

BIOINFORMATIKA: Mengawinkan Teknologi Informasi dengan Bioteknologi

Trendnya di Dunia dan Prospeknya di Indonesia¹

Dr. Arief B. Witarto, M.Eng.

Laboratorium Rekayasa Protein, Pusat Penelitian Bioteknologi – LIPI
Jl. Raya Bogor Km.46, Cibinong 16911, Kabupaten Bogor
E-mail, witarto@biotek-indonesia.net

ABSTRAK

Ledakan informasi dari kemajuan bioteknologi seperti data sekuen DNA dari pembacaan genom, data sekuen dan struktur protein sampai kepada data transkripsi RNA berkat teknologi DNA chip, telah mendorong lahirnya Bioinformatika yang digunakan untuk mengorganisasi dan menganalisa data-data tersebut menjadi sebuah informasi biologis yang bermakna. Berbagai database telah dibuat dan banyak perangkat lunak telah diciptakan yang menunjukkan trend kepada spesialisasi tujuan. Walaupun negara berkembang kurang dapat berpartisipasi dalam eksperimen bioteknologi yang padat informasi untuk pengumpulan informasi dalam database-database itu, peluang untuk memanfaatkannya melalui Bioinformatika terbuka lebar karena sifatnya yang terbuka.

PENDAHULUAN

Tak disangkal, saat ini adalah jamannya Teknologi Informasi (TI). Berbagai produk dan jasa dalam bidang TI mulai dari komputer pribadi, Internet, handphone, dsb sudah dinikmati oleh masyarakat luas. Dengan itu semua, TI tidak hanya telah membangkitkan gelombang *new-economy* tapi juga merubah pola pikir sampai kepada gaya hidup manusia modern sehingga serasa hidup dalam "kampung dunia". Kekuatan inovasi teknologi yang disepadankan dengan TI di masa depan adalah bioteknologi [1]. Bioteknologi modern ditandai dengan kemampuan manusia untuk memanipulasi kode genetik DNA, "cetak biru kehidupan". Berbagai aplikasinya telah merambah sektor kedokteran, pangan, lingkungan, dsb [2]. Pembacaan sekuen genom manusia oleh perusahaan bioteknologi Amerika Serikat (AS) Celera Genomics dalam waktu singkat (beberapa tahun) dibanding usaha konsorsium lembaga riset publik AS, Eropa, dll (lebih dari 10 tahun) a.l. berkat kontribusi TI melalui perangkat komputasinya (perangkat keras maupun lunak). Aplikasi TI dalam bidang biologi/*life sciences* yang melahirkan bidang Bioinformatika akan menjadi semakin penting di masa depan, tidak hanya mengakselerasi kemajuan bioteknologi namun juga menjembatani dua

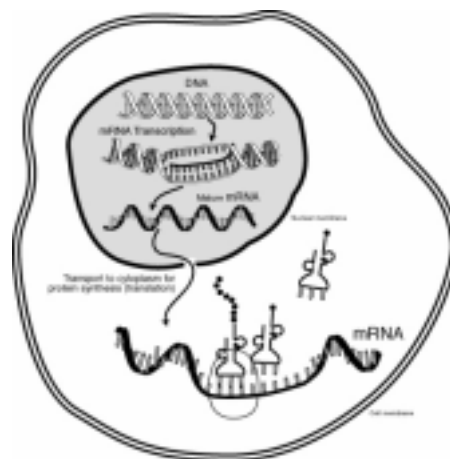
¹ Disampaikan dalam Seminar Teknologi Informasi - MIFTA, Bogor, 9 Januari 2003.

gelombang ekonomi baru tersebut (TI & bioteknologi). Dalam tulisan ini diulas perkembangan Bioinformatika di dunia dengan didahului oleh latar belakang “ledakan” informasi dalam bioteknologi, kemudian ditutup dengan prediksi prospek Bioinformatika di Indonesia melalui pertanyaan “dari mana kita harus mulai?”

BIOTEKNOLOGI MODERN

Bioteknologi modern lahir tahun 70-an diawali dengan inovasi ilmuwan AS mengembangkan teknologi DNA rekombinan. Berkat penemuan ini lahirlah perusahaan bioteknologi pertama di dunia, Genentech di AS yang segera memproduksi protein hormon, insulin yang dibutuhkan penderita diabetes, dalam bakteri. Selama ini insulin hanya bisa didapatkan dalam jumlah sangat terbatas dari organ pankreas sapi. Sebagaimana TI, saat ini produk bioteknologi telah mengimbas bahkan kepada kebutuhan hidup sehari-hari masyarakat seperti pangan, kosmetika, dsb.

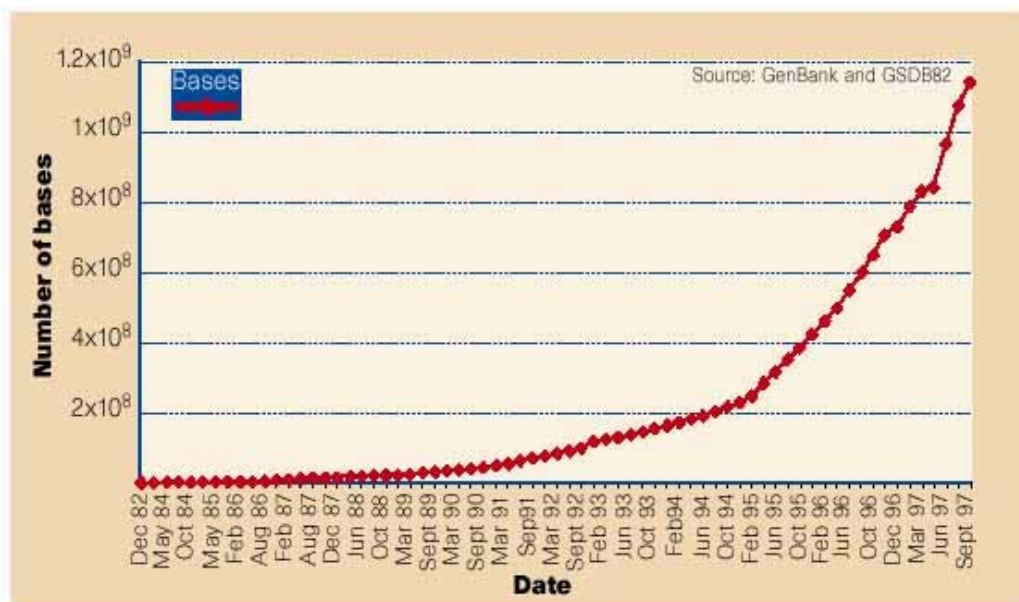
Ciri dari bioteknologi modern tadi adalah kemampuan pada manipulasi DNA. Rantai/sekuen DNA yang mengkode protein disebut gen. Gen itu ditranskripsikan menjadi mRNA yang selanjutnya mRNA ditranslasikan menjadi protein (Gambar 1). Protein sebagai produk akhir adalah yang bertugas menunjang seluruh proses kehidupan antara lain sebagai katalis reaksi biokimia dalam tubuh (protein ini disebut enzim), ikut serta dalam sistem pertahanan tubuh melawan virus, parasit dll (disebut antibodi), menyusun struktur tubuh dari ujung kaki (otot terbentuk dari protein actin, myosin, dsb) sampai ujung rambut (rambut tersusun dari protein keratin), dll. Arus informasi, DNA -> RNA -> Protein, inilah yang disebut sentral dogma dalam biologi.



Gambar 1. Sentral dogma: Arus informasi dalam sel.

Hanya 20-an tahun sejak bioteknologi modern lahir, terjadilah ledakan data biologis yang mencengangkan. Hal ini disebabkan oleh kemajuan teknologi biologi molekuler itu sendiri (misalnya DNA rekombinan, PCR, dsb) dan ditunjang dengan peralatan yang memadai membuat waktu dan biaya lebih pendek/murah. Ledakan awal dimulai dari data DNA (Gambar 2). Tahun 1977 untuk pertamakalinya sekuen DNA satu organisme dibaca secara menyeluruh yaitu pada sejenis virus yang memiliki kurang lebih 5.000 nukleotida/molekul DNA atau sekitar 11 gen. Sekarang sudah ada milyaran data nukleotida tersimpan dalam database DNA, GenBank di AS yang didirikan tahun 1982 [3]. Sekuen seluruh DNA manusia yang terdiri dari 3 milyar nukleotida dirampungkan dalam waktu 3 tahun. Di Indonesia, dengan membayar \$15, kita bisa membaca sekuen 500-an nukleotida di Lembaga Biologi Molekuler Eijkman, Jakarta. Trend yang sama juga nampak pada database lain seperti database sekuen asam amino penyusun protein, database struktur 3D protein, dsb. Inovasi teknologi

DNA chip yang dipelopori oleh perusahaan bioteknologi AS, Affymetrix di Silicon Valley telah mendorong munculnya database baru mengenai RNA. Dengan ini, riset tidak dilakukan lagi satu persatu terhadap molekul (DNA/RNA/protein) yang diminati, namun pada keseluruhan/satu set masing-masing molekul (untuk DNA dari gen ke genom, untuk RNA disebut transkriptom dan proteom untuk protein).



Gambar 2. Pertumbuhan data nukleotida/basa DNA dalam GenBank.

TREND BIOINFORMATIKA DUNIA

Ledakan data/informasi biologi itu yang mendorong lahirnya Bioinformatika. Karena Bioinformatika adalah bidang yang relatif baru, masih banyak kesalahpahaman mengenai definisinya. Komputer sudah lama digunakan untuk menganalisa data biologi, misalnya terhadap data-data kristalografi sinar X dan NMR (Nuclear Magnetic Resonance) dalam melakukan penghitungan transformasi Fourier, dsb [4]. Bidang ini disebut sebagai Biologi Komputasi. Bioinformatika muncul atas desakan kebutuhan untuk mengumpulkan, menyimpan dan menganalisa data-data biologis dari database DNA, RNA maupun protein tadi. Untuk mewadahnya beberapa jurnal baru bermunculan (misalnya Applied Bioinformatics), atau berubah nama seperti Computer Applications in the Biosciences (CABIOS) menjadi BIOInformatik yang menjadi *official journal* dari International Society for Computational Biology (ICSB) (nama himpunan tidak ikut berubah) [5]. Beberapa topik utama dalam Bioinformatika dijelaskan di bawah ini.

Keberadaan database adalah syarat utama dalam analisa Bioinformatika. Database informasi dasar telah tersedia saat ini. Untuk database DNA yang utama adalah GenBank di AS (Gambar 3) [6]. Sementara itu bagi protein, databasenya dapat ditemukan di Swiss-Prot (Swiss) [7] untuk sekuen asam aminonya dan di Protein Data Bank (PDB) (AS) [8] untuk

struktur 3D-nya. Data yang berada dalam database itu hanya kumpulan/arsip data yang biasanya dikoleksi secara sukarela oleh para peneliti, namun saat ini banyak jurnal atau lembaga pemberi dana penelitian mewajibkan penyimpanan dalam database. Trend yang ada dalam pembuatan database saat ini adalah isinya yang makin spesialis. Misalnya untuk protein struktur, ada SCOP [9] dan CATH [10] yang mengklasifikasikan protein berdasarkan struktur 3D-nya, selain itu ada pula PROSITE [11], Blocks [12], dll yang berdasar pada motif struktur sekunder protein.

Tak kalah penting dari data eksperimen tersebut adalah keberadaan database paper yang terletak di Medline [13]. Link terhadap publikasi asli biasanya selalu tercantum dalam data asli sekuen. Perkembangan Pubmed terakhir yang penting adalah tersedianya fungsi mencari paper dengan topik sejenis dan link kepada situs jurnal on-line sehingga dapat membaca keseluruhan isi paper tersebut.

Setelah informasi terkumpul dalam database, langkah berikutnya adalah menganalisa data. Pencarian database umumnya berdasar hasil alignment/pensejajaran sekuen, baik sekuen DNA maupun protein. Metode ini digunakan berdasar kenyataan bahwa sekuen DNA/protein bisa berbeda sedikit tetapi memiliki fungsi yang sama. Misalnya protein hemoglobin dari manusia hanya sedikit berbeda dengan yang berasal dari ikan paus. Kegunaan dari pencarian ini adalah ketika mendapatkan suatu sekuen DNA/protein yang belum diketahui fungsinya maka dengan membandingkannya dengan yang ada dalam database bisa diperkirakan fungsi daripadanya. Algoritma untuk *pattern recognition* seperti Neural Network, Genetic Algorithm dll telah dipakai dengan sukses untuk pencarian database ini. Salah satu perangkat lunak pencari database yang paling berhasil dan bisa dikatakan menjadi standar sekarang adalah BLAST (Basic Local Alignment Search Tool) [14]. Perangkat lunak ini telah diadaptasi untuk melakukan alignment terhadap berbagai sekuen seperti DNA (blastn), protein (blastp), dsb. Baru-baru versi yang fleksibel untuk dapat beradaptasi dengan database yang lebih variatif telah dikembangkan dan disebut Gapped BLAST serta PSI (Position Specific Iterated)-BLAST [15]. Sementara itu perangkat lunak yang digunakan

```

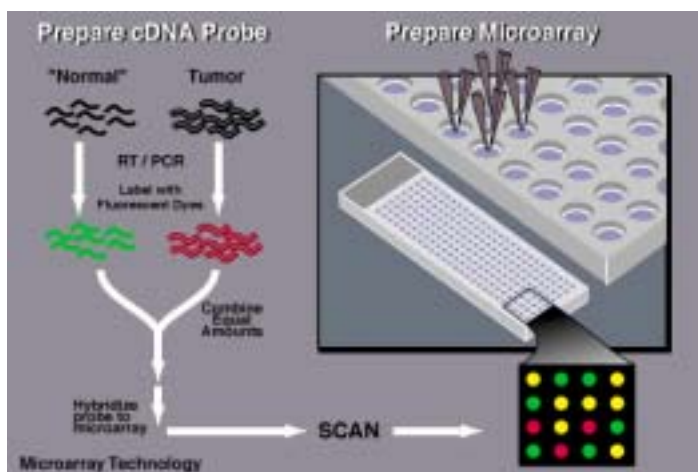
LOCUS       D3029FC          4903 bp    mRNA           287       13-MAR-1990
DEFINITION  D.melanogaster developmental gene complex (DPP-C), complete cds.
ACCESSION  M8115
VERSION    gi17291
KEYWORDS   .
SOURCE     D.melanogaster, cDNA to mRNA.
ORGANISM   Drosophila melanogaster
            Eukaryota; Eukaryota; Eukaryotes; Metazoa; Arthropoda;
            Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscophaga;
            Sigalphoidea; Drosophilidae; Drosophila.
REFERENCE  1. Ganes I to 4901
AUTHORS    Padgett, R.W., St Johnston, S.D., and Gilbert, W.K.
TITLE      A c-myc-like protein from a Drosophila pattern gene produces a protein
            homologous to the transducing growth factor-beta family
            (Nature 325, 81-84 (1987))
JOURNAL    Nature 325, 81-84 (1987)
MEDLINE    8708498
COMMENT    The initiation codes could be at either 1288-1290 or 1287-1289.
FEATURES   Location/Qualifiers
            1..4903
             /organism="Drosophila melanogaster"
             /db_xref="Genbank:U221"
             +3..3950
             /name="Dpp"
             /note="Developmental gene complex (DPP-C) mRNA"
             /db_xref="FlyBase:FBgn000490"
             1..4903
             /name="Developmental"
             /name="Dpp"
             /alias=""
             /db_xref="FlyBase:FBgn000490"
             1288..1294
             /name="Dpp"
             /note="Developmental protein (DPP) could be 1287"
             /codon_start=1
             /db_xref="FlyBase:FBgn000490"
             /db_xref="F02:gi17291"
             /translation="MSKELLVPLATQTIVYVDTTEIQGFIALAIPAAIPILA
            SAGDSDDSDSDVVAHTALAAKAPPPAPAPVQGRGIRDFNEDKADKAD
            .....
            LQGRFYVNDKCFPLACRFETBAVQGLVMSKPKYFACCVTGLDVMKEL
            SDGTVPLRNYSTVYVQDQK"
BASE COUNT 1276 a 1876 c 354 g 197 t
ORIGIN
1  ggtgttcaaa agpcttgcct gacttcaaac ctatcccgaa atpccgagc acaatgacc
2  cacttgcctt gactcccaag ctctccaga aactctccg acrcrcatc acatctatg
3  gaaatgactt gaaatgcttt ctgagcagc agcttccagc tggcccaag agctccagtc
4  gggcccaaaa caatccagc accttccagc agctctgtct ctgagcagc gactctctc
5  actctgctat gttcaatct aactccagc ttptcaagc aactcttca atctgagc
6  atgagcaaaa caataaagc tctcaatag acttccagc caaacatct tctgcaaac
7  gttccagctc cgtctccag tggagcaacc gttactgag atgcccacc aacccctcc
8  atgcaaaa atgagctgt ccaagctgc agtcaaaa tctgagctt actctgagc
9  agtcccaaaa agpcttgcct ggttcaagc tcttccccc agtcaaaa cctgagca
10 atgagcaaaa caatccagc tcccaagc agtctccag tctgcaatg gctccagct
11 atgagcaaaa gaaatgctt ctgagcagc ggtctccag tggagcaaa caatcttca
12 atgagcaaaa caaacatgct tggcttca atactgact caatctatc atctctatg
13 atgagcaaaa agpcttgcct aactcttca agttcaaa atgagcaaaa gttgagca
14 atgagcaaaa tccctgctt gaaatctca atgagcaaaa

```

Gambar 3. Data sekuen DNA dalam GenBank.

untuk melakukan alignment terhadap sekuen terbatas di antaranya yang lazim digunakan adalah CLUSTAL dan CLUSTAL W [16].

Data yang memerlukan analisa bioinformatika dan cukup mendapat banyak perhatian saat ini adalah data hasil DNA chip (Gambar 4). Menggunakan perangkat ini dapat diketahui kuantitas maupun kualitas transkripsi satu gen sehingga bisa menunjukkan gen-gen apa saja yang aktif terhadap perlakuan tertentu, misalnya



Gambar 4. DNA chip.

timbulnya kanker, dll. mRNA yang diisolasi dari sampel dikembalikan dulu dalam bentuk DNA menggunakan reaksi reverse transcription. Selanjutnya melalui proses hibridisasi, hanya DNA yang komplementer saja yang akan berikatan dengan DNA di atas chip. DNA yang telah diberi label warna berbeda ini akan menunjukkan pattern yang unik. Berbagai algoritma pattern recognition telah digunakan untuk mengenali gen-gen yang aktif dari eksperimen DNA chip ini, salah satunya yang paling ampuh adalah Support Vector Machine (SVM) [17].

Bioinformatika sudah menjadi bisnis besar sekarang. Perusahaan bioteknologi yang menghasilkan data besar seperti perusahaan sekuen genom, senantiasa memerlukan bagian analisa Bioinformatika. Produk Bioinformatika pun sudah dipatenkan baik di AS, Eropa maupun Asia [18]. Berdasar jenisnya produk yang dipatenkan itu bisa dibagi menjadi tiga yaitu (1) perangkat lunak Bioinformatika, termasuk diantaranya adalah perangkat lunak pencarian database dsb dengan contoh misalnya paten no. 6,125,331 di AS berjudul "Structural alignment method making use of a double dynamic programming algorithm", (2) metode Bioinformatika, ini menggunakan analogi metode bisnis telah dapat dipatenkan di AS seperti pada kasus pematenan Amazon.com, sebagai contoh adalah paten no. 6,125,383 di AS tentang "Research system using multi-platform object oriented program language for providing objects at runtime for creating and manipulating biological or chemical data", terakhir (3) produk Bioinformatika itu sendiri yaitu informasi biologis hasil analisisnya.

DARI MANA KITA MULAI?

Di Indonesia Bioinformatika masih belum dikenal oleh masyarakat luas. Di kalangan peneliti sendiri, mungkin hanya para peneliti biologi molekuler yang sedikit banyak mengikuti perkembangannya karena keharusan

menggunakan perangkat-perangkat Bioinformatika untuk analisa data. Sementara itu di kalangan TI masih kurang mendapat perhatian.

Ketersediaan database dasar (DNA, protein) yang bersifat terbuka/gratis merupakan peluang besar untuk menggali informasi berharga daripadanya. Sudah disepakati, database genom manusia misalnya akan bersifat terbuka untuk seluruh kalangan. Dari padanya bisa digali kandidat-kandidat gen yang memiliki potensi kedokteran/farmasi. Dari sinilah Indonesia dapat ikut berperan mengembangkan bioinformatika. Kerjasama antara peneliti bioteknologi yang memahami makna biologis data tersebut dengan praktisi IT seperti programmer, dsb akan sangat berperan dalam kemajuan Bioinformatika Indonesia nantinya.

REFERENSI

1. S. Giesecke (2000) The contrasting role of government in the development of biotechnology industry in the US and Germany. *Research Policy* 29, 205-223.
2. C. Robbins-Roth (2000) From Alchemy to IPO: The business of biotechnology. Perseus Publishing, New York, USA.
3. D. Gershon (1997) Bioinformatics in a post-genome age. *Nature* 389, 417-418.
4. H. Gavaghan (1997) Running to catch up in Europe. *Nature* 389, 420-422.
5. C. Sander (2001) Bioinformatics challenges in 2001. *Bioinformatics* 17, 1-2.
6. D. Benson, D.J. Lipman, and J. Ostell (1993) GenBank. *Nuclei Acid Research* 21, 2963-2965.
7. A. Bairoch, and R. Apweiler (1998) The SWISS-PROT protein sequence data bank and its supplement. *Nucleic Acid Research* 26, 38-42.
8. J.L. Sussman, et al. (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. D. Biol. Crystallogr.*, 54, 1078-1084.
9. G.J. Barton (1994) SCOP: structural classification of proteins. *Trends in Biochemical Sciences* 19, 554-555.
10. <http://www.biochem.ucl.ac.uk/bsm/cath>
11. A. Bairoch, P. Bucher, and K. Hofmann (1997) The PROSITE database, its status in 1997. *Nucleic Acid Research* 25, 217-221.
12. S.E. Brenner (1995) BLAST, Blitz, BLOCKS and BEAUTY: sequence comparison on the net. *Trends in Genetics*, 11, 330-331.
13. R.M. Woodsmall, and D.A. Benson (1993) Information resources at the National Center for Biotechnology Information. *Bull. Med. Libr. Assoc.* 81, 282-284.
14. S.F. Altschul, et al. (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-410.
15. S.F. Altschul, et al. (1997) Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research* 25, 3389-3402.
16. J.D. Thompson, D.G. Higgins, and T.J. Gibson (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research* 22, 4673-4680.
17. T.S. Furey, et al. (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 16, 906-914.
18. T. Maschio, and T. Kowalski (2001) Bioinformatics – a patenting view. *Trends in Biotechnology* 19, 334-339.

MENGENAI PENULIS



Dr. Arief B. Witarto dilahirkan di Lahat (Sumatera Selatan), Mei 1971. Menyelesaikan pendidikan sampai S-3 di Department of Biotechnology, Tokyo University of Agriculture and Technology, Jepang tahun 2000. Setelah itu bekerja sebagai peneliti di Japan Advanced Institute of Science and Technology, Jepang selama 2 tahun sebelum menempati posisinya sekarang. Terpilih sebagai Peneliti Muda Indonesia tahun 2002 terbaik untuk bidang Ilmu Pengetahuan Teknik dan Rekayasa. Spesialisasi pada bidang Rekayasa Protein yang dalam penelitiannya menggunakan *tool-tool* bioinformatika. Salah satu topik penelitiannya saat ini adalah analisa bioinformatika protein *folding motif*.