

Amino Acid Properties

Elements of Biophysics

Emidio Capriotti

<http://biofold.org/>



Biomolecules
Folding and
Disease

Department of Pharmacy and
Biotechnology (FaBiT)
University of Bologna



Amino Acid Properties

Physico-chemical and biochemical **properties of amino acids** are defined by **indexes or propensity scales**.

This properties can be used to perform **simple protein structure predictions** by associating each residue to different statistically evaluated features.

AAIndex Database

AAindex is a database of numerical indices representing various physicochemical and biochemical properties of amino acids and pairs of amino acids.



AAindex

Amino acid indices, substitution matrices and pair-wise contact potentials

AAindex is a database of numerical indices representing various physicochemical and biochemical properties of amino acids and pairs of amino acids. AAindex consists of three sections now: AAindex1 for the amino acid index of 20 numerical values, AAindex2 for the amino acid mutation matrix and AAindex3 for the statistical protein contact potentials. All data are derived from published literature.

Search or Download

Search by DBGET bfind for

<https://www.genome.jp/aaindex/>

Amino Acid Volume

Some scales define the volume of the amino acids

H GRAR740103

D Volume (Grantham, 1974)

R PMID:4843792

A Grantham, R.

T Amino acid difference formula to help explain protein evolution

J Science 185, 862-864 (1974)

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|------|------|------|------|-----|-----|------|------|------|
| | 31. | 124. | 56. | 54. | 55. | 85. | 83. | 3. | 96. | 111. |
| | 111. | 119. | 105. | 132. | 32.5 | 32. | 61. | 170. | 136. | 84. |

//

Amino Acid Surface

Some scales define the surface of the amino acids

H JANJ780101

D Average accessible surface area (Janin et al., 1978)

R PMID:731698

A Janin, J., Wodak, S., Levitt, M. and Maigret, B.

T Conformation of amino acid side-chains in proteins

J J. Mol. Biol. 125, 357-386 (1978)

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|-------|------|------|------|------|------|------|------|------|
| | 27.8 | 94.7 | 60.1 | 60.6 | 15.5 | 68.7 | 68.2 | 24.5 | 50.7 | 22.8 |
| | 27.6 | 103.0 | 33.5 | 25.5 | 51.5 | 42.0 | 45.0 | 34.7 | 55.2 | 23.7 |

//

H CHOC760101

D Residue accessible surface area in tripeptide (Chothia, 1976)

R PMID:994183

A Chothia, C.

T The nature of the accessible and buried surfaces in proteins

J J. Mol. Biol. 105, 1-14 (1976)

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|------|------|------|------|------|------|------|------|------|
| | 115. | 225. | 160. | 150. | 135. | 180. | 190. | 75. | 195. | 175. |
| | 170. | 200. | 185. | 210. | 145. | 115. | 140. | 255. | 230. | 155. |

//

Amino Acid Energies

Some scales provides an estimation of the energies of transfer or contact

H JANJ790102
D Transfer free energy (Janin, 1979)
R PMID:[763335](#)
A Janin, J.
T Surface and inside volumes in globular proteins
J Nature 277, 491-492 (1979)

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|-----|------|------|------|------|------|------|-----|------|-----|
| | 0.3 | -1.4 | -0.5 | -0.6 | 0.9 | -0.7 | -0.7 | 0.3 | -0.1 | 0.7 |
| | 0.5 | -1.8 | 0.4 | 0.5 | -0.3 | -0.1 | -0.2 | 0.3 | -0.4 | 0.6 |

//

H MIYS850101
D Effective partition energy (Miyazawa-Jernigan, 1985)
R
A Miyazawa, S. and Jernigan, R.L.
T Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation
J Macromolecules 18, 534-552 (1985)

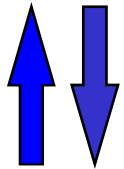
| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|------|------|------|------|------|------|------|------|------|
| | 2.36 | 1.92 | 1.70 | 1.67 | 3.36 | 1.75 | 1.74 | 2.06 | 2.41 | 4.17 |
| | 3.93 | 1.23 | 4.22 | 4.37 | 1.89 | 1.81 | 2.04 | 3.82 | 2.91 | 3.49 |

//

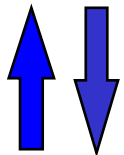
Secondary Structure

Covalent structure

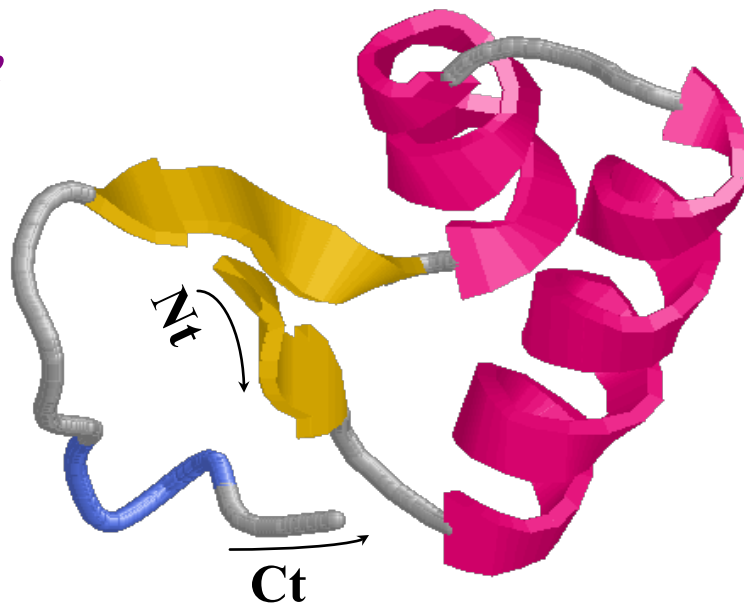
TTCCPSIVARSNFNVCRLPGTPEAICATYTGCIIPGATCPGDYAN



Secondary structure



3D structure



Chou-Fasman (I)

Given a set of known structures we can count how many times a residue is associated to a structure.

Example:

ALAKSLAKPSDTLAKSDFREKWEWLKLLKALACCKLSAAL
hhhhhhhhccccccccccccchhhhhhhhhhhhhhhhh

$$N(A,h) = 7, N(A,c) = 1, N = 40$$

$$P(A,h) = 7/40, P(A,c) = 1/40$$

Is that enough for estimating a propensity?

Chou-Fasman (II)

We need to estimate how much independent the residue-to-structure association is.

$$P(h) = 27/40, P(c) = 13/40, P(A) = 8/40$$

If the structure is independent of the residue:

$$P(A,h) = P(A) \times P(h)$$

The propensity is:

$$\frac{P(A,h)}{P(A) \times P(h)}$$

The prediction method

The Chou-Fasman method was published in 1974 and the propensity scales were calculated on a set of 19 proteins.

| Helical Residues ^b | P_α | β -Sheet Residues ^c | P_β |
|-------------------------------|------------|--------------------------------------|-----------|
| Glu ⁽⁻⁾ | 1.53 | Met | 1.67 |
| Ala | 1.45 | Val | 1.65 |
| Leu | 1.34 | Ile | 1.60 |
| His ⁽⁺⁾ | 1.24 | Cys | 1.30 |
| Met | 1.20 | Tyr | 1.29 |
| Gln | 1.17 | Phe | 1.28 |
| Trp | 1.14 | Gln | 1.23 |
| Val | 1.14 | Leu | 1.22 |
| Phe | 1.12 | Thr | 1.20 |
| Lys ⁽⁺⁾ | 1.07 | Trp | 1.19 |
| Ile | 1.00 | Ala | 0.97 |
| Asp ⁽⁻⁾ | 0.98 | Arg ⁽⁺⁾ | 0.90 |
| Thr | 0.82 | Gly | 0.81 |
| Ser | 0.79 | Asp ⁽⁻⁾ | 0.80 |
| Arg ⁽⁺⁾ | 0.79 | Lys ⁽⁺⁾ | 0.74 |
| Cys | 0.77 | Ser | 0.72 |
| Asn | 0.73 | His ⁽⁺⁾ | 0.71 |
| Tyr | 0.61 | Asn | 0.65 |
| Pro | 0.59 | Pro | 0.62 |
| Gly | 0.53 | Glu ⁽⁻⁾ | 0.26 |

^a Chou and Fasman (1974). ^b Helical assignments: H_α , strong α former; h_α , α former; I_α , weak α former; i_α , α indifferent; b_α , α breaker; B_α , strong α breaker. I_α assignments are also given to Pro and Asp (near the N-terminal helix) as well as Arg (near the C-terminal helix). ^c β -sheet assignments: H_β , strong β former; h_β , β former; I_β , weak β former; i_β , β indifferent; b_β , β breaker; B_β , strong β breaker. b_β assignment is also given to Trp (near the C-terminal β region).

Updated Chou-Fasman

An update version of the Chou-Fasman propensity scales are available at the AAIndex database.

```
H CHOP780201
D Normalized frequency of alpha-helix (Chou-Fasman, 1978b)
R PMID:364941
A Chou, P.Y. and Fasman, G.D.
T Prediction of the secondary structure of proteins from their amino acid
  sequence
J Adv. Enzymol. 47, 45-148 (1978)
```

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|------|------|------|------|------|------|------|------|------|
| | 1.42 | 0.98 | 0.67 | 1.01 | 0.70 | 1.11 | 1.51 | 0.57 | 1.00 | 1.08 |
| | 1.21 | 1.16 | 1.45 | 1.13 | 0.57 | 0.77 | 0.83 | 1.08 | 0.69 | 1.06 |

//

```
H CHOP780202
D Normalized frequency of beta-sheet (Chou-Fasman, 1978b)
R PMID:364941
A Chou, P.Y. and Fasman, G.D.
T Prediction of the secondary structure of proteins from their amino acid
  sequence
J Adv. Enzymol. 47, 45-148 (1978)
```

| I | A/L | R/K | N/M | D/F | C/P | Q/S | E/T | G/W | H/Y | I/V |
|---|------|------|------|------|------|------|------|------|------|------|
| | 0.83 | 0.93 | 0.89 | 0.54 | 1.19 | 1.10 | 0.37 | 0.75 | 0.87 | 1.60 |
| | 1.30 | 0.74 | 1.05 | 1.38 | 0.55 | 0.75 | 1.19 | 1.37 | 1.47 | 1.70 |

//

Trans Membrane Regions

Predicting the position of Trans Membrane Segments along the sequence

ALALMLCMLTYRHKELKCLKLKK ALALMLCMLTYRHKELKCLKLKK ALALMLCMLTYRHKELKCLKLKK

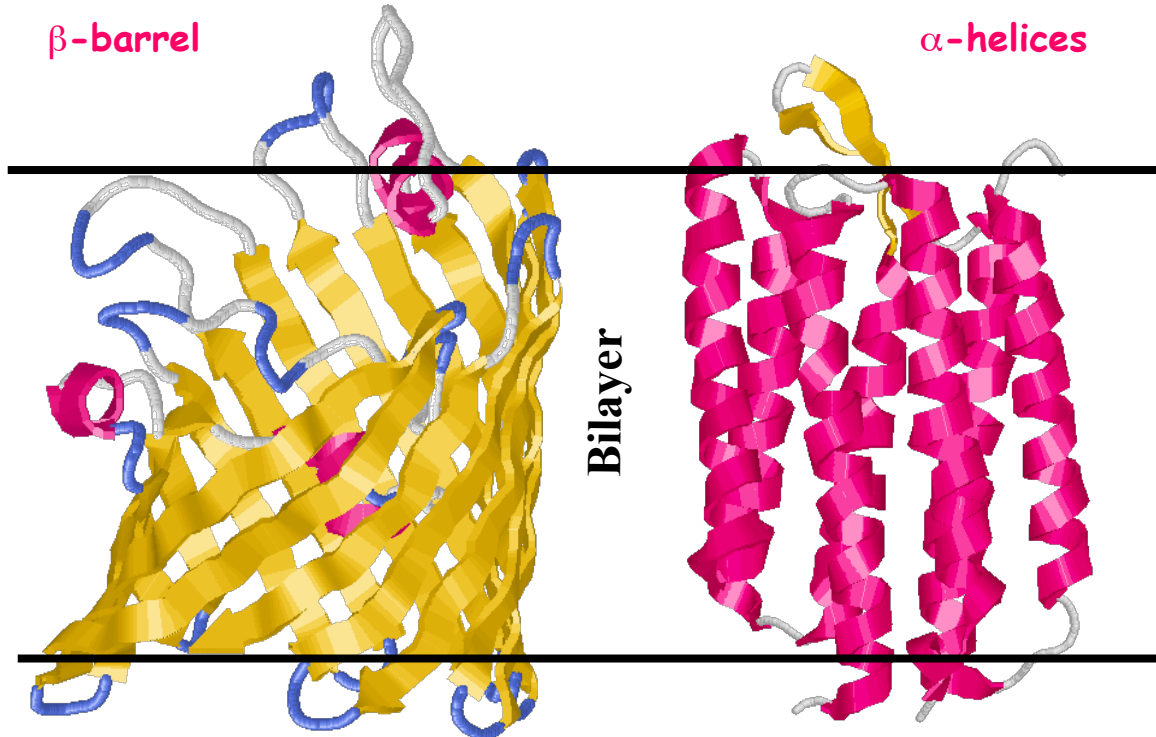


Outer Membrane

Inner Membrane

β -barrel

α -helices

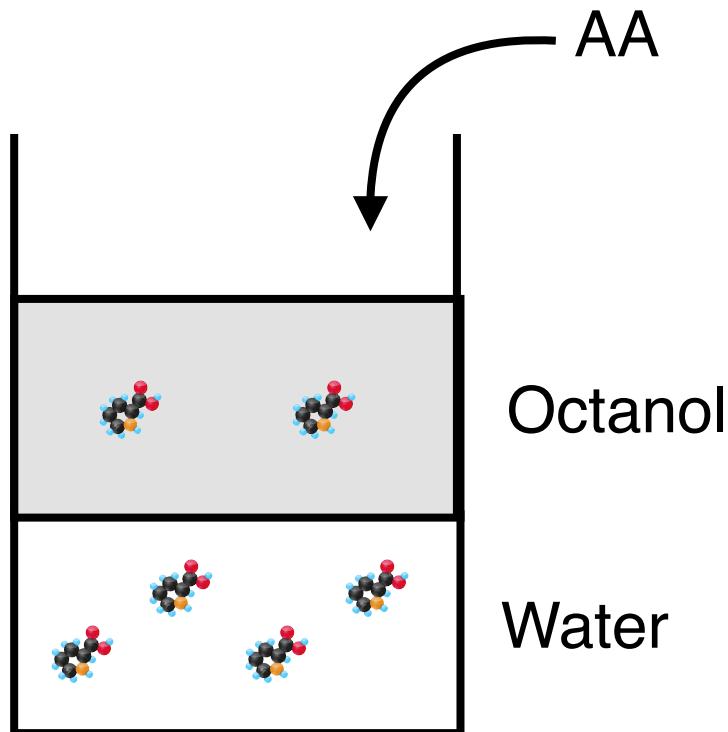


Porin
(*Rhodobacter capsulatus*)

Bacteriorhodopsin
(*Halobacterium salinarum*)

Partition coefficient

The partition coefficient (P) is the ratio of concentrations of a compound in a mixture of two immiscible solvents at equilibrium.



$$P = \frac{[AA]_{\text{Octanol}}}{[AA]_{\text{water}}}$$


Kyte-Doolittle scale

It is computed taking into consideration the octanol-water partition coefficient, combined with the propensity of the residues to be found in known transmembrane helices

```
H KYTJ820101
D Hydropathy index (Kyte-Doolittle, 1982)
R PMID:7108955
A Kyte, J. and Doolittle, R.F.
T A simple method for displaying the hydrophobic character of a protein
J J. Mol. Biol. 157, 105-132 (1982)
I  A/L      R/K      N/M      D/F      C/P      Q/S      E/T      G/W      H/Y      I/V
   1.8      -4.5     -3.5     -3.5     2.5      -3.5     -3.5     -0.4     -3.2     4.5
   3.8      -3.9     1.9      2.8      -1.6     -0.8     -0.7     -0.9     -1.3     4.2
//
```

ProtScale at ExPASy

ExPASy webserver plots protein plots based on different scales

ProtScaleHome | [Contact](#)

ProtScale

ProtScale [[Reference](#) / [Documentation](#)] allows you to compute and represent the profile produced by any amino acid scale on a selected protein.

An **amino acid scale** is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales entered from the literature.

Enter a **UniProtKB/Swiss-Prot** or **UniProtKB/TrEMBL** accession number (AC) (e.g. **P05130**) or a sequence identifier (ID) (e.g. **KPC1_DROME**):

Or you can paste your own sequence in the box below:

```
AGFGHIKLMNPRRFTKWTGGFGRNDEALLALAVRAIALK
PRA
```

Please choose an amino acid scale from the following list. To display information about a scale (author, reference, amino acid scale values) you can click on its name.

| | |
|--|---|
| <input type="radio"/> Molecular weight | <input type="radio"/> Number of codon(s) |
| <input type="radio"/> Bulkiness | <input type="radio"/> Polarity / Zimmerman |
| <input type="radio"/> Polarity / Grantham | <input type="radio"/> Refractivity |
| <input type="radio"/> Recognition factors | <input type="radio"/> Hphob. / Eisenberg et al. |
| <input type="radio"/> Hphob. OMH / Sweet et al. | <input type="radio"/> Hphob. / Hopp & Woods |
| <input checked="" type="radio"/> Hphob. / Kyte & Doolittle | <input type="radio"/> Hphob. / Manavalan et al. |
| <input type="radio"/> Hphob. / Abraham & Leo | <input type="radio"/> Hphob. / Black |
| <input type="radio"/> Hphob. / Bull & Breese | <input type="radio"/> Hphob. / Fauchere et al. |
| <input type="radio"/> Hphob. / Guy | <input type="radio"/> Hphob. / Janin |
| <input type="radio"/> Hphob. / Miyazawa et al. | <input type="radio"/> Hphob. / Rao & Argos |
| <input type="radio"/> Hphob. / Roseman | <input type="radio"/> Hphob. / Tanford |
| <input type="radio"/> Hphob. / Wolfenden et al. | <input type="radio"/> Hphob. / Welling & al |
| <input type="radio"/> Hphob. HPLC / Wilson & al | <input type="radio"/> Hphob. HPLC / Parker & al |

<https://web.expasy.org/protscale/>

Exercise 1

Extract data from the average accessible surface (JANJ780101), the effective partition energy (MIYS850101) and the hydrophobicity (KYTJ820101) scales.

Calculate the correlation coefficient for all the possible pairs.

Which pair of scales are the most correlated ones?

Suggestion: Parse the text files and import the data in excel.

Exercise 2

Consider the Bacterial Rhodopsin structure 1FBB extract the chain A and calculate the secondary structure using [dssp-web](#).

Identify the regions corresponding to the seven alpha helices.

Use the scale of Kyte-Doolittle and Chou-Fasman to verify that helical regions correspond to high level of helical propensity and hydrophobicity.

Suggestion: In the dssp file generated from the PDB ,the residue type and the secondary structure are reported in columns 14 and 17 respectively. Replace blank secondary structure with C (coil). A single residue in turn conformation (T) in the middle of long stretch of helical residues should be considered as helical.