CABIOS APPLICATIONS NOTE

Vol. 12 no. 3 1996
Pages 251–252

PM — Protein music

Ross D.King* and Colin G.Angus[1]

We present the program PM for the analysis of protein sequence information using audification. Audification is the technique of using the sense of hearing to analyse data (Jackson, 1993). The advantage of audification over visualisation in data analysis is that sound has the property that when different notes are played together they can still be individually heard: in vision colours blend to form new colours. This distinction can be very useful when studying multivariate data. The one dimensional structure of DNA maps naturally onto musical sequences (Hayashi and Munakata, 1984; Ohno and Ohno, 1986). The mapping algorithm in PM is designed to translate coding sequences into sound in a way that is more harmonious and more open to analysis than previous methods.

The PM algorithm uses the DNA nucleotide sequence to form the musical top line, and the properties of the translated amino-acids to form the bass line. The notes are in the scale of C Major and there are three beats to the bar. Each codon corresponds to a bar of music and the notes are played on the beat in order of nucleotide sequence. The mapping of the top line to musical notes is as follows:

a → A4 submediant
t → E3 mediant
c → C3 tonic
g → G3 dominant

The mapping of the bass line to notes is more complicated and is based on the designation of amino-acid physico-chemical properties of Taylor (1983). Each of the properties has a note assigned to it. The mapping is as follows:

polar	(s, n, t, d, e, y, w, h, k, r, q)	→ A2 mediant
hydrophobic	(a, g, c, v, t, i, l, m, y, f, w, h, k)	→ C1 tonic
charged	(d, e, h, k, r)	→ F1 subdominant
positive	(h, k, r)	→ E1 mediant
aliphatic	(i, v, l)	→ G1 dominant
aromatic	(f, y, w, h)	→ D1 supertonic
tiny	(a, g, c, s)	→ all +1 octave

Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, PO Box 123, London, WC2A 3PX, UK and [1]The Shamen, Moksha Management, PO Box 102, London, E15 2HH, UK

**To whom correspondence should be addressed*

The number of notes in the bass line varies with the particular amino-acid because each amino acid has a characteristic combination of properties. The following timing and ordering of notes for each residue was found to allow properties and residues to be identified as well as sounding harmonious (X+ = a double beat):

a → C+_+_+
c → C+_+_+
d → A+_+F+
e → A+_+C+
f → D+_+C+
g → C+_+_+
h → E+ACDF
i → C+_+G+
k → A+C+FE
l → C+_+G+
m → C+_+_+
n → A+_+_+
p → _+_+_+
q → A+_+_+
r → A+E+F+
s → A+_+_+
t → A+_+C+
v → C+_+G+
w → D+C+A+
y → D+C+A+

Note: C is considered polar, and P has no bass line (emphasising its unique properties).

For each residue position it is possible to hear the different residue properties. To do this equivalently using vision would be cumbersome as each position would have to be mapped into seven colours.

PM inputs standard GenBank files which are translated into music via the MIDI standard. The PM program was developed for Macintosh computers and the program was written in C. Communication with the MIDI interface was implemented using MIDIPACAL from Altech Systems (Altech, Systems, 122 Faries Indl Pk Dr., Shreveport, LA 71106, USA, (1)-318-868-8036). The PM program is polyphonic and should work with any MIDI instrument. To use PM and listen to sequences the hardware requires are: a Macintosh computer, a MIDI interface, and a MIDI instrument. The program

simultaneously displays the DNA, the translated residue sequence, and a simplified musical score, as it plays the music.

The potential of applying audification data analysis methods to biological data is almost completely unexplored and it is possible that for some niche applications audification will have advantages over other methods.

We have applied PM to analysing the human alpha globin gene cluster on chromosome 16 (zeta gene, with 3 exons), and the G-protein coupled receptor protein serotinin receptor type 2 (5HT2). A musical track based on the serotonin receptor has sold over 100 000 copies world wide.

The PM program is available at the EBI software depositary at: http://www.ebi.ac.uk

References

Hayashi,K. and Munakata,N. (1984) Basically musical *Nature*, **310**, 96.

Ohno,S. and Ohno,M. (1986) The all persuasive principle of repititious recurrences governs not only coding sequence construction but also human endeavour in musical composition. *Immunogenetics*, **24**, 71–78.

Jackson,J. (1993) Proceedings of the First International Conference on Auditory Display, Santa Fe Institute Studies in the Science of Complexity, Proceedings Series, Addison Wesley.

Taylor,W.R. (1983) The classification of amino-acid conservation. *J. Theor. Biol.*, **119**, 205–221.

Received on October 26, 1995; accepted on February 8, 1996