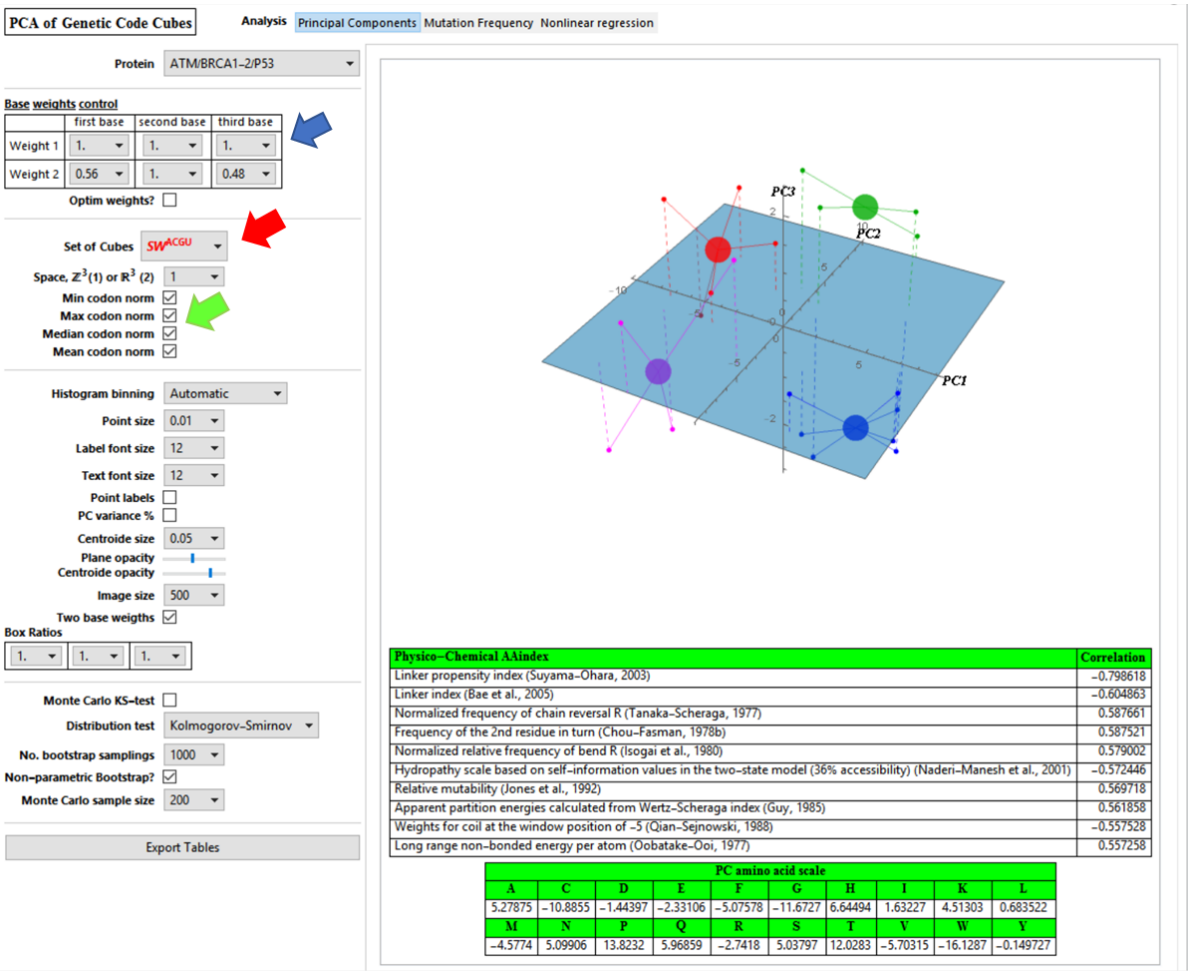
# Introductory Snapshots to the Genetic-Code-Scales_of_Amino-Acids CDF

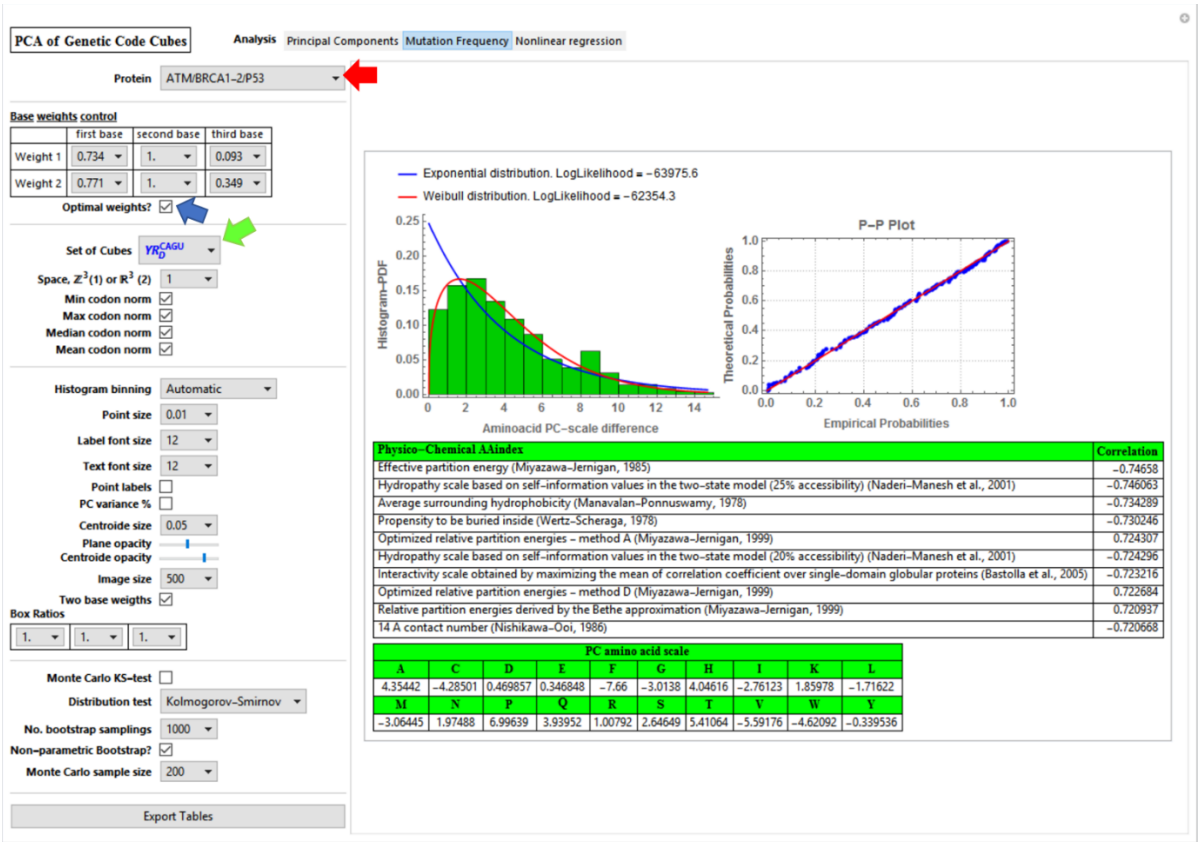
Two sets of weight (blue arrow) can be selected to evaluated Eq. 5 from the main text. The subset of genetic-code cubes where Eq. 5 will be evaluated can be selected as well (red arrow). By default, all the statistics (Min, Max, Mean and Median) of codon norms are selected, but the user could choose between them (green arrow).




| Physico-Chemical AAindex | Correlation |
|---|---|
| Linker propensity index (Suyama-Ohara, 2003) | -0.798618 |
| Linker index (Bae et al., 2005) | -0.604863 |
| Normalized frequency of chain reversal R (Tanaka-Scheraga, 1977) | 0.587661 |
| Frequency of the 2nd residue in turn (Chou-Fasman, 1978b) | 0.587521 |
| Normalized relative frequency of bend R (Isogai et al., 1980) | 0.579002 |
| Hydropathy scale based on self-information values in the two-state model (36% accessibility) (Naderi-Manesh et al., 2001) | -0.572446 |
| Relative mutability (Jones et al., 1992) | 0.569718 |
| Apparent partition energies calculated from Wertz-Scheraga index (Guy, 1985) | 0.561858 |
| Weights for coil at the window position of -5 (Qian-Sejnowski, 1988) | -0.557528 |
| Long range non-bonded energy per atom (Oobatake-Ooi, 1977) | 0.557258 |

| PC amino acid scale | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | C | D | E | F | G | H | I | K | L |
| 5.27875 | -10.8855 | -1.44397 | -2.33106 | -5.07578 | -11.6727 | 6.64494 | 1.63227 | 4.51303 | 0.683522 |
| M | N | P | Q | R | S | T | V | W | Y |
| -4.5774 | 5.09906 | 13.8232 | 5.96859 | -2.7418 | 5.03797 | 12.0283 | -5.70315 | -16.1287 | -0.149727 |

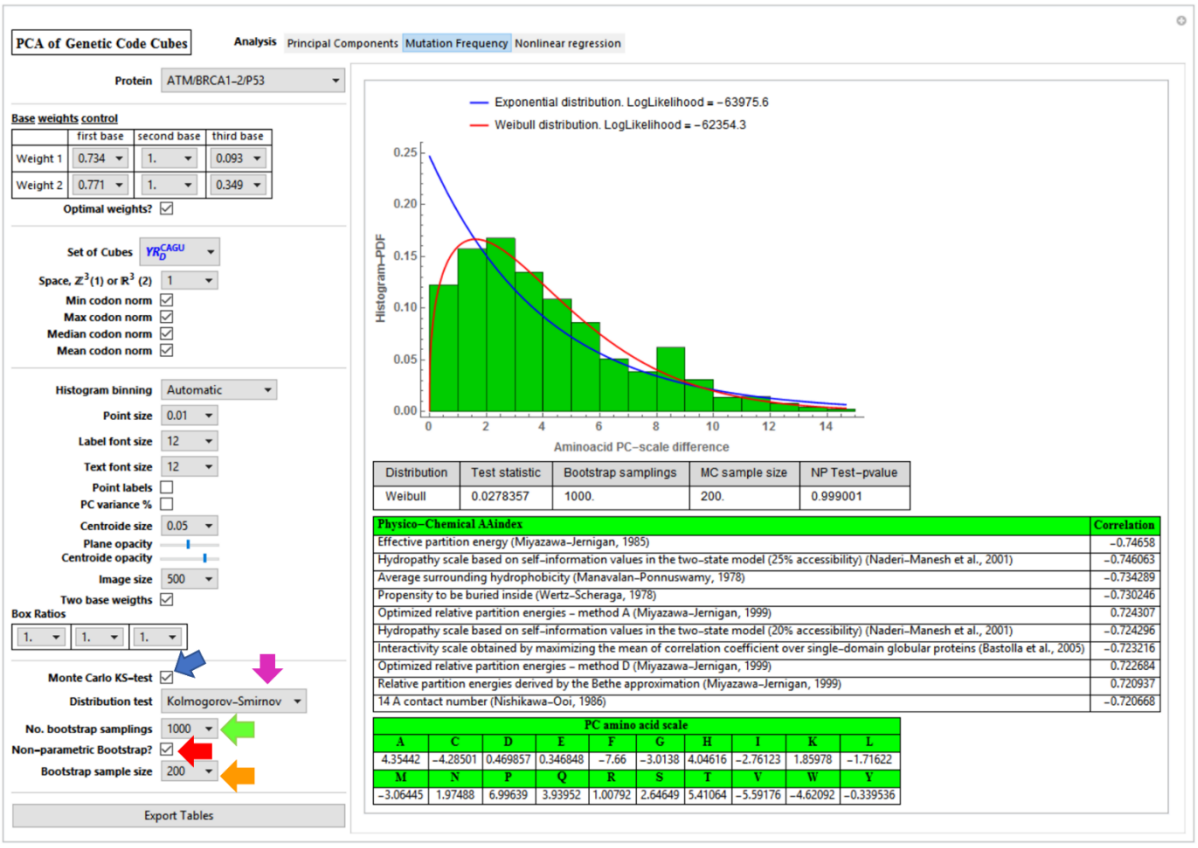
If “Optimal weights” is selected (below, blue arrow), then weights are automatically changed to the best values found for the selected set of genetic-code cubes (green arrow) and on the set of protein sequence alignments of the protein selected (red arrow). In general, these weights are local optima, which can be found by applying genetic algorithm (in the present case) or any other suitable optimization algorithm. If the “Mutation Frequency” tab is selected in the “Analysis” menu, some fitting details of Weibull distribution model estimated for the given set of protein sequence alignment are given, as well as, the amino acid PC-scale and its correlations with amino acid physicochemical indexes found in AAindex.




| Physico-Chemical AAindex | Correlation |
|---|---|
| Effective partition energy (Miyazawa-Jernigan, 1985) | -0.74658 |
| Hydropathy scale based on self-information values in the two-state model (25% accessibility) (Naderi-Manesh et al., 2001) | -0.746063 |
| Average surrounding hydrophobicity (Manavalan-Ponnuswamy, 1978) | -0.734289 |
| Propensity to be buried inside (Wertz-Scheraga, 1978) | -0.730246 |
| Optimized relative partition energies - method A (Miyazawa-Jernigan, 1999) | 0.724307 |
| Hydropathy scale based on self-information values in the two-state model (20% accessibility) (Naderi-Manesh et al., 2001) | -0.724296 |
| Interactivity scale obtained by maximizing the mean of correlation coefficient over single-domain globular proteins (Bastolla et al., 2005) | -0.723216 |
| Optimized relative partition energies - method D (Miyazawa-Jernigan, 1999) | 0.722684 |
| Relative partition energies derived by the Bethe approximation (Miyazawa-Jernigan, 1999) | 0.720937 |
| 14 A contact number (Nishikawa-Ooi, 1986) | -0.720668 |

| PC amino acid scale | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | C | D | E | F | G | H | I | K | L |
| 4.35442 | -4.28501 | 0.469857 | 0.346848 | -7.66 | -3.0138 | 4.04616 | -2.76123 | 1.85978 | -1.71622 |
| M | N | P | Q | R | S | T | V | W | Y |
| -3.06445 | 1.97488 | 6.99639 | 3.93952 | 1.00792 | 2.64649 | 5.41064 | -5.59176 | -4.62092 | -0.339536 |

If Monte Carlo KS-test is selected (below, blue arrow), then a table with the test results is shown. We can choose between Kolmogorov-Smirnov and Kuiper tests (magenta arrow), the number of bootstrap (green arrow) samplings and sample size (orange arrow). Parametric and non-parametric bootstrap options are available as well.




The “Nonlinear regression” tab from “Analysis” menu summarize the fitting of Weibull model on the set of protein sequence alignment selected. Different setting for weight values can be given unselecting “Optimal weights” and selecting a distinct set of genetic-code cubes. Additional “cosmetic” features are added. It is worthy to notice that the computation could take variable time for each protein selected. The processing time will vary depending on the computer processor power used to run the CDF. Unfortunately, I could not find the way to add a progress indicator that could work inside of the “Manipulate” function from Wolfram Mathematica.

**PCA of Genetic Code Cubes**

**Analysis** Principal Components | Mutation Frequency | Nonlinear regression

**Protein** ATM/BRCA1-2/P53

**Base weights control**

| | first base | second base | third base |
|---|---|---|---|
| Weight 1 | 0.734 | 1. | 0.093 |
| Weight 2 | 0.771 | 1. | 0.349 |

**Optimal weights?** ☑

**Set of Cubes** $YR_D^{CAGU}$

**Space, $Z^3$ (1) or $R^3$ (2)** 1

**Min codon norm** ☑
**Max codon norm** ☑
**Median codon norm** ☑
**Mean codon norm** ☑

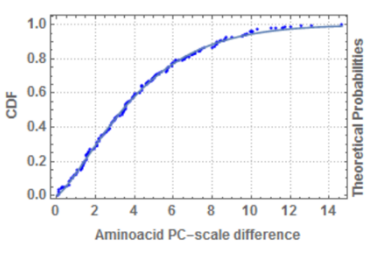
**Histogram binning** Automatic
**Point size** 0.01
**Label font size** 12
**Text font size** 12
**Point labels** ☐
**PC variance %** ☐
**Centroide size** 0.05
**Plane opacity**
**Centroide opacity**
**Image size** 500
**Two base weigths** ☑

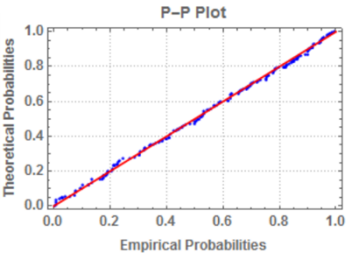
**Box Ratios**

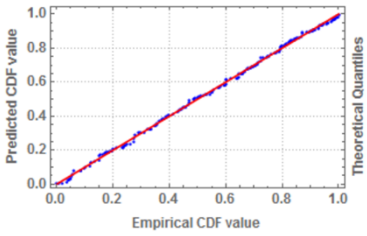
| | | |
|---|---|---|
| 1. | 1. | 1. |

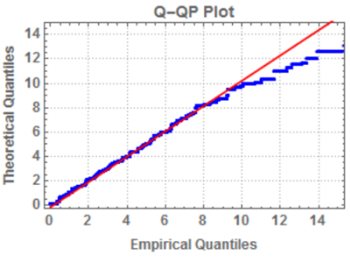
**Monte Carlo KS-test** ☑
**Distribution test** Kuiper
**No. bootstrap samplings** 1000
**Non-parametric Bootstrap?** ☐
**Bootstrap sample size** 200

Export Tables




| | Estimate | Standard Error | t | P[Abs[T] > Abs[t]] | Goodness-Of-Fit measures | |
|---|---|---|---|---|---|---|
| $\alpha$ | 1.31772 | 0.00568461 | 231.805 | $5.62738 \times 10^{-233}$ | AdjustedRSquared | 0.999739 |
| $\beta$ | 4.45567 | 0.00918486 | 485.11 | $3.6703 \times 10^{-293}$ | AIC | −1198.1 |

| Distribution | Test statistic | Bootstrap samplings | MC sample size | P Test-pvalue |
|---|---|---|---|---|
| Weibull | 0.0550939 | 1000. | 200. | 0.95005 |

| PC amino acid scale | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A | C | D | E | F | G | H | I | K | L |
| 4.35442 | -4.28501 | 0.469857 | 0.346848 | -7.66 | -3.0138 | 4.04616 | -2.76123 | 1.85978 | -1.71622 |
| M | N | P | Q | R | S | T | V | W | Y |
| -3.06445 | 1.97488 | 6.99639 | 3.93952 | 1.00792 | 2.64649 | 5.41064 | -5.59176 | -4.62092 | -0.339536 |