R-squared	<i>P</i> -value	R-squared*	<i>P</i> -value*
<b>Old</b> GC UTR and disorder CDS	0.0052	0.0321	0.0348	0.0022
<b>New</b> GC UTR and disorder CDS	0.0071	0.0094	0.0364	0.002
<b>Old</b> GC UTR and disorder C-termini	0.0078	0.0058	0.0379	0.0013
<b>New</b> GC UTR and disorder C-termini	0.007	0.0105	0.0367	0.0019
<b>Old</b> Gene expression and disorder CDS	0.0088	0.0029	ins	ins
<b>New</b> Gene expression and disorder CDS	ins	ins	0.0182	0.0464
<b>Old</b> Gene expression and disorder C-termini	0.0237	0.0	ins	ins
<b>New</b> Gene expression and disorder C-termini	ins	ins	ins	ins
<b>Old</b> length and disorder CDS	0.005	0.0363	0.0515	0.0001
<b>New</b> length and disorder CDS	ins	ins	0.0892	0.0
<b>Old</b> length and disorder C-termini	0.0185	0.0	0.07	0.0
<b>New</b> length and disorder C-termini	ins	ins	0.0571	0.0001