

# DNA code breaker tested theory on Jane Austen text

November 17 2006

---

A researcher at the University of Bradford has perfected a computer programme that could unlock the secrets of the human genome and pave the way towards new treatments and drugs sooner than had been expected.

As reported in this week's edition of the journal *Nature* Professor Simon Shepherd has constructed an algorithm that can unpick the sequences of As, Gs, Cs and Ts that make up the world's genomes.

Professor Shepherd, who is Professor of Computational Mathematics at the University of Bradford, has been working on genomics with Professor Clive Beggs (Professor of Medical Technology) and Dr Sue Jones (Lecturer in Biomedical Science) in Bradford's Medical Biophysics Research Group.

Professor Shepherd originally tested his computer programme on the entire text of *Emma* by Jane Austen after removing all the spaces and punctuation, leaving just a long impenetrable line of letters. Despite having no knowledge of the English vocabulary or syntax, the programme managed to identify 80 per cent of the words and separate them back into sentences.

Professor Shepherd believes that this can be applied to the genetic sequence, which contains around 3 billion letters and is currently baffling scientists as to how to interpret it. Within these sequence there is information that nobody knows how to extract – codes that regulate,

control or describe all kinds of cellular processes.

Professor Shepherd believes that his method of number crunching will be able to make an interpretation. He said: “We are treating DNA as we used to treat problems in intelligence. We want to break the code at the most fundamental level.”

A human cell has to fit about two metres of DNA into a nucleus a few micrometres in diameter, which requires packing it together with proteins in a complex hierarchy of ‘folding back and wrapping around’. The fundamental element underlying all this packaging is the nucleosome – 147 base pairs of DNA wrapped around a globule of eight proteins called ‘histones’.

Professor Shepherd added: “The protein folding problem is regarded as one of the three grand challenge problems of 21st century science. Its resolution is crucial to the development of the new drugs and medical therapies that the Human Genome project promises one day to deliver.

“I believe that the combination of insights from the hard, numerate sciences such as mathematics and engineering, coupled with expert knowledge of the biochemistry at the cellular level, will prove to be the most fruitful approach.

“Although results will not happen overnight, we can expect to see the promise of the Human Genome project bearing fruit within the next 20 to 50 years.”

Citation: *Nature*, pg 259, Vol. 444, 16 November 2006

Source: University of Bradford

Citation: DNA code breaker tested theory on Jane Austen text (2006, November 17) retrieved 4 May 2024 from <https://medicalxpress.com/news/2006-11-dna-code-breaker-theory-jane.html>

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.