

Uemura A, (竹原)	Natural killer cell is a major producer of interferon gamma that is critical for the IL-12-induced anti-tumor effect in mice.	Cancer Immunol Immunother	59(3)	453-463	2010
Yamaguchi S, (竹原)	EphA2 derived peptide vaccine with amphiphilic poly (γ -glutamic acid) nanoparticles elicits an antitumor effect against mouse liver tumor.	Cancer Immunol Immunother	(In press)		
Kanmura S, (宇都)	The complement component C3a fragment is a potential biomarker for hepatitis C virus-related hepatocellular carcinoma.	J Gastroenterol	(In press)		2010
Takami Y, (宇都)	Identification of a novel biomarker for oxidative stress induced by hydrogen peroxide in primary human hepatocytes using the 2-nitrobenzenesulfonyl chloride isotope labeling method	Hepatol Res	(In press)		2010
Nishida C, (宇都)	Clinical significance of alanine aminotransferase levels and the effect of ursodeoxycholic acid in hemodialysis patients with chronic hepatitis C	J Gastroenterol	(In press)		2010
Takami Y, (宇都)	Proanthocyanidin derived from the leaves of <i>Vaccinium virgatum</i> suppresses platelet-derived growth factor-induced proliferation of the human hepatic stellate cell line LI90	Hepatol Res	(In press)		2010
Uto H, (宇都)	Increased rate of death related to presence of viremia among hepatitis C virus antibody-positive subjects in a community-based cohort study.	Hepatology	50(2)	393-9	2009
Takeshita M, (宇都)	Proanthocyanidin from blueberry leaves suppresses expression of subgenomic hepatitis C virus RNA.	J Biol Chem	284(32)	21165-76	2009
Enomoto N, (前川)	HCV genetic elements determining the early response to peginterferon and ribavirin therapy.	Intervirology	53(1)	66-9	2010
Maekawa S, (前川)	Viral factors influencing the response to the combination therapy of peginterferon plus ribavirin in chronic hepatitis C.	J Gastroenterol	44(10)	1009-15	2009
Itakura J, (前川)	Reproducibility and usability of chronic virus infection model using agent-based simulation; comparing with a mathematical model.	Biosystems	99(1)	70-8	2010

Sekine-Osajima Y, (前川)	Two flavonoids extracts from Glycyrrhizae radix inhibit in vitro hepatitis C virus replication.	Hepato Res	39(1)	60-9	2009
Kusakabe A, (田中)	Case-control study for the identification of virological factors associated with fulminant hepatitis B.	Hepato Res	39(7)	648-656	2009
Tanida I, (花田)	Knockdown of autophagy-related gene decreases the production of infectious hepatitis C virus particles.	<i>Autophagy</i>	5 (7)	937-45	2009
T Kawakami, (菅)	Diverse backbone-cyclized peptides via codon reprogramming	<i>Nature Chemical Biology</i>	5(12)	888-890	2009
N Niwa, (菅)	A flexizyme that selectively charges amino acids activated by a water-friendly leaving group	<i>Bioorganic Medicinal Chemistry Letter</i>	19	3892-3894	2009
Y Goto, (菅)	Ribosomal synthesis of dehydrobutyrine- and methyllanthionine-containing peptides	<i>Chemical Communication</i>	-	3419-3421	2009
Y Yamagishi, (菅)	Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction	<i>ChemBioChem</i>	10	1469-1472	2009
E Nakajima, (菅)	Ribosomal synthesis of peptides with C-terminal lactams, thiolactones, and alkylamides	<i>ChemBioChem</i>	10	1186-1192	2009
Y Goto, (菅)	Translation initiation with initiator tRNA charged with exotic peptides	Journal of American Chemical Society	131(14)	5040-5041	2009
H Murakami, (菅)	Bases in the anticodon loop of tRNA(GGC)(Ala) prevent misreading	<i>Nature Structural & Molecular Biology</i>	16(4)	353-358	2009
Tokumoto Y, (堀本)	TRAIL inhibited the cyclic AMP responsible element mediated gene expression	Biochem Biophys Res Commun	381	533-536	2009

IV. 研究成果の刊行物・別刷

13. 肝疾患の血清プロテオミクスを用いた診断

鹿児島大学大学院医歯学総合研究科消化器疾患・生活習慣病学講師 宇都浩文

同 消化器疾患・生活習慣病学 上村修司

同 教授 坪内博仁

key words hepatocellular carcinoma, liver cirrhosis, proteomics, 2-DE, ProteinChip SELDI system, MALDI-TOF MS

動 向

現在、肝細胞癌 hepatocellular carcinoma (HCC) の血液腫瘍マーカーとして alpha-fetoprotein (以下AFP), protein induced by Vitamin K absence or antagonist-II (PIVKA-II), AFPレクチン分画L3 (AFP-L3) が臨床で用いられ、その有用性が広く知られている。しかし、早期HCCにおけるAFPやPIVKA-IIの感度は約30~60%と、決して高感度ではない¹⁻⁴⁾。AFPとPIVKA-IIの同時測定でHCC診断能は上昇するものの、2cm以下の早期HCCでは診断能は低く、早期発見には、新たな診断法の開発、新規の腫瘍マーカーの同定が必要と考えられる⁵⁻⁷⁾。また、肝硬変診断のための血清マーカーとしては、ヒアルロン酸、IV型コラーゲンなどの肝線維化マーカーが用いられているが、必ずしも肝臓の線維化を反映しない^{8,9)}。さらに、欧米諸国をはじめ本邦でも増加傾向にある非アルコール性脂肪肝疾患 nonalcoholic fatty liver disease (NAFLD)、とくに非アルコール性脂肪肝炎 nonalcoholic steatohepatitis (NASH) の確定診断には侵襲的な肝生検が必須で、特異性のある血清マーカーはない^{10,11)}。したがって、HCCの早期診断あるいは肝硬変からの発癌予測、肝硬変と慢性肝炎の鑑

別、さらにはNASHの診断などに有用な、かつ特異的な血清マーカーの同定が必要である。

プロテオミクス (Proteomics)、またはプロテオーム解析 (Proteomic analysis) とは、生物のタンパク質の構造や機能を網羅的に研究することである。遺伝子発現量とタンパク質濃度は、生体において必ずしも比例せず、さらに、タンパク質のもつ生理活性はリン酸化などの翻訳後修飾において制御されていることが多い^{12,13)}。そのため、肝疾患だけでなく様々な疾患の病態解明やバイオマーカー探索においてプロテオミクスは有用と考えられる。また、血清や血漿などは、生体から低侵襲、かつ容易に得ることができ、臨床サンプルとして利用しやすい。患者血清を用いて高感度かつ特異度の高い診断や病態と関連するバイオマーカーを見出せば、臨床的な有用性が高い。肝疾患においては、以前から二次元電気泳動法 2-dimensional polyacrylamide gel electrophoresis (2-DE) を中心としたプロテオミクスが行われてきたが、その大部分がHCCを含む肝組織を用いた解析である¹⁴⁻¹⁶⁾。近年、質量分析を中心とした解析装置だけでなく、タンパク標識法やアミノ酸解析法などのプロテオミクス技術の進歩により、患者血清などに含まれるタンパク質

を網羅的に比較的容易、かつ感度よく解析できるようになってきている^{17,18)}。このようなプロテオミクス技術の進歩により、HCCだけでなく、肝硬変やNAFLDなどを対象としたバイオマーカー探索も進んでいる^{8,9,11,19-40)}。

最近の肝疾患における血清マーカー探索を目的としたプロテオミクスは、2DEで分離し、発現量の変化のあるスポット（タンパク質）を質量分析計により同定する方法¹⁹⁻²¹⁾とProteinChip SELDI システム（バイオ・ラッド ラボラトリーズ株式会社）により低分子領域のタンパク質発現をピーク値として数値化し、比較する方法²²⁻³²⁾が主に用いられている。しかし、このシステムによるタンパクのピーク強度には解析装置間や施設間でばらつきがある可能性があり、現時点ではピーク強度の比較のみで臨床応用するのは困難と考えられる。なお、ProteinChip SELDI システムで得られたピークタンパクの一部は同定が可能となってきた^{22,23,26-30)}。

A. 肝細胞癌の血清プロテオミクス

1. 二次元電気泳動による血清解析

従来から用いられてきた2DEによるタンパク質分離法は、低分子領域のタンパク質解析には不向きであるが、扱いやすさや再現性の面では最も確立している手法であり、現在も広く用いられている。Yangらは、2-DEとnano-HPLC-ESI-MS/MSを用いて、5例（HBs抗原陽性2例、HCV抗体陽性2例、いずれも陰性1例）のHCC患者血清と健常者血清を比較し、発現に差のある317のタンパク質を分離・同定し、そのうち、6個のタンパク質（Annexin VI isoform 1, Complement component 9, Ceruloplasmin, Serum amyloid A4, A2, A1 isoform 2）はHCCのマーカーとなる可能性があることを報告している¹⁹⁾。また、Angらは、HCCと慢性肝疾患 chronic liver

disease (CLD) 患者血清を用いて糖化ハプトグロビンの発現量を比較検討し、特異的糖化ハプトグロビンはHCC群で発現上昇し、早期HCCより進行期HCCで高発現していることから、糖化ハプトグロビンはHCCの診断や病期判定に有用であると報告している²⁰⁾。FengらはHBV-HCCの患者血清を2DEで比較検討し、有意差のある8個のタンパク質〔Transferrin, Transthyretin, α 1-antitrypsin, Clusterin, Haptoglobin α 2chain, Ceruloplasmin, Heat-shock protein 27 (HSP27), AFP〕を同定した。HSP27はHCCの90%で陽性であり、HCCのスクリーニング検査に有用である可能性を示した²¹⁾。このように、2DEを用いた血清解析では、ゲルから抽出したタンパク質スポットをpeptide-mass fingerprinting (PMF) 解析やMS/MS解析することにより直接タンパク質を同定でき、有用な診断マーカーを発見できる可能性がある。今後、AFP, PIVKA-IIといった既存のマーカーとの比較や前向き研究によりその臨床的有用性を検証することが必要である。

2. ProteinChip SELDIシステムによる血清解析（表1）

ProteinChip SELDI システムは、プロテインチップ、飛行時間型質量分析計 time-of-flight mass spectrometry (TOF-MS) およびデータ解析用コンピューターから構成されており、化学的性質を表面にもつ特殊なチップに試料中のタンパク質を結合させ、結合したタンパク質を質量分析装置で解析する方法である。また、このシステムは、少量のサンプルを迅速かつ網羅的に解析できるため、近年の血清プロテオミクスに頻用されている²²⁻³¹⁾。さらに、2DEでは低分子領域（6kDa以下）のタンパク質・ペプチドや等電点（pI）が3以下もしくは11以上のタンパク質の解析が困難であるが、ProteinChip SELDI システムでは、

表1 ProteinChip SELDIシステムを用いた肝細胞癌、肝硬変患者の血清プロテオミクスの最近の報告例

著者	発表年	症例	比較対象	感度/特異度	その他の特徴	文献
Göbelら	2006	HCV-HCC 34例	HCV-LC 44例 (HCV-CH 39例)	HCC vs LC: 68%/80% (LC vs CH: 80%/ 67%)	肝硬変で減少する6646Da のピークタンパクを apolipoprotein C-Iと同定.	26
Wardら	2006	HCV-HCC 77例	HCV-CLD 105例	94%/86%	ピークタンパク (23/23.5 kDa) をκ/λimmunoglobulin light chains と同定.	27
Kanmuraら	2007	HCV-HCC 64例	HCV-LC 77例	83%/76%	AFP, PIVKA-IIより, 早期診 断に有用.	24
Cuiら	2007	HBV-HCC 81例 HBV-LC 54例 HBV-CH 75例	健常者87例	HCC vs 健常者: 90.5%/89.4% LC vs 健常者: 100%/86.5% CH vs 健常者: 85.3%/84.2%	HBV-HCCもしくはHBV-CLD を健常者と判別可能.	25
Gengら	2007	HBV-HCC 25例 (未加療)	健常者50例	92%/100%	HBV-HCC加療, 未加療と健常 者とで有意差のあった7ピーク タンパクを同定.	29
Zinkinら	2008	HCC 41例	HCV-LC 51例	79%/86%	最大径2cm以内の小肝細胞癌 の診断では, 既存マーカーより も, 高率に診断可能. 既存マ ーカーと組み合わせることによ り診断率はさらに向上.	22
Heら	2008	HBV-HCC 50例	HBV-非HCC 45例	100%/87%	11615 Daのピークタンパクを Serum Amyloid A (SAA) と 同定.	23
Heら	2008	HBV-HCC 81例	健常者 33例	95.92%/100%	6個のピークタンパク (157.33, 4177.02, 4284.79, 4300.80, 7789.87, 7984.14 m/z) を検 出, 7984 m/zのピークタンパ クをNeutrophil-activating peptide 2 (NAP-2) と同定し, 肝癌組織で発現を証明.	30
Poonら	2005	HBV-LC 46例	なし	89%/89%	肝線維化スコア (Ishakスコア) と相関あり.	32
Morraら	2007	HCV-LC 191例	なし	81%/80% (fibrosis scoreが 0.5の時)	8個のピークタンパクを用いた 判別法が, FibroTest (γ-GTP などを用いて肝線維化を評価 する方法) より肝線維化の評価 に有用.	8

HCC; 肝細胞癌, LC; 肝硬変, CH; 慢性肝炎, HCV; C型肝炎ウイルス (関連), HBV; B型肝炎ウイルス (関連)

再現性良く解析できる。しかし、解析で得られたピークタンパクは直接同定できないため、タンパク質同定にはSDS-PAGEや2-DEを用いたゲル泳動およびPMF法などによる解析が必要である^{22,23,26,28,30}。

ProteinChip SELDI システムを用いた血清学的診断には、シングルマーカーを同定し、その機能解析を行って診断に用いる手法と、データマイニングによりクラス分類を行い診断に用いる手法がある。ProteinChip SELDI システムでは、一般に目的ピークタンパクの同定が困難であり、個々のピークタンパクを同定せず、複数のピーク強度により、疾患群と対象群を判別していくクラス分類による診断法（マルチマーカー解析法）が用いられることが多い。我々は、HCV-HCCとHCV-CLD間で判別能が高かった7個のピークを用いて、クラス分類法を確立した²⁴。この方法では、腹部超音波検査により腫瘍性病変が検出される以前にHCCの存在を予想することが可能であり、AFPやPIVKA-IIといった既存の腫瘍マーカーより、HCCの早期診断に有用であった²⁴。Zinkinらは、ProteinChip SELDI システムにより検出した11個のピークを用いた診断法を開発し、この方法を用いるとHCV-HCCを感度79%、特異度86%で診断可能であった報告している²²。この診断法では感度、特異度ともAFP、AFP-L3分画やPIVKA-IIと診断率に大きな差は無かったが、最大径2cm以内のHCCの診断では、既存マーカーよりも、高率に診断可能であった。また、Heらは、HBV-HCC患者の血清解析で、3個の有意差のあるピーク(5890, 11615, 11724 Da)を選別し、HBV-HCCをほぼ100%診断できたと報告した²³。さらに、2DEと組み合わせて11615 Daのピークタンパクを、Serum Amyloid A (SAA)と同定している。しかし、この解析ではHBV-CLD患者でもこのピークが高率に陽性であり、HCC特異的な診断法ではないと考えられ

る。また、Cuiらは、HBV-HCCもしくはHBV-CLDの患者血清を用い、マルチマーカー解析法により、感度、特異度90%以上でHBV-HCCもしくはHBV-CLDを健常者と判別可能であったと報告しているが²⁵、HBV-HCCとHBV-CLDとの鑑別における有用性については疑問が残る。同じように、HBV-HCCの診断の感度は100%、特異度は92%もしくは97%であったとする報告もあるが^{29,30}、対照群が健常者であり、肝硬変患者におけるHCCの早期診断に有用であるかは疑問である。一方、Göbelらは、7486, 12843, 44293および53598 Daの4つのピークを用いたHCV-HCCと肝硬変群の判別法（マルチマーカー解析法）を作成し、早期のHCCの診断に有用であると報告した²⁶。同じように、Wardらも、感度94%、特異度86%、AUROC 0.92のマルチマーカー解析法を報告している²⁷。このように、マルチマーカー解析法は診断に有用であるものの、解析にProteinChip SELDIシステムが必要であり、簡便かつ迅速にどこ施設でも行えるものではない。また、このシステムによるピーク強度には解析装置間や施設間でばらつきがある可能性があり、大きな意味で再現性が問題であり、我々の経験ではマルチマーカーによるHCCの早期診断法を臨床応用するのは現時点では困難であると考えられる。

一般に、AFPやPIVKA-IIのように腫瘍が発現しているタンパク質は、局所治療や腫瘍切除後に血中濃度が低下し、治療効果判定や再発の指標になる。しかし、HCV-HCC治療前後の血清タンパク質発現をProteinChip SELDI システムにより解析した結果では、治療後に血中濃度が正常化するピークは見出されていない³¹。この結果は、腫瘍から直接産生されているタンパク質をProteinChip SELDIシステムにより解析することの限界を示しているかもしれない。さらに、慢性肝炎・肝硬変の進展に伴い変化するタンパク質も

存在することから、HCC患者血清のプロテオーム解析結果の解釈には留意すべきである。

B. 肝硬変、肝線維化の血清プロテオミクス

肝硬変の診断マーカーや、肝線維化マーカー探索を目的として肝硬変を対象としたプロテオミクスも行われている。Poonらは、ProteinChip SELDIシステムから得た肝硬変患者血清中の7個のピークとALT、総タンパク、およびビリルビン値を組み合わせて、肝線維化診断スコアを作成している。このスコアを用いることにより感度94%、特異度91%で肝硬変の診断が可能である³²⁾。Morraらも8個のピークを用いた判別法が、FibroTest (α 2-macroglobulin, haptoglobin, apolipoprotein A1, total bilirubin, γ -GTPを用いて肝線維化を評価する方法)より肝線維化の評価に有用であることを報告している⁸⁾(表1)。さらに、Göbelらも血清中の5個のピークを選び、肝硬変と非肝硬変(F1およびF2)を判別するマルチマーカー解析法を作成し、感度80%、特異度67%で肝硬変の診断が可能であることを明らかにした²⁶⁾。彼らは、肝硬変で減少しているタンパクの一つがapolipoprotein C-Iであることも報告している。しかし、多くの研究者により報告されたこれらのピークタンパクは、同じ手法にもかかわらず、必ずしも同一でない。このことは、先に述べたように検査法としてのProteinChip SELDIシステムそのものの限界を示している。

HCV-CLD患者血清を2DEにより解析した成績では、肝硬変群でinter-alpha-trypsin inhibitor heavy chain H4 (ITI4) fragments, α 1-antichymotrypsin, apolipoprotein L1, prealbumin, albumin, paraoxonase/arylesterase 1, zinc- α 2-glycoproteinが減少し、CD5 antigen-like protein (CD5L)と β 2 glycoprotein I (β 2GPI)が上昇している⁹⁾。こ

のように肝硬変で変動する血清タンパクを同定し、肝硬変診断における有用性を検証することにより新しい診断マーカーを見出せる可能性がある。

C. 非アルコール性脂肪肝疾患 (NAFLD) の血清プロテオミクス

最近、患者数の増加やHCCの原因となることから、NASHを含むNAFLDに関心が集まっている。NASHはアルコール性肝疾患と類似の病態を呈し、炎症・線維化を伴って肝硬変やHCCに進展する。NASHの病態には、過剰な栄養、肥満、それを基盤としたインスリン抵抗性、酸化ストレスなどが関与すると考えられているが、その病態は十分明らかにされていない。また、NAFLDの診断には組織学的な診断が必須で、特異的な血清マーカーがない。検診受診者の約30%が血液生化学検査異常を呈し、そのほとんどがNAFLDと推測されていることから、NAFLDやNASHの血清マーカーの同定は極めて重要である¹⁰⁾。

高脂血症と脂肪肝を呈するHcB19マウスの肝組織を用いたプロテオミクスでは、プロピオニルCoAからメチルマロニルCoAへの反応に関わっているpropionyl CoA carboxylaseの α chain (PCCA)が低下し³³⁾、確かに、PCCAノックアウトマウスではケトアシドーシスと脂肪肝の発症が見られる⁴¹⁾。一方、NAFLD患者血清を対象としたProteinChip SELDIシステムによる解析では、NAFLDのない肥満患者と比較してNASH患者で有意に変動する4個のピークが見出されている¹¹⁾。これらのピークタンパクはNASHの診断に有用である可能性はあるものの、残念ながら、タンパク質の同定ができていない。また、この研究では、同じ対象の肝組織中のmRNA発現がマイクロアレイを用いて網羅的に解析されており、NAFLDの病態進展や機序解明の研究には非常に

興味ある解析であるが、血清中のタンパク質発現と肝組織の遺伝子発現との関連は述べられていない。我々も、NAFLD患者血清を対象にProteinChip SELDIシステムを用いて解析し、健常者と比較してNAFLD患者で高値を示すピークを見出している³⁴⁾。真に、NAFLDに特異的か、NASHの診断に有用であるタンパク質かどうかなど今後の研究課題である。

D. 肝疾患におけるプロテオミクスの展望

ProteinChip SELDIシステムではタンパク同定ができないため、最近ではMALDI-TOF MSシステムを用いたHCCマーカー探索も行われるようになってきている³⁵⁻³⁸⁾。このシステムでは、検出したピークのうち低分子領域、特に3000m/z以下では直接タンパク質・ペプチドの同定が可能であり、それによりdes-Ala-fibrinopeptide A³⁵⁾、補体C3³⁶⁾、補体C4a³⁷⁾などがHCCで上昇することが明らかにされている。また、このシステムで得られたピーク強度を用いた肝硬変診断法も報告されている³⁸⁾。さらに、タンパク質の糖鎖修飾の違いに着目したプロテオミクスも行われるようになってきている³⁷⁻³⁹⁾。Comunaleらは、HBV-HCCの血清を解析し、AFPなどがHBV-HCCで高値であることを確認しただけでなく、総 α -1 acid glycoprotein (AGP)濃度はHCCと対照群で差がないにもかかわらず、フコシル化AGPがHCCで増加していることを明らかにしている³⁹⁾。今後、低分子領域のペプチドをターゲットとしたPeptidomics³⁵⁾や糖鎖修飾をターゲットとしたGycoproteomics^{20,37-39)}も肝疾患のマーカー探索に有用であると考えられる。

また、Paradisらはインターフェロンとリバビリン治療を行った96名のC型慢性肝炎患者血清をProteinChip SELDIシステムで解析し、治療経過中に変化したピーク数が、著効患者では無

効患者よりも有意に多いことを報告した⁴⁰⁾。さらに、2つのピークと線維化ステージ、ウイルスGenotypeを組み合わせた判別法で、AUROC 0.75の治療効果予測が可能であったと述べている。このように、プロテオミクスで得られた結果は、肝疾患の診断だけでなく、治療効果予測やその効果に関連するタンパク質の同定も期待できる。

むすび

固体や細胞レベルでのプロセスは、大部分がタンパク質で制御されており、病態や生体の反応を解析するには、タンパク質発現の変化を解析することが必須である。アルブミンや免疫グロブリンなど約20種類のタンパク質が血清中総タンパクの99%を占め、残り1%の微量タンパク質の中に、バイオマーカーとなりうるタンパク質が存在している。今後、プロテオミクス技術がさらに向上し、微量なバイオマーカー探索に有用で、より感度の高い検出系が開発され、あらゆる肝疾患における有用な血清マーカーの発見が期待される。また、本稿では血清プロテオミクスを中心に述べたが、最近、HCC組織を2-DEで解析し、HCC組織中のclathrin heavy chainの増加とformiminotransferase cyclodeaminaseの低下が早期HCCの診断に非常に有用であること⁴²⁾、切除HCCのAPC-binding protein EB1発現は、HCC切除後の生存率と再発率に関係することが報告されている⁴³⁾。HCCにおける有用な組織マーカーの報告も今後さらに増加すると考えられ、用いる試料や解析手法の利点・欠点を十分吟味した解析が必要と考えられる。

文献

- 1) Oka H, Tamori A, Kuroki T, et al. Prospective study of alpha-fetoprotein in cirrhotic patients monitored for development of hepatocellular carcinoma. *Hepatology*. 1994; 19(1): 61-6.

- 2) Ishii M, Gama H, Chida N, et al. Simultaneous measurements of serum alpha-fetoprotein and protein induced by vitamin K absence for detecting hepatocellular carcinoma. South Tohoku District Study Group. *Am J Gastroenterol.* 2000; 95(4): 1036-40.
- 3) Okuda H, Nakanishi T, Takatsu K, et al. Serum levels of des-gamma-carboxy prothrombin measured using the revised enzyme immunoassay kit with increased sensitivity in relation to clinicopathologic features of solitary hepatocellular carcinoma. *Cancer.* 2000; 88(3): 544-9.
- 4) Marrero JA, Su GL, Wei W, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in american patients. *Hepatology.* 2003; 37(5): 1114-21.
- 5) Ikoma J, Kaito M, Ishihara T, et al. Early diagnosis of hepatocellular carcinoma using a sensitive assay for serum des-gamma-carboxy prothrombin: a prospective study. *Hepato-gastroenterology.* 2002; 49(43): 235-8.
- 6) Toyoda H, Kumada T, Kiriya S, et al. Prognostic significance of simultaneous measurement of three tumor markers in patients with hepatocellular carcinoma. *Clin Gastroenterol Hepatol.* 2006; 4(1): 111-7.
- 7) Nakamura S, Nouse K, Sakaguchi K, et al. Sensitivity and specificity of des-gamma-carboxy prothrombin for diagnosis of patients with hepatocellular carcinomas varies according to tumor size. *Am J Gastroenterol.* 2006; 101(9): 2038-43.
- 8) Morra R, Munteanu M, Bedossa P, et al. Diagnostic value of serum protein profiling by SELDI-TOF ProteinChip compared with a biochemical marker, FibroTest, for the diagnosis of advanced fibrosis in patients with chronic hepatitis C. *Aliment Pharmacol Ther.* 2007; 26(6): 847-58.
- 9) Gangadharan B, Antrobus R, Dwek RA, et al. Serum biomarker candidates for liver fibrosis in hepatitis C patients. *Clin Chem.* 2007; 53(10): 1792-9.
- 10) Nugent C, Younossi ZM. Evaluation and management of obesity-related nonalcoholic fatty liver disease. *Nat Clin Pract Gastroenterol Hepatol.* 2007; 4(8): 432-41.
- 11) Younossi ZM, Baranova A, Ziegler K, et al. A genomic and proteomic study of the spectrum of nonalcoholic fatty liver disease. *Hepatology.* 2005; 42(3): 665-74.
- 12) Espina VA, Edmiston KH, Heiby M, et al. A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. *Mol Cell Proteomics.* 2008; 7(10): 1998-2018.
- 13) Kriegsheim A, Preisinger C, Kolch W. Mapping of signaling pathways by functional interaction proteomics. *Methods Mol Biol.* 2008; 484: 177-92.
- 14) Kuramitsu Y, Nakamura K. Proteomic analysis of cancer tissues: shedding light on carcinogenesis and possible biomarkers. *Proteomics.* 2006; 6(20): 5650-61.
- 15) Lee IN, Chen CH, Sheu JC, et al. Identification of human hepatocellular carcinoma-related biomarkers by two-dimensional difference gel electrophoresis and mass spectrometry. *J Proteome Res.* 2005; 4(6): 2062-9.
- 16) Melle C, Ernst G, Scheibner O, et al. Identification of specific protein markers in microdissected hepatocellular carcinoma. *J Proteome Res.* 2007; 6(1): 306-15.
- 17) Tiss A, Smith C, Camuzeaux S, et al. Serum peptide profiling using MALDI mass spectrometry: avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology. *Proteomics.* 2007; 7 Suppl 1: 77-89.
- 18) Shoshan SH, Admon A. Novel technologies for cancer biomarker discovery: humoral proteomics. *Cancer Biomark.* 2007; 3(3): 141-52.
- 19) Yang MH, Tyan YC, Jong SB, et al. Identification of human hepatocellular carcinoma-related proteins by proteomic approaches. *Anal Bioanal Chem.* 2007; 388(3): 637-43.
- 20) Ang IL, Poon TC, Lai PB, et al. Study of serum haptoglobin and its glycoforms in the diagnosis of hepatocellular carcinoma: a glycoproteomic approach. *J Proteome Res.* 2006; 5(10): 2691-700.
- 21) Feng JT, Liu YK, Song HY, et al. Heat-shock protein 27: a potential biomarker for hepatocellular carcinoma identified by serum proteome analysis. *Proteomics.* 2005; 5(17):

- 4581-8.
- 22) Zinkin NT, Grall F, Bhaskar K, et al. Serum proteomics and biomarkers in hepatocellular carcinoma and chronic liver disease. *Clin Cancer Res.* 2008; 14(2): 470-7.
 - 23) He QY, Zhu R, Lei T, et al. Toward the proteomic identification of biomarkers for the prediction of HBV related hepatocellular carcinoma. *J Cell Biochem.* 2008; 103(3): 740-52.
 - 24) Kanmura S, Uto H, Kusumoto K, et al. Early diagnostic potential for hepatocellular carcinoma using the SELDI ProteinChip System. *Hepatology.* 2007; 45(4): 948-56.
 - 25) Cui J, Kang X, Dai Z, et al. Prediction of chronic hepatitis B, liver cirrhosis and hepatocellular carcinoma by SELDI-based serum decision tree classification. *J Cancer Res Clin Oncol.* 2007; 133(11): 825-34.
 - 26) Göbel T, Vorderwülbecke S, Hauck K, et al. New multi protein patterns differentiate liver fibrosis stages and hepatocellular carcinoma in chronic hepatitis C serum samples. *World J Gastroenterol.* 2006; 12(47): 7604-12.
 - 27) Ward DG, Cheng Y, N'Kontchou G, et al. Changes in the serum proteome associated with the development of hepatocellular carcinoma in hepatitis C-related cirrhosis. *Br J Can.* 2006; 94(2): 287-93.
 - 28) Lee IN, Chen CH, Sheu JC, et al. Identification of complement C3a as a candidate biomarker in human chronic hepatitis C and HCV-related hepatocellular carcinoma using a proteomics approach. *Proteomics.* 2006; 6(9): 2865-73.
 - 29) Geng X, Wang F, Li YG, et al. SELDI-TOF MS proteinchip technology for screening of serum markers of HBV-induced hepatocellular carcinoma. *J Exp Clin Cancer Res.* 2007; 26(4): 505-8.
 - 30) He M, Qin J, Ahai R, et al. Detection and identification of NAP-2 as biomarker in hepatitis B-related hepatocellular carcinoma by proteomic approach. *Proteome Science.* 2008; 6: 10.
 - 31) Ward DG, Cheng Y, N'Kontchou G, et al. Preclinical and post-treatment changes in the HCC-associated serum proteome. *Br J Cancer.* 2006; 95(10): 1379-83.
 - 32) Poon TCW, Hui A, Chan HLY, et al. Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: A pilot study. *Clinical Chemistry.* 2005; 51(2): 328-35.
 - 33) Van Greevenbroek MM, Vermeulen VM, De Bruin TW. Identification of novel molecular candidates for fatty liver in the hyperlipidemic mouse model, HcB19. *J Lipid Res.* 2004; 45: 1148-54.
 - 34) Uto H, Sato Y, Ishida Y, et al. Proteomic analysis of serum biomarkers in patients with nonalcoholic steatohepatitis using SELDI-TOF/MS or MALDI-TOF/MS. *Hepatology.* 2007; 46 (Supple): 735A.
 - 35) Orvisky E, Drake SK, Martin BM, et al. Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma. *Proteomics.* 2006; 6(9): 2895-902.
 - 36) Goldmann R, Ressom HW, Abdel-Hamid M, et al. Candidate markers for the detection of hepatocellular carcinoma in low-molecular weight fraction of serum. *Carcinogenesis.* 2007; 28(10): 2149-53.
 - 37) Kam RKT, Poon TCW, Chan HLY, et al. High-throughput quantitative profiling of serum N-glycome by MALDI-TOF mass spectrometry and N-glycomic fingerprint of liver fibrosis. *Clinical Chemistry.* 2007; 53(7): 1254-63.
 - 38) Resson HW, Varghese RS, Goldman L, et al. Analysis of MALDI-TOF mass spectrometry data for discovery of peptide and glycan biomarkers of hepatocellular carcinoma. *J Proteome Res.* 2008; 7(2): 603-10.
 - 39) Comunale MA, Lowman M, Long RE, et al. Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. *J Proteome Res.* 2006; 5(2): 308-15.
 - 40) Paradis V, Asselah T, Dargere D, et al. Serum proteome to predict virologic response in patients with hepatitis C treated by pegylated interferon plus ribavirin. *Gastroenterology.* 2006; 130(7): 2189-97.
 - 41) Miyazaki T, Ohura T, Kobayashi M, et al. Fatal propionic acidemia in mice lacking propionyl-CoA carboxylase and its rescue by postnatal,

- liver-specific supplementation via a transgene. *J Biol Chem.* 2001; 276: 35995-9.
- 42) Seimiya M, Tomonaga T, Matsushita K, et al. Identification of novel immunohistochemical tumor markers for primary hepatocellular carcinoma; clathrin heavy chain and formiminotransferase cyclodeaminase. *Hepatology.* 2008; 48(2): 519-30.
- 43) Orimo T, Ojima H, Hiraoka N, et al. Proteomic profiling reveals the prognostic value of APC-binding protein EB1 in hepatocellular carcinoma. *Hepatology.* 2008(in press).

Parameter Optimization in Network Dynamics Including Unmeasured Variables by the Symbolic-numeric Approach*

Masahiko Nakatsui¹

Katsuhisa Horimoto^{1,2}

¹Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan

²Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Abstract In this report, we propose a new symbolic-numeric method of differential algebra and a numerical parameter optimization algorithm. First, we utilize differential elimination, which is an algebraic approach for rewriting a system of differential equations into another equivalent system, to derive the constraints between the kinetic parameters from the original system. Second, we introduce these constraints to effectively optimize the parameters into a genetic algorithm, Real-Coded Genetic Algorithms (RCGAs), which is a numerical parameter optimizing method. To evaluate the ability of our method, we performed a simulation study for an artificial biological network including one measured and three unmeasured molecules. As a result, our method, the symbolic-numeric method of differential elimination and RCGAs, precisely estimated the kinetic parameters in the simulated network, while RCGAs failed. Thus, our method is useful for analyzing the dynamics of a biological network including unmeasured molecules.

Keywords symbolic-numeric method, differential elimination, real-coded genetic algorithms (RCGAs)

1 Introduction

The investigation of network dynamics is a major issue in systems biology [1]. A network model for describing the kinetics of constituent molecules is usually first constructed with reference to the biological knowledge, and then the model is mathematically expressed by differential equations, based on the chemical reactions underlying the kinetics. Finally, the kinetic parameters in the model are estimated by various parameter optimization techniques [2], from the time series data measured for the constituent molecules. In the last stage, we cannot always obtain the data measured for all of the constituent molecules, due to limitations of measurement techniques and ethical constraints. Thus, one of the issues we should resolve is that the parameters are estimated from the data for only some of the constituent molecules. Unfortunately, it is frequently difficult to estimate the parameters in such a network model including unmeasured variables.

*This work was partly supported by a project grant, 'Development of Analysis Technology for Gene Functions with Cell Arrays', from The New Energy and Industrial Technology Development Organization (NEDO).

Recently, Boulier and his colleagues developed differential elimination [3], derived from the Rosenfeld-Gröbner base [4], and applied it to solve the issue of network dynamics including unmeasured variables [3, 5]. In their application, the equations rewritten by differential elimination from the system of differential equations were utilized to estimate the initial values for the parameter optimization, by Newton-type numerical optimization.

Here, we propose a new method for optimizing the parameters in a network including unmeasured variables, in combination with a genetic algorithm [6, 7] and differential elimination [3]. Our method partially utilizes a technique from a previous study [3] about the introduction of differential elimination into the parameter optimization in a network including unmeasured variables. Instead of using differential elimination for estimating the initial values for the following parameter optimization, the equations reduced by differential elimination are directly introduced as the constraints into the genetic algorithm for the parameter optimization. To validate the effectiveness of the constraint introduction, we performed a simulation where we assumed that a network was composed of four molecules, and the data for only one molecule among them were measured. The parameter values estimated by our method were compared with those generated by the genetic algorithm without the constraints.

2 Results and Discussion

2.1 Analyzed Model

We prepared a network model containing four molecules, as shown in Fig. 1 (A). The differential equation representation of Fig. 1 is shown in eq. 1.

$$\begin{aligned} \frac{dx_1(t)}{dt} &= k_{21}x_2 + k_{31}x_3 + k_{41}x_4 - k_{e1}x_1(t) \\ \frac{dx_2(t)}{dt} &= -k_{e2}x_2 - k_{21}x_2 \\ \frac{dx_3(t)}{dt} &= -k_{e2}x_3 - k_{31}x_3 \\ \frac{dx_4(t)}{dt} &= -k_{e2}x_4 - k_{41}x_4 \end{aligned} \quad (1)$$

In the model, we assume that the molecules, x_2 , x_3 , and x_4 , activate x_1 , with linear relationships. Notably, only the amount of molecule x_1 among the four molecules is measured.

We generated the reference curve of x_1 for estimating the kinetic parameter set under the following initial conditions for each molecule and kinetic constant: $x_1(0) = 10.0, x_2(0) = 130.0, x_3(0) = 80.0, x_4(0) = 170.0, k_{21} = 0.01, k_{31} = 0.1, k_{41} = 10.0, k_{e1} = 5.0, k_{e2} = 3.0$. The generated reference curve ($0 \leq t \leq 1$ at 0.01 intervals) is shown in Fig. 1 (B).

2.2 Constraints from Differential Elimination

We derived the constraint equations from differential equations of the analyzed network model (eq. 1), by using differential elimination (see details in 3.2).

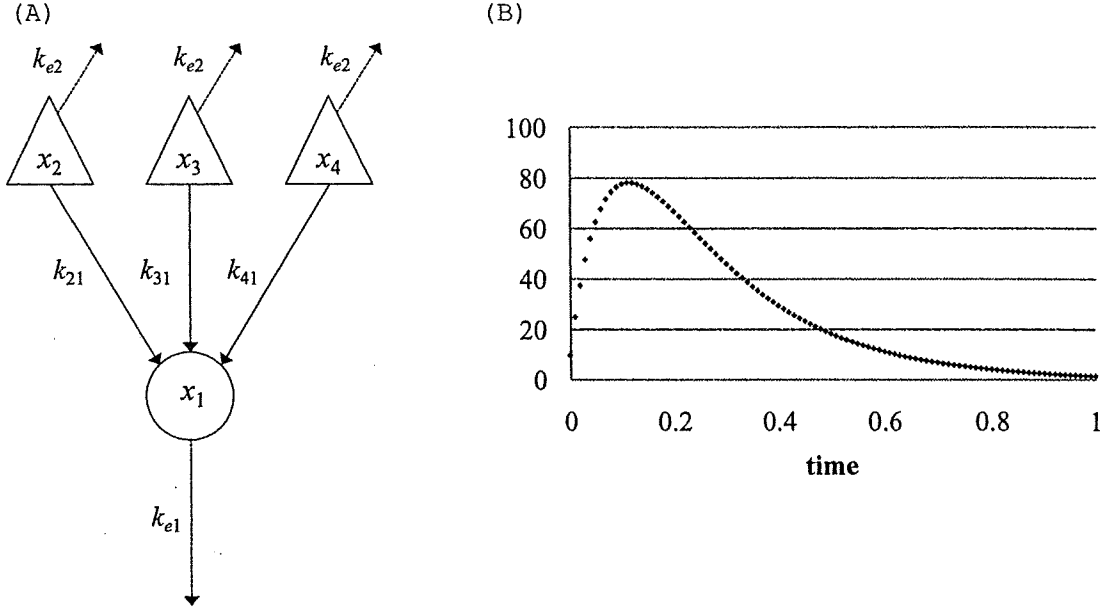


Figure 1: Analyzed model and reference curve

 (A) Analyzed model. (B) Reference curve of x_1 generated by given parameter values.

$$\begin{aligned}
 C_{1,t} = & \frac{1.0}{k_{21}(k_{21}-k_{31})(k_{21}-k_{41})} \left(\frac{d^3}{dt^3} x_1(t) + (k_{31} + k_{41} + k_{e1} + 2k_{e2}) \frac{d^2}{dt^2} x_1(t) \right. \\
 & + (k_{31}k_{41} + k_{31}k_{e1} + k_{41}k_{e1} + k_{31}k_{e2} + k_{41}k_{e2} + 2k_{e1}k_{e2} + k_{e2}^2) \frac{d}{dt} x_1(t) \\
 & \left. + k_{e1}(k_{31} + k_{e2})(k_{41} + k_{e2})x_1(t) - k_{21}(k_{21} - k_{31})(k_{21} - k_{41})x_2(t) \right) = 0
 \end{aligned} \quad (2)$$

$$\begin{aligned}
 C_{2,t} = & \frac{1.0}{(k_{21} - k_{31})k_{31}(k_{31} - k_{41})} \left(\frac{d^3}{dt^3} x_1(t) + (k_{21} + k_{41} + k_{e1} + 2k_{e2}) \frac{d^2}{dt^2} x_1(t) \right. \\
 & + (k_{41}(k_{e1} + k_{e2}) + k_{21}(k_{41} + k_{e1} + k_{e2}) + k_{e2}(2k_{e1} + k_{e2})) \frac{d}{dt} x_1(t) \\
 & \left. + k_{e1}(k_{21} + k_{e2})(k_{41} + k_{e2})x_1(t) + x_3(t) \right) = 0
 \end{aligned} \quad (3)$$

$$\begin{aligned}
 C_{3,t} = & \frac{1.0}{(k_{21} - k_{41})(k_{31} - k_{41})k_{41}} \left(\frac{d^3}{dt^3} x_1(t) + (k_{21} + k_{31} + k_{e1} + 2k_{e2}) \frac{d^2}{dt^2} x_1(t) \right. \\
 & + (k_{21}k_{31} + k_{21}k_{e1} + k_{31}k_{e1} + k_{21}k_{e2} + k_{31}k_{e2} + 2k_{e1}k_{e2} + k_{e2}^2) \frac{d}{dt} x_1(t) \\
 & \left. + k_{e1}(k_{21} + k_{e2})(k_{31} + k_{e2})x_1(t) + (k_{21} - k_{41})k_{41}(-k_{31} + k_{41})x_4(t) \right) = 0
 \end{aligned} \quad (4)$$

$$\begin{aligned}
C_{4,t} = & \frac{d^4}{dt^4}x_1(t) + (k_{21} + k_{31} + k_{41} + k_{e1} + 3k_{e2})\frac{d^3}{dt^3}x_1(t) \\
& + (k_{21}k_{31} + k_{21}k_{41} + k_{31}k_{41} + k_{21}k_{e1} + k_{31}k_{e1} + k_{41}k_{e1} \\
& + 2k_{21}k_{e2} + 2k_{31}k_{e2} + 2k_{41}k_{e2} + 3k_{e1}k_{e2} + 3k_{e2}^2)\frac{d^2}{dt^2}x_1(t) \\
& + (k_{31}(k_{41}(k_{e1} + k_{e2}) + k_{e2}(2k_{e1} + k_{e2})) + k_{21}(k_{41}(k_{e1} + k_{e2}) + k_{31}(k_{41} + k_{e1} + k_{e2}) \\
& + k_{e2}(2k_{e1} + k_{e2})) + k_{e2}(k_{41}(2k_{e1} + k_{e2}) + k_{e2}(3k_{e1} + k_{e2})))\frac{d}{dt}x_1(t) \\
& + k_{e1}(k_{21} + k_{e2})(k_{31} + k_{e2})(k_{41} + k_{e2})x_1(t) = 0
\end{aligned} \tag{5}$$

In the above four equations, $C_{4,t}$ is an equation of x_1 , its derivatives, and the parameters, and the three remaining constraints are equations of x_1 , its derivatives, the parameters, and either x_2 , x_3 , or x_4 . Since the solutions of x_2 , x_3 , and x_4 can be analytically obtained from eq. 1, we can numerically calculate the four constraint values by using the parameter values. The above constraints (eq. 2 - eq. 5) are introduced into the genetic algorithm as the constraints for parameter estimation, as described below.

2.3 Objective Function in Symbolic-Numeric Method

In our study, the objective function is composed of two terms: one is the standard error function between the estimated and sample data, and the other is the constraints (eq. 2 - 5) obtained by differential elimination (DE constraints). The error function is defined as follows: Suppose that $x_{1,t}^c$ is the time-course data at time t of x_1 calculated by using the estimated parameter values, and $x_{1,t}^s$ represents the sampling data (reference curve) at time t of x_1 . The sum of the absolute value of the relative error between $x_{1,t}^c$ and $x_{1,t}^s$ gives the total relative error, E .

$$E = \sum_{t=1}^T \left| \frac{x_{1,t}^c - x_{1,t}^s}{x_{1,t}^s} \right| \tag{6}$$

As usual, the above error function is an objective function for RCGAs (OF_{RCGAs}), i.e.,

$$\text{OF}_{\text{RCGAs}} = E. \tag{7}$$

Next we obtain the DE constraints as the linear combination of eqs. 2 - 5, as follows:

$$C_{\text{DE}} = \sum_{l=1}^L \sum_{t=1}^T |C_{l,t}| \tag{8}$$

where $L = 4$ and $T = 100$.

Finally, we introduce the DE constraints into the objective function of the RCGAs: we defined the objective functions for our symbolic-numeric method (OF_{SN}) as:

$$\text{OF}_{\text{SN}} = \alpha \text{OF}_{\text{RCGAs}} + (1 - \alpha) C_{\text{DE}} \tag{9}$$

where $\alpha = 0.99975$. The computational task is to determine a set of parameter values that minimize to OF_{SN} (see details in 3.3).

2.4 Estimation of Kinetic Parameters

We performed RCGAs and our symbolic-numeric method to estimate the kinetic parameter set (k_{21} , k_{31} , and k_{41}) 200 times from the data in Fig. 1 (B). We successfully obtained 132 parameter sets by RCGAs and 90 sets by the symbolic-numeric method: we regarded the outcome as a success, when E/T in 2.3 satisfies less than 0.01 until 20,000 generations are reached. The number of successful estimations by RCGAs was larger than that by the symbolic-numeric method. This is because the constraints in the symbolic-numeric method might strongly affect the parameter estimation.

The histograms of the estimated values for the three parameters are shown in Fig. 2. As seen in the figures, the three parameters were correctly estimated by the symbolic-numeric method, while all of the estimations by RCGAs failed. Figure 2 (A) shows the histogram of the k_{21} values. The most frequent values estimated by the symbolic-numeric method were found in the bin corresponding to the range from $0.005 < \text{estimated } k_{21} \leq 0.015$, which included the correct value, 0.01. In addition, the estimated values were concentrated around the correct value. In contrast, although the most frequent values generated by RCGAs were also found near the correct value, the estimated values were distributed uniformly in the range from 0.01 to 0.05. A similar situation is shown in Fig. 2 (C). The most frequent values estimated by the symbolic-numeric method were found in the bin including the correct value, 10.0, while the values obtained by RCGAs were widely distributed. In Fig. 2 (B), our symbolic-numeric method correctly estimated the parameter values ($k_{31} = 0.1$), but RCGAs definitely failed. In summary, the parameter values estimated by our method were narrowly distributed in the bin including the correct values in all cases, but those generated by RCGAs were widely distributed.

2.5 Further Remarks

Our symbolic-numeric method correctly estimated the parameter values from the measured data of one molecule, while by RCGAs failed. This clearly indicates that the ability of our method for estimating the parameter values was far superior to that of RCGAs. Although the present study focused on a simple model, our method is a feasible approach for parameter estimation in network dynamics including unmeasured variables.

Finally, we discuss further application possibilities of our method. As indicated in 3.2, the differential elimination is not dependent on linearity, but shows its power in the nonlinear system. By the differential elimination, x_2 in the first nonlinear equation of eq. 10 in 3.2 can be eliminated in the first equation of eq. 11. Thus, our method has different types of differential equations in its application. Another possibility of our method is application to network inference without known structure. Since the present method is designed with the assumption of known network structure, the application range of our method to network inference is naturally restricted. However, our method can select the most possible network structure among the networks with similar structures. Indeed, we designed a similar procedure for evaluating the network structures with the measured data [8]. In our previous approach, we adopted the transformation of a system of differential equations into the equivalent system of algebraic equations by Laplace transformation. In this case, the system must be linear, due to the Laplace transformation. Furthermore, the numeric optimization in the previous approach frequently faces difficulties, due to the existence of the pole in the Laplace domain. In contrast, the above pitfalls are overcome in

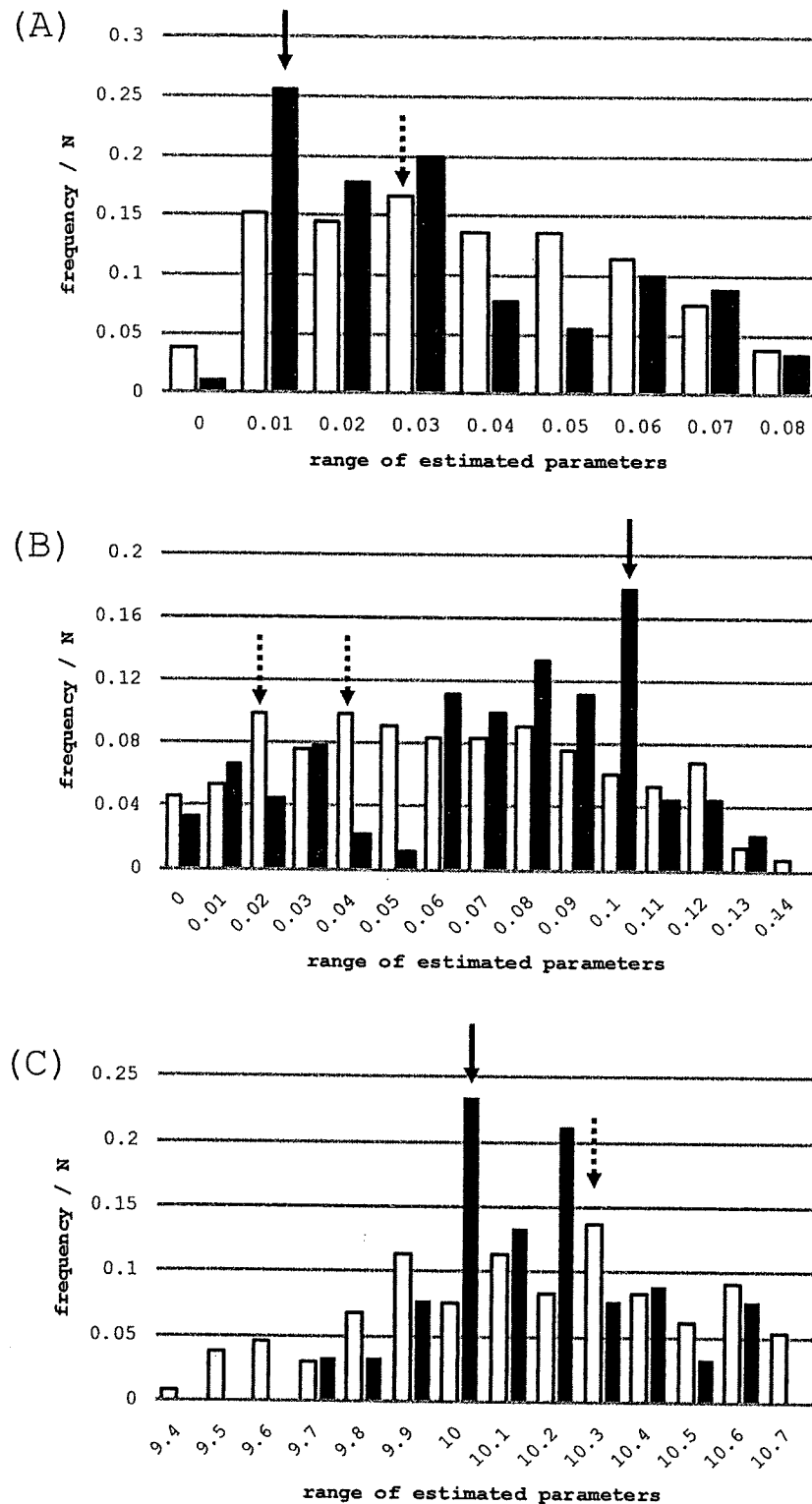


Figure 2: Histograms of estimated parameters. (A) Histogram of estimated k_{21} . (B) Histogram of estimated k_{31} . (C) Histogram of estimated k_{41} . In these histograms, the abscissa represents the bin. The bin range for (A) and (B) is 0.01, and that for (C) is 0.1. The ordinate shows the ratio of frequency (frequency divided by the number of estimated parameter sets (N)). N for RCGAs was 132, and that for the symbolic-numeric method was 90). Black bars represent the ratio of frequency for the symbolic-numeric method. White bars represent that for the RCGAs.

the present method, by the combination of genetic algorithm and differential elimination. In the near future, the application of the present method to various forms of networks will be presented to further investigate the merits and pitfalls of our method, in terms of parameter optimization and model selection.

3 Methods

3.1 Real-Coded Genetic Algorithms

The Genetic Algorithm is a well-known parameter optimization framework, which is inspired by the evolutionary process of biology [6, 7]. We have applied an efficient computational technique based on RCGAs as a nonlinear numerical optimization method [9], by the combination of *unimodal normal distribution crossover* (UNDX) [10] and *minimal generation gap* (MGG) [11]. The generation-alternations are repeated until either the value of the objective function E becomes less than a given threshold (we called this threshold the error allowance on RCGAs) or the number of generation-alternation iterations reaches a given threshold of maximum generation counts.

The general procedure for a typical or conventional GA is as follows:

1. *Generation of Initial Population* Generate initial population randomly.
2. *Selection for Reproduction* Select individuals in population for the *Reproduction* step according to the generation-alternation model.
3. *Reproduction* Generate offspring by the crossover operator from the parent individuals selected in step 2.
4. *Selection* Evaluate the fitness of each individual and select individuals for survival (see also 2.3).
5. *Repeat the procedure between steps 2 to 4 until a certain condition for termination is satisfied.*

3.2 Differential Elimination

Differential algebra aims at studying differential equations from a purely algebraic point of view [12, 13]. The differential elimination theory is a sub theory of differential algebra [3], based on Rosenfeld-Gröbner [4]. The differential elimination process rewrites the inputted system of differential equations to another equivalent system according to ranking (order of terms). Here, we provide an example of the differential elimination process, as shown below, according to Boulier [3, 5].

Assume a system of parametric ordinary differential equations with two variables, x_1 and x_2 , and then build the following equations,

$$\begin{aligned} \dot{x}_1 &= -k_{12}x_1 + k_{21}x_2 - \frac{V_e x_1}{k_e + x_1} \\ \dot{x}_2 &= k_{12}x_1 - k_{21}x_2 \end{aligned} \quad (10)$$

where k_{12} , k_{21} , k_e and V_e are some constants. The differential elimination, then produces the following two equations equivalent to the above system.

$$\begin{aligned} \dot{x}_1(x_1 + k_e)^2 + (k_{12} + k_{21})\dot{x}_1(x_1 + k_e)^2 + V_e \dot{x}_1 k_e + k_{21} V_e x_1(x_1 + x_e) &= 0 \\ \dot{x}_1(k_e + x_1) + k_{21}x_1^2 + (k_{12} + V_e)x_1 - k_{21}(k_e + x_1)x_2 &= 0 \end{aligned} \quad (11)$$

The first equation in eq. 11 is composed of x_1 , its derivatives, and the parameters obtained by eliminating x_2 , and the second one is composed of x_1 , its derivatives, the parameters, and x_2 . Note that x_2 can be expressed by x_1 , its derivatives, and the parameters in the second equation.

All of the symbolic computations for the differential elimination were performed using the *difalg* package of MAPLE 10.

3.3 Combination of RCGAs and Differential Elimination

In general, the typical objective function for evaluating the reproducibility of an experimentally observed time-series for a parameter set is the total relative error, E , such as eq. 6. The parameter set is then estimated when the total relative error falls below a given threshold. However, the immense searching space of parameter values frequently prohibits the correct parameter estimation. To overcome this problem, we introduce the constraints derived by differential elimination into the objective function. In the present study, our strategy is to reduce the immense searching space, by introducing constraints between the estimated parameters derived from differential elimination, C , into the objective function for the Real-Coded Genetic Algorithm, which is a well-known heuristic numerical optimizer, i.e.,

$$\text{ObjectiveFunction} = \alpha E + (1 - \alpha)C \quad (12)$$

where α is a weighting factor in the objective function. Here, α was defined such that E and C were equally weighted.

References

- [1] Kitano, H.: System Biology: A Brief Overview, *Science* Vol. 295, no. 5560, March (2002) 1662 - 1664.
- [2] Nocedal, J., Wright, S. J.: Numerical Optimization, *Springer-Verlag, NY*, (1999).
- [3] Boulier, F.: Differential Elimination and Biological Modeling, *Johann Radon Institute for Computational and Applied Mathematics (RICAM) Book Series*, Vol. 2, (2007) 111-139.
- [4] Boulier, F., Lazard, D., Ollivier, F., Petitot, M.: Representation for the radical of a finitely generated differential ideal, *Proceedings of ISSAC 1995*, (1995), 158-166.
- [5] Boulier, F.: Differential algebra and system modeling in cellular biology, *Proceedings of Algebraic Biology 2008*, (2008) 22-39.
- [6] Holland, J. H.: Adaptation in Natural and Artificial Systems, *The University of Michigan Press, Ann Arbor, MI* (1975).
- [7] Goldberg, D. D.: Genetic Algorithms in Search, Optimization and Machine Learning, *Addison-Wesley Longman Publishing Co., Inc., Boston, MA*, (1989).
- [8] Yoshida, H., Nakagawa, K., Anai, H., and Horimoto, K.: An Algebraic-Numeric Algorithm for the Model Selection in Kinetic Networks, *Proceedings of the 10th CASC, LNCS 4770*, (2007) 433-447.
- [9] Jonikow, G. Z. and Michalewicz, Z.: An Experimental Comparison of Binary and Floating Point Representations in Genetic Algorithms, *Proceedings of the Fourth International Conference on Genetic Algorithms*, (1991) 31-36.
- [10] Ono, I. and Kobayashi, S.: A real-coded genetic algorithm for function optimization using unimodal distribution crossover, *Proc 7th ICGA*, (1997) 249-253.