

Does order matter in protein sequence alignment?

December 2 2015, by Dr Colm Ryan

What if the order in which you performed addition mattered, and the sum of '2 + 3 + 4' gave you a different answer to the sum of '4 + 3 + 2'?

Accountants everywhere would cry as their books ceased to balance. You might find yourself pausing at the supermarket wondering if you could save money by buying your milk before your toothpaste. Eurovision fans may value a 'douze points' early in the voting more than one later on.

Fortunately, there is no new evidence to suggest that order matters in addition. However, researchers at UCD have recently found that order matters a surprising amount for sequence alignment, an important part of modern genetic analyses.

Sequence alignment is used to understand similarities and differences between proteins found in different species. Proteins are the building blocks of life and carry out most of the functions in our cells. Consequently understanding proteins and their function is a key part of biology.

We can use pairwise sequence alignment to identify which parts of a specific [protein](#) are identical between a pair of species (e.g. humans and chimps). Using multiple sequence alignment, we can identify those parts of a protein that are conserved in all mammals or in even larger groups of species.

Histone H1 (residues 120-180)

```

HUMAN KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
MOUSE KKAAPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKVVVKPVKASKPKKAKTVK
RAT   KKAAPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKIVKVKPVKASKPKKAKPVK
COW   KKAAPKKAASKAPSKKPKATPVKKAKKKPAATPKKTKKPKTVKAKPVKASKPKKTKPVK
CHIMP KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
***.*****:***** *****:**** *.*****:* **

```

Sequence alignment showing a region of a protein (Histone H1) that is highly conserved across humans, mice, rats, cows and chimps. Dark grey indicates parts of the protein (amino acids) that are identical across all of these species. Credit: Thomas Shafee, licensed under CC-BY-SA 4.0

Very large sequence alignments help us understand which parts of proteins are important; if part of a protein is identical in all mammals then it's probably important and also gives us some insight into the three dimensional structure of proteins (as parts of a protein that are close together in 3D space tend to change together across species).

PhD student Kieran Boyce, Dr Fabian Sieviars and Professor Des Higgins in UCD Conway Institute & Systems Biology Ireland found that for large protein sequence alignments the order in which sequences are compared matters, i.e. the alignment that you get out of a sequence alignment programme depends on the order that you input your sequences into the programme.

This finding is surprising, as people have been performing sequence alignments for decades without knowing how dependent the results are on the input order. It is important because it suggests that most scientific publications that make use of large multiple sequence alignments

probably have not provided sufficient information to reproduce their results.

Reproducibility is an important part of science. Consequently, when performing sequence alignment, most scientists will provide details of the sequence alignment programme and the settings they used. The findings from Boyce et al suggest that the order is also an important setting that may need to be provided from now on. A question left open by the paper is how we can make sequence alignment programmes ignore order, or how we can choose the best possible ordering.

More information: Kieran Boyce et al. Instability in progressive multiple sequence alignment algorithms, *Algorithms for Molecular Biology* (2015). [DOI: 10.1186/s13015-015-0057-1](https://doi.org/10.1186/s13015-015-0057-1)

Provided by University College Dublin

Citation: Does order matter in protein sequence alignment? (2015, December 2) retrieved 23 April 2024 from <https://phys.org/news/2015-12-protein-sequence-alignment.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--