



DRUG DISCOVERY

Innovations in Bioinformatics

Emerging tools for drug discovery and development

By Biophoenix

 [Table of Contents](#)

Biophoenix

Biophoenix is a biomedical consultancy based in the heart of England, UK, which was founded by Dr Sreten Bogdanovic and Dr Beata Langlands. Established with support from the UK Department of Trade and Industry (DTI) in 1994, Biophoenix provides detailed, accurate, market and technical information to support decision-making by biotechnology-based enterprises in the healthcare industry. In doing so, Biophoenix relies on its comprehensive in-house bioindustry database and a wide network of contacts. Biophoenix also audits biotechnology-based enterprises with regard to the risk of intellectual property litigation on behalf of several Lloyds of London insurers. Further information is available on the internet at www.biophoenix.com.

Dr Bogdanovic is Managing Partner in Biophoenix. He holds a PhD in endocrinology from Imperial College, UK, and has worked as product manager for immunodiagnostics with the UK subsidiary of Boehringer Mannheim (now owned by Roche), and as an independent biomedical market analyst with Cistron, UK. Dr Langlands holds a PhD in biochemistry from the University of Glasgow, UK, and has worked as an immunologist, a project leader with a new biotechnology company based at Warwick University in the UK, and as a university lecturer. Biophoenix has researched and written over 50 in-depth management/market reports.

Copyright © 2007 Business Insights Ltd

This Management Report is published by Business Insights Ltd. All rights reserved. Reproduction or redistribution of this Management Report in any form for any purpose is expressly prohibited without the prior consent of Business Insights Ltd.

The views expressed in this Management Report are those of the publisher, not of Business Insights. Business Insights Ltd accepts no liability for the accuracy or completeness of the information, advice or comment contained in this Management Report nor for any actions taken in reliance thereon.

While information, advice or comment is believed to be correct at the time of publication, no responsibility can be accepted by Business Insights Ltd for its completeness or accuracy.

Table of Contents

Innovations in Bioinformatics: Emerging tools for drug discovery and development

Executive Summary	12	
Bioinformatics in the omics era	12	
Genomics and related ‘omics	13	
Proteomics	14	
Metabolomics and systems biology	15	
Knowledge management solutions	17	
Profiles of selected companies	18	
Trends and opportunities	19	
Chapter 1	Bioinformatics in the omics era	22
What is bioinformatics?		23
End-users of bioinformatics tools and services		23
Life sciences in the omics era		24
Overview of omics		26
Genomics		26
Pharmacogenomics		27
Transcriptomics		27
Proteomics		28
Metabolomics		28
Other omics		28
Background on drug discovery and development		29
Types of drugs under development		29
Small molecules		29
Protein-based biotherapies		30
Nucleic acid-based biotherapies		31
The drug discovery and development process		31
Stages (US)		31
Role of biomarkers		33
Bioinformatics in drug discovery and development		34

Chapter 2 Genomics and related omics 38

Summary	38
Background	39
Categories of genomic analysis	40
Sequencing	40
Next-generation sequencing	42
Genotyping and gene expression	44
DNA microarrays	45
Value of genomic analysis to the drug industry	46
Bioinformatics solutions	50
Analysis of sequencing data	50
Sequence databases	52
Sequence search tools	56
Multiple sequence alignment tools	61
Focus on RNAs	63
Genome finishing and annotating	71
Microarray platforms	74
Affymetrix	75
Illumina	76
Agilent Technologies	78
Applied Biosystems	79
Genome-wide association studies	79
Other advances	81

Chapter 3 Proteomics 86

Summary	86
Background	87
Categories of proteomic analysis	88
Protein separation	89
Protein identification	90
Mass spectrometry	90
Protein microarrays	91
Structure determination	91
Value of proteomic analysis to the drug industry	92
Bioinformatics solutions	94
Analysis of sequencing data	95
Sequence databases	95
Sequence search tools	110
Multiple sequence alignment tools	111

Chapter 5 Knowledge Management Solutions 158

Summary	158
Introduction	159
Providers of high-performance computing	160
Providers of storage systems	164
Web-based solutions	165
Ontologies	166
Knowledge sharing	167
The Semantic Web	169
Software to support R&D labs	170
Abrevity	172
Accelrys	172
Agilent Technologies	172
BioWisdom	173
CambridgeSoft	173
Elsevier MDL	174
Geospiza	174
GeneLogics	174
IO Informatics	174
KOOPrime	175
LabVantage	175
MathWorks	175
NextBio	176
Oracle	176
SAS	176
Symyx	177
Teranode	177
Thermo Fisher Scientific	177
Text searching and mining	178
QUOSA	178
Linguamatics	178
Inforsense	179
Insightful	179
Nervana	180
Velocity	180
Clinical trials solutions	180
Adobe Systems	182
Infosys	182
Oracle	182
Pharsight	183

Chapter 6 Profiles of selected companies 186

Summary	186
Accelrys Inc	187
Company overview	187
Bioinformatics tools and services	188
Bioinformatics-related collaborations	189
Affymetrix Inc	189
Company overview	189
Bioinformatics tools and services	191
Agilent Technologies Inc	191
Company overview	192
Bioinformatics tools and services	193
Avalon Pharmaceuticals Inc	194
Company overview	194
Bioinformatics-related collaborations	195
BG Medicine Inc	196
Company overview	196
Bioinformatics tools and services	197
BioWisdom Ltd	198
Company overview	198
Bioinformatics tools and services	198
Caliper Life Sciences Inc	200
Company overview	200
Bioinformatics tools and services	201
Compugen Ltd	202
Company overview	202
Bioinformatics tools and services	203
Bioinformatics-related collaborations	204
CuraGen Corporation	205
Company overview	205
Bioinformatics tools and services	206
Entelos Inc	207
Company overview	207
Bioinformatics tools and services	208
Genstruct Inc	209
Company overview	209
Bioinformatics tools and services	210
Bioinformatics-related collaborations	211
Gene Network Sciences Inc	212
Company overview	212
Bioinformatics tools and services	212
Bioinformatics-related collaborations	213
GeneGo Inc	213
Company overview	214

Bioinformatics tools and services	214
Bioinformatics-related collaborations	215
Health Discovery Corporation	216
Company overview	216
Bioinformatics tools and services	217
Illumina Inc	218
Company overview	219
Bioinformatics tools and services	220
Bioinformatics-related collaborations	220
Ingenuity Systems Inc	221
Company overview	221
Bioinformatics tools and services	222
Bioinformatics-related collaborations	223
Insightful Corp	224
Company overview	224
Bioinformatics tools and services	225
Bioinformatics-related collaborations	226
NextBio	226
Company overview	226
Bioinformatics tools and services	227
Bioinformatics-related collaborations	228
SAS Institute Inc	228
Company overview	228
Bioinformatics tools and services	229
Bioinformatics-related collaborations	229
Vermillion Inc	230
Company overview	230
Bioinformatics tools and services	231

Chapter 7 Trends and opportunities 234

Summary	234
End-user needs and sentiment	235
Emerging trends in market evolution	237
Bioinformatics-enabled biomarker discovery	239
The promise of Semantic Web technology	240
Facilitating implementation of IT systems	242
Market snapshot and forecasts	243
Genomics and related omics	243
Proteomics	244
Systems biology	245
Life science enterprise knowledge management	245
Market estimates	246

Research methodology	249
Index	250

List of Figures

Figure 2.1:	An entry from Entrez GENE, the US NCBI's web-based interface to GENBANK	55
Figure 2.2:	Genomic context of the myostatin gene (GDF8) using the NCBI Map Viewer	56
Figure 2.3:	Search page of microInspector	67
Figure 2.4:	Search page of the miRNA database (miRBase)	68
Figure 2.5:	miRBase Targets database, a new resource for predicting miRNA targets in animals	69
Figure 2.6:	Predicted Human Genomic Targets for the hsa-let-7g* miRNA	69
Figure 2.7:	Dharmacon siRNA Designer Search Page	70
Figure 3.8:	Advanced Search page of the Protein Data Bank (PDB), the central repository for 3D protein structure information	99
Figure 3.9:	PDB Entry for the complex of the acetylcholine receptor with carbamylcholine (1UV6)	99
Figure 3.10:	Results of a search in the UniProt KnowledgeBase (Swiss-PROT and TrEMBL) for MAP kinase phosphatases	101
Figure 3.11:	Partial UniProt record for Dual Specificity Protein Phosphatase 4 (DUSP4) / MAP Kinase Phosphatase 2	103
Figure 3.12:	UniProt Feature Aligner for Dual Specificity Protein Phosphatase 4 (DUSP4)	104
Figure 3.13:	UniProt sequences at least 90% similar to Dual Specificity Protein Phosphatase 4 (DUSP4)	105
Figure 3.14:	Front page of the Protein Kinase Resource, University of California at San Diego	106
Figure 3.15:	UniProt sequences at least 90% similar to Dual Specificity Protein Phosphatase 4 (DUSP4)	107
Figure 4.16:	Genomatica's SimPheny™ Systems Biology Model Development Process	153

List of Tables

Table 2.1:	Bioinformatic Analyses of DNA/RNA Sequences	50
Table 2.2:	Contribution of Bioinformatics to Genomics-Based Drug Discovery	50
Table 2.3:	Databases and On-line Tools for Analyzing DNA Sequences and Signals	53
Table 2.4:	Leading Sequence Comparison Servers	57
Table 2.5:	BLAST Applications and Flavors	58
Table 2.6:	Online Pairwise Alignment Programs	60
Table 2.7:	Multiple Sequence Alignment Tools	62
Table 2.8:	RNA Secondary Structure Prediction	64
Table 2.9:	miRNA and siRNA Resources	65

Table 2.10:	Selected genome-sequencing packages	72
Table 2.11:	Phylogeny and Orthology	73
Table 2.12:	Summary of patent-related genes in the major organisms	74
Table 3.13:	Role of Bioinformatics in Protein Diagnostics and Therapeutics	94
Table 3.14:	Selected Protein Databases	96
Table 3.15:	On-line tools to test for protein transmembrane segments	108
Table 3.16:	Principal Protein Domain Recognition Resources	109
Table 3.17:	Protein Structure Prediction	110
Table 4.18:	Software for Systems Biology	138
Table 7.19:	World Bioinformatics Market, 2006-2011	248