

# Dr. Grant Jacobs

## Senior bioinformatics scientist / consultant

Email [gjacobs@bioinfotools.com](mailto:gjacobs@bioinfotools.com)  
Mail BioinfoTools, PO Box 6129, Dunedin, New Zealand  
WWW <http://www.bioinfotools.com>  
Telephone +64 3 476 1820 (New Zealand, after 10am)

### Skills

- Bioinformatics scientist with over ten years experience.
- Excellent communication and planning skills.
- Actively follows developments in software development, bioinformatics and gene regulation.
- Knowledge of higher eukaryote gene regulation and chromatin structure literature from atomic interactions to organelle (nucleus).
- Molecular structure analysis, especially proteins and protein–DNA interactions.
- Molecular modelling.
- Sequence analysis, especially protein sequences and protein binding sites in DNA.
- Experienced programmer. Have used more than 10 programming languages since 1982.
- Algorithm development, including those based on suffix trees and computational geometry.

### Current projects and research interests

- The structural bioinformatics of higher eukaryote gene regulation. This major project binds together several aspects including the nature of protein–DNA interactions, available atomic structures of protein–DNA complexes and the role of chromatin in higher eukaryote gene regulation.
- Extracting evolutionary, protein–substrate, protein–protein interaction and protein structure information from alignments of protein sequences.
- High–speed data stores for biological sequence and structure data.
- Protein sequence–structure matching.
- Other work–related interests include computer language design, cognitive neuroscience, science management and disability–related issues.

## Previous experience

*Programming* (see separate section below for examples of software written):

- A simple operating system for a 4-CPU overlapping RAM system – programmed in C, cross-compiled, burned into an EPROM and debugged with an oscilloscope.
- Parsing Genbank and EMBL flat-file databases.
- Development of interactive web sites.
- Suffix trees for sequence analysis.
- New sorting algorithm, novel variant on heapsort.
- Implemented own database to analyse multiple sequence alignments from ground up. Not a relational database-based product. Supports many operations.
- Experienced in a wide range of computer languages including C, Perl, Pascal, HTML, CSS, Unix shell scripting.

*Bioinformatics science:*

- Studying motifs in, and structures of, DNA-binding proteins.
- First method to identify functional residues from multiple sequence alignment of functional divergent proteins. Applied to correctly predict the substrate binding residues of CCHH zinc fingers from protein sequence data.
- Molecular dynamic simulations, molecular modelling and sequence analysis of the “leucine zipper” coiled-coil domain of the bZIP family of transcriptional activators.
- Molecular modelling and sequence analyses of CCHH zinc finger proteins, using own multiple sequence analysis software written for the purpose.
- Modelling a proposed interchelating protein-DNA interaction.
- Techniques for analysing protein sequences, including multiple alignment, sequence-structure matching, identifying active sites and so on.
- Examining the role of water molecules in protein-DNA interactions.
- Presenting a survey of international bioinformatics initiatives on behalf of the New Zealand Foresight Programme.
- Linkage analysis of vesico-ureteric reflux (VUR).
- Development and maintenance of Transterm, a database of mRNA regions and signal sequences.

□

## Previous appointments

|  |           |
|--|-----------|
| Independent scientist / consultant   | 2001–     |
| Research Fellow, Department of Biochemistry, University of Otago   | 1999–2000 |
| Post-doctoral fellow, Cancer Genetics Laboratory, University of Otago  | 1998–1999 |
| Computer consultant, including maintaining bioinformatics resources for the Christchurch School of Medicine        | 1997–1998 |
| FRST post-doctoral research fellow, Christchurch School of Medicine  | 1994–1996 |
| Short-term research fellow, MRC Laboratory of Molecular Biology, Cambridge, England                                | 1994      |
| Dept of Computer Science, University of Canterbury; tutor (terminated prematurely to take up studies at Cambridge) | 1988      |
| Computing Technology Limited; computer programmer  | 1987      |

□  
□

## Academic qualifications and awards

|   |           |
|---|-----------|
| NZ FRST Postdoctoral fellowship   | 1994–1996 |
| Max Perutz Postgraduate Student Award (UK)  | 1992      |
| Ph.D., MRC Laboratory of Molecular Biology, Cambridge, England  | 1988–1992 |
| ORS Award (UK)  | 1988–1991 |
| NZ MRC (now HRC) Postgraduate Scholarship   | 1988–1991 |
| University of Canterbury (NZ) Postgraduate Scholarship (declined due to plans to study overseas)  | 1987      |
| BSc(Hons) 1st Class, University of Canterbury, Christchurch, New Zealand.<br>Studied both biology and computer science to graduate level. | 1982–1986 |

## Programming background

### *Programming languages used previously:*

- Perl
- C
- Pascal
- FORTRAN-77
- BASIC
- SQL
- Web “programming” with HTML, CSS and Javascript
- Various assembly languages (6502, 68000, 6805, etc.)

Several other languages (e.g. Modula-2, Prolog, Lisp) have been used occasionally. Currently I am learning Java, XML, JXTA, threads and socket programming.

### *Operating systems used:*

- Apple (from Apple II through to OS X)
- DOS
- VMS
- Unix (various flavours, Solaris, SGI, etc.)
- Linux

### *Computer systems used:*

- Tandy TRS-80
- Apple II+
- IBM XT
- NEC PC
- VAX/VMS
- Alliant
- Silicon Graphics
- Cray YMP
- Sun (Sun OS & Solaris)
- Apple Macintoshes: 68000 (LCIII, LC475), PowerPC (7200, 7600, iMac, G4)

## Examples of software written

(Current projects are confidential and are not included in this list.)

- **Various** (Pre 1986) My first programs were written in BASIC on TRS-80 computers at high school (1981), then in UCSD Pascal on an Apple II+ I purchased as an undergraduate student.
- **CROWDY** (1986) Modelling density effects on plant growth for ecological studies being done at the University of Canterbury.
- **Unnamed** (1987) A simple operating system for a 4-processor overlapping-RAM computer developed at Computing Technology Limited (Christchurch, NZ). Programmed in C on a PC, cross-compiled to 6805 code, “burned” into an EPROM, installed into the multi-processor hardware and debugged using an oscilloscope.
- **NWAlign** (1990) Needleman-Wunsch alignments of sequences.
- **DOTTY** (1990) Dot plots for sequence comparisons.
- **DAWG, DAWGAlign and others** (1990-) Suffix tree methods for high-speed sequence searching and locating conserved motifs in unaligned sequences.
- **MotifAnal** (1991-) An interactive database-style system for analysing large multiple alignments of protein sequences. This program (>20,000 lines of source code) was written in Pascal on Vax computers over several years. It has many features, including:
  - Construction of databases, including annotation options;
  - Use of arbitrary motif weight, amino acid property and amino acid similarity tables;
  - Conversion of amino acids property tables to amino acid similarity tables. Standard operations on tables such as scaling, normalisation, symmetrising;
  - User-specified position referencing schemes. Allows users to refer to positions in an alignment independent of the actual alignment position. Provides a referencing scheme that withstands later revision of alignments such as later inclusion of longer loop sequences.
  - Related positions that are not sequentially adjacent in the alignment (eg. active site residues) can be referred to in convenient manner;
  - Calculation and plots of the conservation of each position in an alignment, output in plain text or Postscript. Conservation tables can be compared to determine co-conservation and the like;
  - Complex comparisons can be generated using a range of user-specified options, filtering “masks” and commands. Subsets of the alignment (certain sequences, certain positions within those sequences) can be passed to each of the analysis options, based on criteria such as named sequences and positions, the amino acids present at named positions, the score of the selected positions against a mask sequence, optional use of position weight tables and so on;
  - Statistics of properties can be output;
  - Correlations between alignment positions can be calculated and output;
  - The number of amino acids between positions can be used as a property;
  - Duplicate entries can be eliminated if desired;
  - Phylogenies of motifs can be constructed, including “group phylogenies”; an early pre-descendent of the evolutionary trace methods that are now widely available
- **WHEEL (1991)** Depiction of protein sequence (family) conservation on “helical wheels” with PostScript™ output.
- **SITEPRED** (1991-1992) Prediction of active site residues from large multiple sequence alignments of protein families with divergent functions. Successful application of

this approach to CCHH zinc fingers resulted in a single author article in EMBO Journal (Jacobs, 1992). Predates methods subsequently published to predict functional residues from alignments.

- **MACC** (1991–1992) Plotting conservation and co-conservation of multiple sequence alignments.
- **MALIGN** (1991–1992) Exploration of a “two dimensional” approach to multiple sequence alignment using a graph theoretic approach.
- **WatCons** (1995) Fast identification of conserved atoms in atomic structures. Used to determine water molecules conserved in related protein–DNA complexes.
- **Various** (1998–1999) Macintosh software to assist linkage analysis, including :
  - **Make .dat files from DBs**
  - **Make .pre files from DBs**
  - **Make marker map files**
  - **Extract marker lists**
- **Various** (1999–2000) Over 50 programs used in the development and maintenance of the Transterm database, including:
  - Software to check the contents of the NCBI taxonomy database and convert into the locally–used format (All 1999–): **BuildTaxaIndices**, **CheckNodesFormat**, **MakeSpeciesList**, **MakeSppTaxidList**, **MakeTaxid2Div**, **MakeTaxid2Names**, **MakeTaxid2Org**, **MakeTaxid2SSN**, **MakeTaxid2SppTaxid**, **RemoveTaxidLines**, **ShowStrains**, **StripOKTaxids**, **StudyTaxid**, **Taxid2SppTaxid**. Recently merged into a single program **BuildTaxaDB** (2002–).
  - The multi–frame, multi–form web interface for Transterm, including perl modules: **GHJ\_TTCGI** (1999–) and **GHJ\_TTPG** (1999–);
  - **ExtractFeatures** (2000–) – software to process arbitrary flat–file database files, including Genbank, SWISSPROT and the like;
  - Programs to automatically download large sets of files from ftp sites (used to update local copies of the genome database files and the NCBI taxonomy database), eg (all 1999–): **GetGenomes**, **GetTaxaDB**, **PrepGetGenomes**, **UnpackGenomes**, **UpdateGBKGenome**, **getGB**, **getGenomes**, **getTaxdump**;
  - Perl module to process Unix command line & shell options: **GHJ\_UnixShell** (1999–);
  - Many other programs used in the construction of the Transterm database, written in both C and Perl, totalling over 12,000 lines of code, eg (All 1999–2000): **PrepTransTermBuild**, **makeGB**, **BuildGB**, **BuildIndices**, **BuildListFiles**, **BuildLocusFiles**, **CalcBit**, **CalcBit2**, **CalcChiSq**, **CountBases**, **CountCodons**, **CountData**, **CountDivisions**, **DeStrain**, **DeStrain2**, **DoPepchi**, **DoTransTerm**, **DoubleStop**, **EmptyDB**, **ExtractLocusDataOrg**, **Fasta2Fasta**, **FilterN**, **FinRptLn**, **FinalReport**, **Fish2Fasta**, **FishError**, **FixFishSeq**, **FixNc**, **FullPathList**, **GetCDS**, **GetGenomes**, **GetGenomes.ftp**, **GetLocus**, **GetOrganism**, **LineLengths**, **ListStrains**, **MakeFTPScript**, **MakeGenomeLocusTable**, **MakeGenomes2SSN**, **MakeLCDS2protID**, **MakeListFiles**, **MakeLocus2Taxid**, **MakeLocusData**, **MakeLocusErrCounts**, **MakeSSN.err**, **MakeSpeciesList**, **MakeSppTaxidList**, **MakeTransterm**, **MergeLocusDataOrg**, **MkDirs.csh**, **PatchLocusData**, **PlotBases**, **PlotChiSq**, **PrepGetGenomes**, **PrepareSpecies**, **RemoveEntry**, **RemoveTaxidLines**, **ShowStrains**, **StripOKTaxids**, **StudyTaxid**, **SummariseSeqs**, **TidyListFiles**, **TidyLocusData**, **UTRfish**, **UniqueSSN**, **WWW\_Clean**, **WWW\_Make**, **fixDIV**, **fixTAXID**, **run\_fish**, **split40.csh**, **ttofasta**, **ttofastahead**.
- **CodeDoc** (2003–) A computer language–independent source–code documentation application.

## Selected publications in refereed journals

(conference proceedings omitted)

Jacobs, G.H.

NZ BioScience 12(5):15–18 (2003).

*Bioinformatics — Computing with Biotechnology and Molecular Biology data.*

Jacobs, G.H. Stockell, P.A., Brown, C. M.

*Applied Bioinformatics, in press.*

Jacobs, G.H., Rackham, O., Stockwell, P.A., Tate, W., Brown, C.M.

Nucleic Acids Research 30(1):310–311 (2002).

*Transterm: a database of mRNAs and translational control elements*

Eccles, M.R, Jacobs, G.H.

Annals, Academy of Medicine, Singapore. Special Issue: "Complex Genetic Diseases", 2000 Vol. 29 (3):337–345 (2000, invited review).

*The genetics of primary vesico-ureteric reflux*

Jacobs, G.H., Stockwell, P., Schreiber, M., Tate, W.P. and Brown, C.M.

Nucleic Acids Research 28: 293–295 (2000).

*Transterm: a database of messenger RNA components and signals*

Brown, C., Jacobs, G.H., Schreiber, M.J., Magnum, J., McNaughton, J.C., Cambray, M., Futschik, M., Major, L.L., Rackham, O., Tate, W.P., Thompson, C. and Kasabov, N.K.

New Zealand BioScience 7(4):11–12 (1999)

*Using bioinformatics to investigate post-transcriptional control of gene expression.*

Brown, C.M., Schreiber, M., Chapman, B. and Jacobs, G.H.

Springer-Verlag. Series title: "Studies in Fuzziness and Soft Computing. Series Ed. Prof. Janusz Kacprzyk. Issue Title: Future Directions for Intelligent Systems and Information Science. Issue editor: Prof. N. Kasabov. Chapter 13. (1999).

*Information Science and Bioinformatics*

Eccles, M.R., Jacobs, G.H. et al.

Am. J. Hum. Genet. (1999, conference proceedings).

*Linkage analysis studies of primary vesicouteric reflux*

Jacobs, G.H.

EMBO J. 11(12):4507–4517 (1992).

*Determination of the base recognition position of zinc fingers from sequence analysis. (Front cover, over 100 citations.)*

Jacobs, G. Michaels, G.

The New Biologist 2(8):583–584 (1990).

*Zinc finger gene database.*

□

□

□

□

□

□

□

□

## Biography

In the final year of my B.Sc.(Hons) (1986, 1st Class) in which I studied both biology and computer science, I “discovered” bioinformatics which offered a niche where I could exploit my interests in both molecular biology and computer science.

After completing my degree I worked as a computer programmer (Computing Technology Ltd.). After hours I taught myself bioinformatics by reading the research literature at the local university library. From this I drew up my own research proposal, eventually obtaining a Ph.D. studentship in the Structural Studies section of the [MRC Laboratory of Molecular Biology](#) at Cambridge University. There I studied under Dr. Andrew McLachlan (FRS), one of the founders of bioinformatics who published his first bioinformatics paper in 1969, along with Sir Aaron Klug (FRS, OM, Nobel laureate, Chemistry, 1982) and Dr. Daniela Rhodes and others.

My Ph.D. research focused on DNA-binding proteins, in particular the bZIP and CCHH zinc finger protein families. This research included molecular dynamic simulations of proteins, sequence analysis, studies of protein-DNA complex structures, basic phylogenetics and development of a large program to analyse protein motifs written over several years. This work includes correctly predicting the DNA-binding residues of zinc fingers from sequence analysis (Jacobs, GH, EMBO J. 1992).

Since leaving Cambridge, I have continued to study protein-DNA interactions, spent a period doing genetic linkage analysis and programming maintaining the [Transterm database](#) and presenting it as an interactive website.

More recently, I have established myself as an independent scientist, setting up BioinfoTools as a vehicle to deliver my bioinformatics software and consulting services. Having reviewed where bioinformatics and computer programming is headed, a portfolio of projects has been developed from a log of research ideas which has been maintained over many years. Using this, I am now bringing the most promising projects to life.

Supported by with Amonida Zadissa and Anar Khan I co-coordinate the local Bioinformatics Club whose members are drawn from several departments of the local university and local biotechnology companies. I am a member of the local biotechnology cluster ([bioSouth](#)) and frequently contribute to national taskforces in bioinformatics and biotechnology.

- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
- 
-