Bioinformatics for the Korean Functional Genomics Project

Sangsoo Kim
Korea Research Institute of Bioscience & Biotechnology
52 Oun-dong, Yusung-gu, Taejon, 305-333 KOREA

Abstract

Genomic approach produces massive amount of data within a short time period. New high-throughput automatic sequencers can generate over a million nucleotide sequence information overnight. A typical DNA chip experiment produces tens of thousands expression information, not to mention the tens of megabyte image files. These data must be handled automatically by computer and stored in electronic database. Thus there is a need for systematic approach of data collection, processing, and analysis. DNA sequence information is translated into amino acid sequence and is analyzed for key motif related to its biological and/or biochemical function. Functional genomics will play a significant role in identifying novel drug targets and diagnostic markers for serious diseases. As an enabling technology for functional genomics, bioinformatics is in great need worldwide. In Korea, a new functional genomics project has been recently launched and it focuses on identifying genes associated with cancers prevalent in Korea, namely gastric and hepatic cancers. This involves gene discovery by high throughput sequencing of cancer cDNA libraries, gene expression profiling by DNA microarray and proteomics, and SNP profiling in Korea patient population. Our bioinformatics team will support all these activities by collecting, processing and analyzing these data.

Key Ideas :

EST sequencing of stomach and liver cDNA libraries will produce many redundant sequences. These will be clustered in order to identify unique clones. The goal is to identify 10-20k unique genes for each of stomach and liver tissues. These clones will be used in the production of stomach- or liver-specific DNA chips. These expression profile data will be clustered over both genes and tissues. The goal is to identify cancer marker genes and to sub-classify cancer-types. Patient information will be collected in conjunction with clinical tissues and will be used to draw correlation between gene expression profiles and epidemiological factors. It is also hoped that this expression profile data give clues to identifying novel anti-cancer drug targets.

Methods & Research Contents :

EST sequence data will be collected from our MegaBACE and ABI 3700 automatics sequencers and transfered to Linux servers for automatics base-calling and vector trimming. Autonomous

BLAST servers are being set up for parallel annotation of more than 2,000 sequences per day. EST sequences are first compared to NCBI's UniGene database, subsequently to non-redundant peptide sequence database via BLASTX. As genomic sequences become available, each EST sequence will be mapped to the chromosome loci.

As an initial attempt, we will produce DNA microarray containing about 10,000 known human genes. Gene annotaion information as well as expression profile data from 6 clinical pathology teams will be collected and stored in a relational database system. The system will be scalable as the amount of data grows and provide easy retrieval. There are several published clustering methods for DNA chip data. Employing these methods, the expression data will be clustered in two ways: one over genes and the other over tissues. Statistical analysis will be implemented in order to identify cancer-specific genes.

In proteomics and SNP areas, our role is primarily on data archiving.

Curriculum vitae

Personal data

Sangsoo Kim

Born in Seoul on Aug. 16, 1958

Work address

Korea Research Institute of Bioscience & Biotechnology

Genome Research Center

Team Leader of Bioinformatics

Oun-dong 52, Yusong-gu, Taejon 305-333

Tel 042-860-4177

Fax 042-860-4409

E-mail sskimb@mail.kribb.re.kr

Experience

Korea Research Institute of Bioscience & Biotechnology

Mar. 1, 2000 ~ present

Team Leader of **Bioinformatics**

LG Chem., LTD

Jan. 1, 1999 ~ Mar. 1, 2000

Team Leader of **Bioinformatics**

DNA chip, Proteomics, Data Mining

Drug target identification, validation

Dec. 12, 1989 ~ Dec. 31, 1998

Team Leader of **Biopharmaceutical Design**

Structure-based Drug Design

HIV protease inhibitors, thrombin inhibitors

Purdue University, West Lafayette, IN

Dec. 29, 1986 ~ Dec. 10, 1988

Postdoctoral Res. Ass., Dept. of Bio. Sci. (Prof. M.G.Rossmann)

Virus crystallography

Human Rhinovirus type 1A, Mengo Virus

Education

Iowa State University

Jan., 1983 ~ Dec. 20, 1986

Ph.D. in Physical Chemistry

Thesis Title: X-ray diffraction studies of bismuth-doped lead dioxide electrodes, of the radiation-damaged benzene chromium tricarbonyl crystal structures, and of selected organometallic compounds

Thesis Advisor: Prof. R.A. Jacobson

Seoul National University, Seoul, Korea

Mar. 1, 1981 ~ Feb. 26, 1983

M.S. in Physical Chemistry

Thesis Title: Crystal and Molecular Structure of Thiamphenicol"

Advisor: Prof. Whanchul Shin

Mar. 1, 1977 ~ Feb. 26, 1981

B.S. in Chemistry

Academic societies

The Biochemical Society of the republic of Korea

The Korean Chemical Society

International Society for Computational Biology

Publications

Discovery of LB30057, A Benzamidrazone-based selective oral thrombin inhibitor. Yeong Soo Oh, Mkyung Yun, Sang Yeul Hwang, Seongwon Hong, Youseung Shin, Koo Lee, Kyung Hee Yoon, Yung Joon Yoo, Dong Soo Kim, Sun Hwa Lee, Yong Hee Lee, Hee Dong Park, Chang Ho Lee, Sang Koo Lee and Sangsoo Kim *Bioorganic & Medicinal Chemistry Letters* (1998) 8:631-634.

Structural modification of an orally active thrombin inhibitor, LB30057: replacement of the D-pocket-binding naphthyl moiety. Koo Lee, Sang Yeul Hwang, Seongwon Hong, Chang Yong Hong, Chang-seok Lee, Youseung Shin, Sangsoo Kim, Mikyung Yun, Yung Joon Yoo, Myunggyun Kang and Yeong Soo Oh *Bioorganic & Medicinal Chemistry* (1998) 6:869-876.

Rational design of selective thrombin inhibitors. Sangsoo Kim, Sang Yeul Hwang, Young Kwan Kim, Mikyung Yun and Yeong Soo Oh *Bioorganic & Medicinal Chemistry Letters* (1997) 7:769-774.

Total synthesis of (+)-cladantholide and (-)-estafiatin: 5.exo,7-endo radical cyclization strategy for the construction of guaianolide skeleton. Eun Lee, Jong Won Lim, Cheol Hwan Yoon, Yong-suk Sung, Young Kwon Kim, Mikyung Yun and Sangsoo Kim *Journal of the American Chemical Society* (1997) 119:8391-8392.

Design, synthesis, and characterization of dipeptide isostere containing cis-epoxide for the irreversible inactivation of HIV protease. Chang Sun Lee, Nakyen Choy, Chihyo Park, Hoil Choi, Young Chan Son, Sangsoo Kim, Jong Hoa Ok, Heungsik Yoon and Sung Chun Kim *Bioorganic & Medicinal Chemistry Letters* (1996) 6:589-594.

Fused Oxacycle synthesis via radical cyclization of β-alkoxyacrylates. Eun Lee, Jin Sung Tae, You Hoon Chong, Yong Cheol Park, Mikyung Yun and Sangsoo Kim *Tetrahedron Letters* (1994) 35:129-132.

Orthocyclophanes. 3. Ketonands, novel ketonic crowns of polyoxo[1_n]orthocyclophane constitution. Woo Young Lee, Chang Hee Park, Hyo-Jung Kim and Sangsoo Kim *The Journal of Organic Chemistry* (1994) 59:878-884.

Crystallization and preliminary X-ray crystallographic analysis of phospholipid transfer protein from maize seedings. Dong Hae Shin, Kwang Yeon Hwang, Kyeong Kyu Kim, Sangsoo Kim,

Robert M. Sweet and Se Won Suh *Proteins* (1994) 19:80-83.

Solid-state reactions of iridium(I)-1,5-cyclooctadiene compounds with CO: synthesis of cationic (1,5-cyclooctadiene)carbonyliridium(I) complexes. Chong Shik Chin, Byeongno Lee and Sangsoo Kim *Organometallics* (1993) 12:1462-1466.

A novel star-shaped crown ether: spontaneous isomerization of a macrocyclic polyketone to a spirobicyclic polyketal. Woo Young Lee, Chang Hee Park and Sangsoo Kim *Journal of the American Chemical Society* (1993) 115:1184-1185.

Preparation of a cyclopropane-containing analogue of artemisinin. Soo-Un Kim, Jin-Hyuk Choi and Sangsoo Kim *Journal of Natural Products* (1993) 56:857-863.

Solution dynamics and crystal structure of $CpMoOs_3(CO)_{10}(\mu\text{-}H)_2[\mu_3\text{-}\eta^2\text{-}C(O)CH_2Tol]$. Joon T. Park, Jeong-Ju Cho, Kang-Moon Chun, Sock-Sung Yun and Sangsoo Kim *Bull. Korean Chem. Soc.* (1993) 14:137.

A novel (Tricobaltiomethoxy)silatrane: Synthesis and structure of $[Co_3(CO)_9\{\mu_2\text{-}COSi(OCH_2CH_2)_3N\}]$. Myong W. Kim, Dong Seon Uh, Sangsoo Kim and Youngkyu Do *Inorganic Chemistry* (1993) 32:5883.

Novel exo-nido-cupracarborane wih an unusual solution fluxionality. Synthesis and structure of $[9,10\text{-exo-}\{Cu(PPh_3)_2\}\text{-}9,10\text{-}(\mu\text{-}H)_2\text{-}7,8\text{-nido-}C_2B_9H_{10}]$. Young-Whan Park, Jinkwon Kim, Sangsoo Kim and Youngkyu Do *Chemistry Letters* (1993) 121-124.

X-ray crystallographic study of covalently modified carboxypeptidase A by 2-benzyl-3,4-epoxybutanoic acid, a pseudomechanism-based inactivator. Mikyung Yun, Chihyo Park, Sangsoo Kim, Doohyun Nam, Sung Chun Kim and Dong H. Kim *Journal of the American Chemical Society* (1992) 114:2281-2282.

Crystallization and preliminary X-ray crystallographic analysis of lipase from Pseudomonas cepacia. Kyeong Kyu Kim, Kwang Yeon Hwang, Hye Sung Jeon, Sangsoo Kim, Robert M. Sweet, Chul Hak Yang and Se Won Suh *Journal of Molecular Biology* (1992) 227:1258-1262.

Synthesis, reactivity, and molecular structure of [(phenylsilatrane)$Mn(CO)_2P(OMe)_3$]ClO_4. Ae-Soon Oh, Young Keun Chung and Sangsoo Kim *Organometallics* (1992) 11:1394-1398.

Endo-σ-bonded group 14 heterodicarboranes: Syntheis of $[Ph_3MC_2B_9H_{11}]^-$ [M=Germanium(IV), Tin(IV) and structure of [10-endo-($SnPh_3$)-10-μ-H-7,8-nido-$C_2B_9H_{10}$][trans-Ir(CO)(PPh_3)$_2$(MeCN)] Jinkwon Kim, Sangsoo Kim and YoungKyu Do *J. Chem. Soc., Chem. Commun.* (1992) 938.

Internal hydride assisted C-O bond scission of a μ_3-η^2-coordinated acyl and an apparent rotation of μ-alkylidene carbon in molybdenum-triosmium cluster compounds. Joon T. Park, Kang M. Chun, Sock S. Yun and Sangsoo Kim *Bulletin of the Korean Chemical Society* (1991) 12:249-251.

Crystallization and preliminary structure analysis of an insect virus with T=4 quasi-symmetry: Nudurelia capensis ω virus. Jean Cavaralli, Wu Bomu, Lars Liljas, Sangsoo Kim, Wladek Minor, Sanjeev Munshi, Steve Muchmore, Tim Schmidt and John E. Johnson *Acta. Cryst.* (1991) B47:23-29.

Crystallization and a preliminary X-ray crystallographic study of α-amylase from Bacillus licheniformis. Suk Yeong Lee, Sangsoo Kim, Robert M. Sweet and Se Won Suh *Archiv. Biochem. Biophys.* (1991) 291:255-257.

Conformational variability of a picornavirus capsid: pH-dependent structural changes of Mengo virus related to its host receptor attachment site and disassembly. Kim S, Boege U, Krishnaswamy S, Minor I, Smith TJ, Luo M, Scraba DG, Rossmann MG *Virology.* 1990 Mar;175(1):176-90.

Crystal structure of human rhinovirus serotype 1A (HRV1A). Kim S, Smith TJ, Chapman MS, Rossmann MC, Pevear DC, Dutko FJ, Felock PJ, Diana GD, McKinlay MA *J Mol Biol.* 1989 Nov 5;210(1):91-111.

Patents issued

Novel HIV protease inhibitors having anti-AIDS activity and their composition and process thereof *Korea* 97-09-18 132093

Irreversible HIV protease inhibitors *Korea* 97-12-3 134497

Selective thrombin inhibitors *Korea* 98-09-08 163669

Non-peptide thrombin inhibitors *Korea* 98-09-08 163668

Selective thrombin inhibitors *Korea* 96-01-06 153490

Selective thrombin inhibitors *Korea* 98-10-27 173035

Selective thrombin inhibitors *USA* 98-5-5 5,747,535

Selective thrombin inhibitors *USA* 98-11-2 5,977,114

Selective thrombin inhibitors *USA* 98-11-16 5,985,899

Selective thrombin inhibitors *Korea* 98-10-27 173034

Irreversible HIV protease inhibitors, intermediates, compositions and process thereof *USA* 97-12-09 5,696,134

Irreversible HIV protease inhibitors, intermediates, compositions and process thereof *USA* 98-4-28 5,744,621

Irreversible HIV protease inhibitors, intermediates, compositions and process thereof *USA* 98-6-9 5,763,631

Cis-epoxide-based irreversible HIV protease inhibitors, intermediates, compositions and process thereof *Europe* 97-10-29 EP0601486B1