

# 第 35 回構造活性相関シンポジウム

## 講演要旨集

会 期 2007 年 11 月 15 日 (木)・16 日 (金)

会 場 京都大学 百周年時計台記念館

主 催 日本薬学会構造活性相関部会

共 催 日本化学会

日本農芸化学会

日本分析化学会

日本農薬学会

## 第 35 回構造活性相関シンポジウム

## 第 1 日(11 月 15 日)

10:25 - 10:30 開会 (京大院・薬)藤井 信孝

共通セッション(会場:百周年記念ホール)

座長: 山下 富義

10:30 - 10:50 JK01 活性構造知識ベース構築とそのソフトウェア基盤  
(関学大・理工)○大森 紀人, (豊橋技科大・工)藤島 悟志, (関学大・理工)  
森 幸雄, 堀川 裕志, 山川 眞透, 岡田 孝.....21

10:50 - 11:10 JK02 100 クラス SVM による薬物の活性クラス分類  
(豊橋技科大・工)○河合 健太郎, 藤島 悟志, 高橋 由雅.....23

座長: 河合 隆利

11:10 - 11:30 JK03 PLS アルゴリズムを用いたノンパラメトリック回帰手法の改良  
(大阪大谷大・薬, 阪大・薬, 阪大院・薬, 田辺製薬, 阪大・RI センター, 阪大・遺伝  
情報実験センター, 阪大・タイ感染症研究センター)○錦織 理華, 米倉 聡,  
岡本 晃典, 大軽 貴典, 松浦 晶子, 越智 雪乃, 森本 正太郎, 斉藤 直,  
川下 理日人, 安永 照雄, 川瀬 雅也, 高木 達也.....27

11:30 - 11:50 JK04 Generative topographic mapping (GTM) によるドラッグライクネス評価モデルの構築  
(東大院・工, 東大・工)○荒川 正幹, 宮尾 知幸, 尾崎 由隆, 青木 宏,  
船津 公人.....29

12:00 - 13:00 情報化学部会・構造活性相関部会 合同幹事会(会場:会議室Ⅲ)

特別講演(会場:百周年記念ホール)

座長: 西岡 孝明

13:15 - 14:15 JS FT-ICR/MS でみる悉皆的生物代謝  
(大阪府大院・生命環境)太田 大策.....3

ポスターセッション(会場:国際交流ホールⅡ・Ⅲ)

14:15 - 16:45 KP01~KP33

14:15 - 15:15 奇数番発表

15:15 - 16:15 偶数番発表

特別講演(会場:百周年記念ホール)

座長: 藤井 信孝

16:45 - 17:45 KS Drug Discovery and Development in the Kinase Inhibitory Field with Nested  
Chemical Library<sup>TM</sup>  
(Peptide Biochemistry Res. Group of Hung. Acad. Sci. and Semmelweis University)  
György Kéri.....1

17:45 - 18:00 ポスター賞表彰式(情報化学討論会と合同、会場:百周年記念ホール)

18:30 - 20:30 懇親会(情報化学討論会と合同、会場:芝蘭会館山内ホール)

KP13	セフェム系抗生物質のラット脳脊髄腔液移行性に関する QSAR 解析 (徳島大院・薬)○吉田 麻衣, 坂和 園子, 阿部 寛, 木原 勝, 山内 あい子, (大塚製薬)内藤 真策 .....	103
KP14	アゾール系化合物の CYP2B/3A との結合および酵素活性阻害様式の解析 (徳島大院・薬)○糸川 大祐, 山内 あい子, 中馬 寛.....	105
KP15	化学物質の 28 日間反復投与試験のデータ解析 (製品評価技術基盤機構)○櫻谷 祐企, 佐藤 佐和子, 張 慧琪, 西川 智, 山田 隼, 前川 昭彦 .....	109
KP16	計算化学的手法を用いた新規 $\beta$ -セクレターゼ阻害剤の探索 (北里大院・薬)○藤本 拓, 松下 泰雄, 合田 浩明, 広野 修一.....	111
KP17	Structure-based Drug Design における仮想的水和リガンドを用いたドッキングスタディの有用性 (北里大・薬)○弦間 格, 山乙 教之, 酒匂 佑介, 合田 浩明, 広野 修一.....	115
KP18	X 線構造解析及び熱力学的解析に基づく CK2 $\alpha$ 阻害剤の構造活性相関 (大阪府大院・理)○仲庭 哲津子, (大阪府大・生命)深田 はるみ, (大阪府大院・理) 関口 雄介, 木下 誉富, (京大院・薬)仲西 功, 北浦 和夫, 平澤 明, 辻本 豪三, (大阪府大院・理)多田 俊治 .....	119
KP19	Protein Data Bank に蓄積されたリガンド-蛋白質相互作用情報のデータベース化に向けて ~リガンド認識様式の整理 (長浜バイオ大)○齊藤 美保子, 白井 剛 .....	121
KP20	Virtual screening による UCH-L3 の新規阻害剤同定 (早大院・先進理工, NCNP 神経研)○平山 和徳, (NCNP 神経研)青木 俊介, 西川 香里, (早大院・先進理工)松本 隆, (NCNP 神経研)和田 圭司.....	123
KP21	FAMSmulti: 複数の鋳型タンパク質を用いた全自動ホモロジーモデリング (北里大・薬)○加納 和彦, 岩館 満雄, 寺師 玄記, 竹田-志鷹 真由子, 高谷 大輔, 酒井 博子, 梅山 秀明 .....	125
KP22	SKE-DOCK: 相補性を考慮したドッキングと最適構造評価法による、タンパク質複合体立体 構造予測法の開発 (北里大・薬)○寺師 玄記, 竹田-志鷹 真由子, 加納 和彦, 岩館 満雄, 高谷 大輔, 酒井 博子, 梅山 秀明 .....	127
KP23	fams-ace: CASP7 におけるタンパク質立体構造の再構築と、最適構造の評価法の開発 (北里大・薬)○竹田-志鷹 真由子, 寺師 玄記, 加納 和彦, 岩館 満雄, 高谷 大輔, 酒井 博子, 梅山 秀明 .....	129

KP24	エイコサノイドとグルタチオンの代謝に関係する一群の3量体膜タンパク質の網羅的モデリング:MAPEG superfamily complex modeling (北里大・薬)○加納 和彦, 竹田-志鷹 真由子, 岩館 満雄, 寺師 玄記, 高谷 大輔, 酒井 博子, 梅山 秀明.....	131
KP25	IARC 分類に基づく発癌性化合物の解析と Support Vector Machine によるヒト発癌性予測 (徳島大院・薬)○竹崎 俊晶, 小林 進一, 坂本 久美子, 木原 勝, 山内 あい子.....	133
KP26	廃水処理施設流入水中の BOD 予測に関する研究 (豊橋技科大)○川添 智久, 吉田 雄, 木曾 祥秋, 高橋 由雅, (神鋼電機)坪井 孝緒, 日奔 勉, 平野 新一, 河村 博年.....	135
KP27	MST 部分木を利用した類似空間検索と薬物構造データベース検索への応用 (豊橋技科大)○熊谷 正雪, 藤島 悟志, 高橋 由雅.....	137
KP28	モチーフ構造適合度マップの構築 (関学大・理工)○大村 隆晴, (関学大・理工)岡田 孝, (豊橋技科大)藤島 悟志.....	141
KP29	セロトニンアゴニストの基本的活性構造抽出 (関学大・理工)大森 紀人, ○森 幸雄, 堀川 裕志, 山川 眞透, 岡田 孝.....	143
KP30	チトクローム P450 の基質および阻害剤選択性の情報化学的手法による予測 (徳島大院・薬)○西岡 大貴, 近藤 崇泰, 寺尾 直也, 中馬 寛.....	147
KP31	KY 法による Ames 試験データ解析と考察 (富士通, 阪大・臨床医工)○湯田 浩太郎, (富士通九州)Jose Martin Ciloy, 北島 正人.....	149
KP32	ドーパミンアンタゴニストの基本的活性構造抽出 (関学大・理工)大森 紀人, 森 幸雄, 堀川 裕志, ○山川 眞透, 岡田 孝.....	151
KP33	超偏極キセノンの肺機能診断薬としての利用:肺癌モデルマウスへの適用 (阪大院・医)○今井 宏彦, 伊藤 豪, 木村 敦臣, 藤原 英明.....	155

## 第2日(11月16日)

## 一般講演(会場:百周年記念ホール)

- 座長: 木村 敦臣
- 9:30 - 9:55    K01    環境中に存在する化学物質の人工脂質膜透過性—生態系に対する蓄積性および  
ヒト曝露量予測への応用—  
(京大院・農)○藤川 真章, 赤松 美紀,  
(田辺三菱製薬)中尾 和也, 清水 良.....33
- 9:55 - 10:10    K02    チトクローム P450 の基質および阻害剤選択性の情報化学的手法による予測  
(徳島大院・薬)○西岡 大貴, 近藤 崇泰, 寺尾 直也, 中馬 寛.....37
- 座長: 岩間 年治
- 10:10 - 10:25    K03    ファージディスプレイによるペプチドの多様性と Scaffold  
(中村学園大・薬膳科学研)  
○毛利 彩恵子, 内山 文昭, 千原 智美, 治京 玉記.....39
- 10:25 - 10:40    K04    キチナーゼBに対するペプチド性阻害剤 Argifin の相互作用解析とその誘導体の論  
理的分子設計  
(北里大・薬)○合田 浩明, 広野 修一.....41
- 10:40 - 10:55    K05    SYBYL Zap プログラムを用いた MM/PBSA 結合自由エネルギー計算によるドッキ  
ングポーズ選択の有効性  
(北里大・薬)○山乙 教之, (ワールドフュージョン)木村 敏郎,  
(北里大・薬)広野 修一.....45

## 招待講演(会場:百周年記念ホール)

座長: 広野 修一

- 11:00 - 11:40    K06    創薬バリューチェーンによるインシリコでの阻害剤探索と最適化に関する研究  
(阪大院・工)井上 豪.....7

## 11:50 - 13:20    構造活性相関幹事会(会場:会議室Ⅲ)

## 招待講演(会場:百周年記念ホール)

座長: 仲西 功

- 13:30 - 14:10    K07    多標的インシリコスクリーニングと化合物記述子としてのドッキングスコア  
(産総研)福西 快文.....11

## 一般講演(会場:百周年記念ホール)

- 座長: 辻下 英樹
- 14:10 - 14:35    K08    SWAP MODEL によるタンパク質の部分脱水とエネルギーの解析  
(京大院・薬)○村田 克美, 仲西 功, 北浦 和夫.....49
- 座長: 福島 千晶
- 14:35 - 15:00    K09    fams-ace: CASP7 におけるタンパク質立体構造の再構築と、最適構造の評価法の開発  
(北里大・薬)○竹田-志鷹 真由子, 寺師 玄記, 加納 和彦, 岩館 満雄,  
高谷 大輔, 酒井 博子, 梅山 秀明.....53

# K09

## fams-ace: CASP7 におけるタンパク質立体構造の再構築と、最適構造の評価法の開発

(北里大学・薬) ○竹田一志鷹 真由子, 寺師 玄記, 加納和彦, 岩館 満雄, 高谷 大輔, 酒井 博子, 梅山 秀明

### 1 はじめに

タンパク質立体構造予測コンテストである Critical Assessment of Protein Structure Prediction (CASP7, Pacific Grove, CA, 2006)<sup>1</sup>とは、実験解析による立体構造の情報がホールドされた状態で、アミノ酸配列情報のみが与えられ、その三次元構造の予測精度を競うものである。参加するチームは主に二つに分けられる。一つは全自動で予測を行う Server部門。もう一つは、サーバーが予測した結果を考慮して人間による手動操作による修正などを行う Human expert部門である。

2006年5月から開催されたタンパク質立体構造予測コンテストであるCAS7において、本研究室は立体構造予測部門(TS)で4つの立体構造予測サーバーチーム (FAMS, FAMSD, CIRCLE, FUNCTION) と5つのHuman expertチーム (CHIMERA, fams-ace, CIRCLE-FAMS, fams-multi, Ligand-CIRCLE) で参加した。特に、fams-aceにおいては、当研究室の梅山秀明教授がCAS7国際会議でスコア上位者として招待講演を行った。<sup>2,3</sup>

ここでは、fams-aceのアルゴリズムと、更なる改良を試みた結果を報告する。

### 2 fams-ace

fams-aceの処理プロセスは主に以下の3つに要約される。

- ① サーバー部門で提出されたサーバーの予測構造を収集し、FAMSプログラム<sup>3</sup>を使用したホモロジーモデリング法によって三次元構造の再構築と最適化を行う。
- ② サーバー部門に参加した各サーバーは、それぞれのターゲットに5つの予測構造を提出している。①の構造再構築後、各サーバーの代表構造を選ぶ。代表構造の決定には、本研究室で開発されたタンパク質立体構造評価プログラムのCIRCLEを使用する。CIRCLEはアミノ酸残基の側鎖の環境を評価するプログラムである。<sup>4</sup>

- ③ 代表構造から最も多数意見を反映している構造をコンセンサス法によって5つ選び、fams-aceの予測構造として提出する。

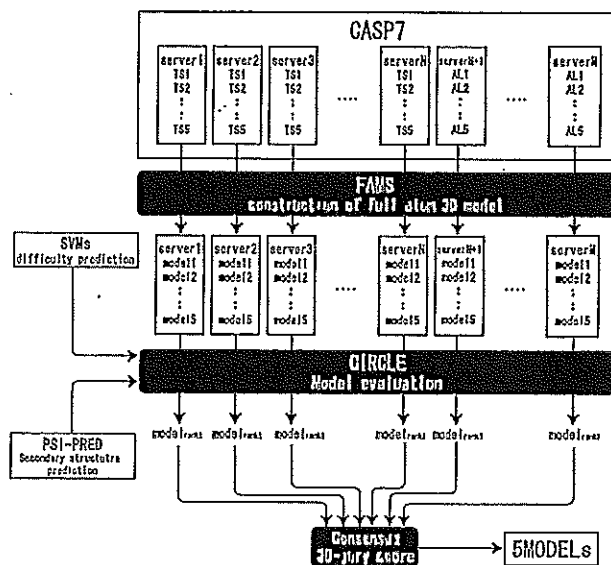


図1 fams-ace のフローチャート

### 3 結果

fams-ace は全自動でありながら GDT\_TS (実験構造とモデル構造の類似度の指標) の高い優れた構造を提出することができた。Zhang らが公表した単純な GDT\_TS の合計値の比較においては、全参加チーム中4位であった。これは fams-ace が、多くの Human expert チーム、つまり人間の手作業を取り入れているチームと比較しても上位の成績を収めることができた事を意味している。このような成績を収めることができたのは以下の3つの要因が考えられる。

- (1) FAMS による構造の再構築により、側鎖の構造の精度 (原子の衝突は90%以上、 $\alpha 1$ ,  $\alpha 1+2$  の精度はそれぞれ 5.3%, 4.2%改善) を高めることができた。
- (2) CIRCLE により間違った構造のフィルタリングが成功した。
- (3) コンセンサス法によって多数意見を反

映した構造を選ぶことで、大きな間違いの無い構造を選ぶことができた。

例として T0371 のドメイン2 (図2) では、CASP7 に参加した全チームの中で最も GDT\_TS が高い構造を提出することができた。

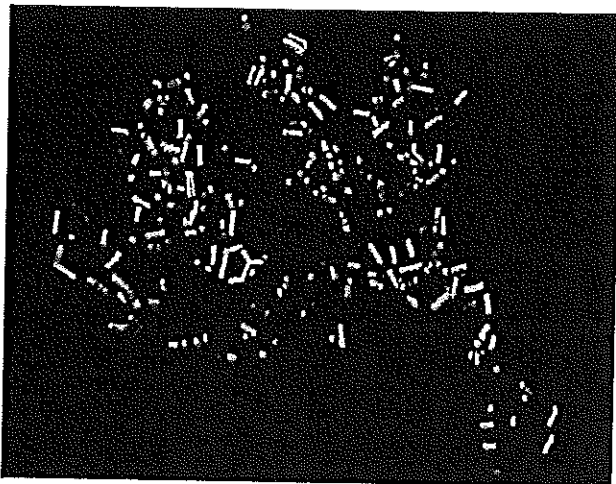


図2 T0371 のドメイン2 magenta: fams-ace, green: 実験構造, cyan: 再構築前の構造

#### 4 今後の展望

fams-ace の方法では、難易度が高いターゲットで精度の高いモデルを提出することができないという問題点があった。コンセンサス法によって最終構造を決定する方法は、サーバー構造の多様性が大きいターゲットでは効果を発揮しないためである。そこで、コンセンサス法を使用してサーバーの構造のフィルタリングを行い、最終的な提出構造は CIRCLE で選ぶという方法を行い検証した(fams-ace(improved))。

図3は、fams-ace, fams-ace(improved), コンセンサス法のみで提出構造を選ぶ 3D-Jury<sup>5</sup> の3つの方法を比較したものである。サーバーの構造の内、GDT\_TS の最高値(bestGDT\_TS)と、各方法が提出したモデル構造の GDT\_TS の比較を行った。図3にあるように、fams-ace(improved)は fams-ace よりも良いモデル構造を選び出し提出することができることが分かる。fams-ace(improved)は、側鎖の精度( $\alpha 1, \alpha 1+2$ )においても、5.5%, 7.3%の改善が観察することができた。

fams-ace は全自動で精度の高い構造の構築ができるシステムを構築することができるシステムであるが、多くの改善の余地があることも分かっている。今後、fams-ace を改良していくことで、より精度の高い構造を自動的に計算し、生物学的

な情報、知識を人間の手で加えることができるシステムを構築する予定である。

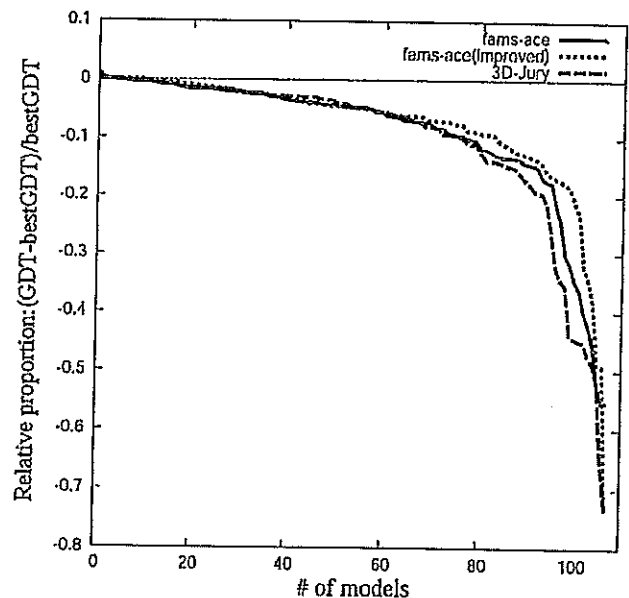


図3 各ターゲットにおけるサーバーの提出構造の中での最も高い GDT\_TS と fams-ace, fams-ace(improved), 3D-Jury が提出した構造の比較。

#### 参考文献

1. <http://predictioncenter.org/casp7/>
2. [http://predictioncenter.org/casp7/meeting\\_docs/meeting\\_program.html](http://predictioncenter.org/casp7/meeting_docs/meeting_program.html)
3. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. *J Mol Graph Model* 2000;18:258-272, 305-256.
4. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, Akio Hosoi, Kazuhiro Ohta, and Hideaki Umeyama. Fams-ace: a combined method to select the best model after remodeling all server models. *Proteins*, accepted.
5. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, and Hideaki Umeyama. The SKE-DOCK server and human teams based on a combined method of shape complementarity and free energy estimation. *Proteins*, accepted.
6. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 2003;19:1015-1018.

## KP21

### FAMSmulti: Automated homology modeling using multiple reference proteins

(<sup>1</sup> - School of Pharmaceutical Sciences, Kitasato University)

○Kazuhiko Kanou<sup>1</sup>, Mitsuo Iwadate<sup>1</sup>, Genki Terashi<sup>1</sup>,  
Mayuko Takeda-Shitaka<sup>1</sup>, Daisuke Takaya<sup>1</sup>, Hiroko Sakai<sup>1</sup>  
and Hideaki Umeyama<sup>1</sup>

#### INTRODUCTION:

We had developed an automated method of the protein structure prediction called FAMS (Full Automatic Modeling System) [1,2]. FAMS is a homology modeling program consisting of database search and simulated annealing, and can construct high accuracy model when appropriate reference protein was detected. For predicting more accurate model (especially loop structure and side-chain torsion angles), we developed new version of FAMS, called FAMS-multi, which use multiple reference proteins.

For the purpose of assessment of this method, we participated in CASP7 (7<sup>th</sup> Critical Assessment of Techniques for Protein Structure Prediction) experiment (our team name is 'fams-multi'). CASP is a world-wide experiment for protein structure prediction held every two years since 1994. CASP provides participants with more than 100 amino acid sequences, and participants must submit the predicted structures within 48 hours (for automatic predictors) or more 2 weeks (for non-automatic predictors). Non-automatic predictors can use models which were predicted by automatic predictors. We participated as a non-automatic predictor for the purpose of using automatic predicted models, but all process had been performed full automatically. Models which were predicted by other automatic predictors were used to generate the better alignments, and we rebuilt models by FAMS-multi program which use multiple reference proteins. In the following, we describe the scheme of this method.

#### METHODS:

##### 1. Generating the 'best' pair wise alignments

The predicted models automatically by the other teams were refined by FAMS for the purpose of removing collision, and these models were evaluated and ranked by the CIRCLE method (later description). Then top 5 models (in excluding models which hasn't described reference PDB code on the 'PARENT' record) were selected to generate alignments. The alignment was generated by structural alignment

between the model and 'parent' PDB using CE program [3].

##### 2. Constructing models by FAMS-multi

First some reference proteins were chosen based on the certain criteria concerning sequence and structural similarity with 'parent' PDB proteins. Maximum number of the reference proteins is 30. Next, a multiple structural alignment based on the superposition of CA atoms was performed among the reference proteins. For this alignment, the target sequence was put on according to sequence alignment generated by CE. This alignment was evaluated to fit if inserted gaps were concentrated in loop and variable regions (VRs), which are defined by residues having the distance between CA atoms greater than 1.0 Å. Thus, we get a result of multiple alignment between a target protein and reference proteins.

Using this alignment, tertiary structure was constructed mainly with next three steps, CA construction, main-chain construction, side-chain construction. In the each step, optimization was executed by the simulated annealing method.

CA construction step: For the initial CA coordinates, first, the weighted average of CA coordinates and the average distance were obtained from pairwise structural alignment based on the superposition of CA atoms between the target and reference proteins. Next, simulated annealing optimized the coordinates of CA atoms.

Main-chain construction step: Initial coordinates of main-chain atoms were constructed in the same method as FAMS. In the simulated annealing step, potential function, which is consisting of (1) the weighted average of the coordinates of main-chain atoms, (2) the average of distance, and (3) the pair of N and O atoms forming the hydrogen bond as structural information, was used.

Side-chain construction step: For the generated main-chain atoms, conserved side-chain torsion angles



were obtained from homologous proteins. The coordinates of side-chain atoms consisting of conserved side-chain torsion angles were placed in relation to the fixed main-chain atoms. The structural information, the weighted average of the coordinates, average of distance, and the pair of N and O atoms forming the hydrogen bond, was derived from homologous proteins, and this information was used in optimization procedure.

**3. Evaluating models (CIRCLE method)**

First the target difficulty was predicted for the purpose of the evaluation method depending on difficulty. For predicting the target difficulty, we used Support Vector Machine (SVM) as the classification tool in the two categories of "CM" and "FR, NF". The training data set was CASP6 targets. The five models constructed by FAMS-multi were evaluated by either the scoring functions for "CM" (easy) or "FR, NF" (difficult), in the following equation (1).

$$TotalScore = \begin{cases} 0.35 \times SSscore + 3DIDscore_{CM} & CM \\ 0.75 \times SSscore + 3DIDscore_{FRNF} & FR \text{ or } NF \end{cases} \quad (1)$$

$$SSscore = \sum f(SS_{PREDICTED}, SS_{MODEL}, confidence) \quad (2)$$

SSscore represents the measure of secondary structure similarity, calculated by comparing secondary structure of model and the result of PSIPRED. SS<sub>PREDICTED</sub> represents the secondary structure predicted by PSIPRED. SS<sub>MODEL</sub> is the secondary structure of model. "confidence" is the confidence of prediction, taken from PSIPRED output. 3DIDscore<sub>CM</sub> and 3DIDscore<sub>FRNF</sub> are scoring function to evaluate side chain environments. These functions were refined by CASP6 models and difficulties of targets. The 3DID score is calculated by 3 parameters (fraction of buried area, fraction of polar area, Secondary structure). As shown in the equation (1) and (2), more weight in difficult targets (FR, NF) than easy targets (CM) is given for SSscore.

**4. Refinement model**

Five selected models were refined using Energy minimize & Molecular dynamics. Refined models are correctly revised for hydrogen bonds, main-chain torsion angles, side-chain torsion angles and the decreasing collision between hydrophobic atoms.



Figure1. Automatic scheme of *fams-multi*.

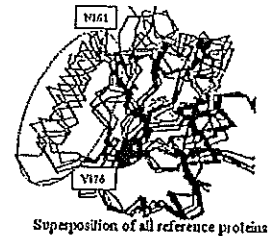
**RESULTS & DISCUSSION:**

In "high accuracy" targets corresponding to 27 domains, *fams-multi* team predicted models having high accuracy backbone geometry and side-chain torsion angles. In the several cases, models constructed using multiple reference proteins improved backbone over the each single template. In accuracy of side-chain torsion angles, more than 70 % of region was predicted successfully. We fixed correct  $\alpha_1$  angle within 30 degree with native structure.

We show an example that *fams-multi* predicted successfully.

**Example: Target T0378 (PDB ID=2I6D)**

This target contains two domains, and, therefore, target sequence was divided into two regions. *Fams-multi* was performed for each domain. The second domain (T0378\_D2) has some reference proteins in PDB, and these structures



were almost similar excluding one helix which consists of about 15 amino acid residues. The relative position of this helix is variable between reference proteins, and many predictors mistook this helix position. *Fams-multi* predicted this helix successfully, so GDT\_TS (accuracy of backbone geometry) of this target (sum of T0378\_D1 and T0378\_D2) was the highest among all predictor's models.

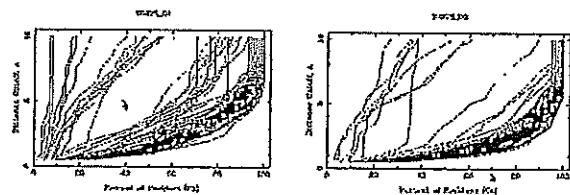


Figure2 Coverage plot of T0378\_D1 (left) and T0378\_D2 (right). Black line indicates model of *fams-multi* and orange lines indicate model of other predictors.

**REFERENCE:**

[1] Ogata, K. and Umeyama, H. (2000). An automatic homology modeling method consisting of database searches and simulated annealing J Mol Graph Model/J Mol Graph Model 18, 258-272, 305-256.  
 [2] Ogata K, Umeyama H. (1998) The role played by environmental residues on sidechain torsional angles within homologous families of proteins: a new method of sidechain modeling. Proteins. 31(4):355-69.  
 [3] Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering 11(9) 739-747.

# KP22

## SKE-DOCK: 相補性を考慮したドッキングと最適構造評価法による、タンパク質複合体立体構造予測法の開発

(北里大学・薬) ○寺師 玄記, 竹田一志鷹 真由子,  
加納 和彦, 岩館 満雄, 高谷 大輔, 酒井 博子, 梅山 秀明

### 1.はじめに

タンパク質-タンパク質相互作用 (Protein-Protein Interaction: PPI) 状態を予測する Protein-Protein Docking は、タンパク質立体構造予測の次の段階として、非常に注目されている分野である。タンパク質が何らかの活性や機能を示している状態は、タンパク質複合体の状態であることが多い。したがって、タンパク質複合体の状態を精度よく予測することができるならば、疾患に関係したタンパク質の機能についての情報を得ることができると考えられている。

PPI 予測のアルゴリズムを、ブラインドテストによって評価するコンテスト Critical assessment of predicted interaction (CAPRI) が現在行われており、本研究室も参加することができた。本発表では、CAPRI に参加した我々の SKE-DOCK サーバー (全自動の予測サーバー) と Human チーム (手作業で生物学的な情報などを使用することができる) のアルゴリズムの説明と、その結果、問題点を述べる予定である。

### 2.SKE-DOCK サーバー

図 1 に SKE-DOCK サーバー<sup>1</sup> と Human チームのフローチャートを示した。SKE-DOCK サーバー (実線の矢印) は主に 4 つの処理を全自動で行う。

- ① タンパク質立体構造の表面を、膨大な数の四角錐上のパッチで表現し、そのパッチ同士をフィッティングする。
- ② 形の相補性によって順位をつけ、上位 5000 構造をクラスタリングし冗長性を除く。
- ③ 側鎖の環境を評価するプログラム CIRCLE<sup>2</sup> でドッキング構造を評価する。上位 10 個の候補構造を選ぶ。
- ④ 側鎖の再構築、最適化を行い提出する。

Human チーム (破線矢印) は、③の前に側鎖の再構築を行い、さらに文献情報を用いて手動で順位付けを行う。

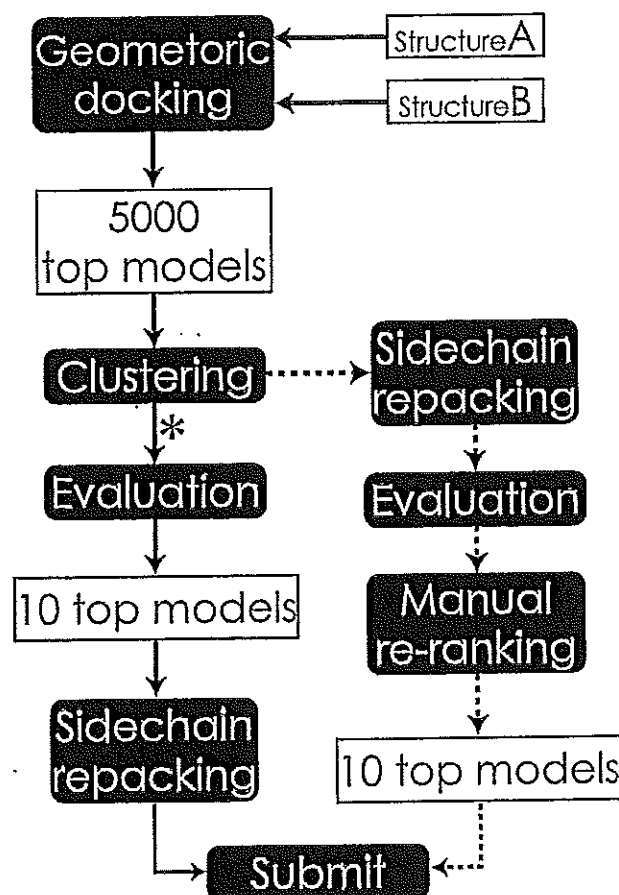


図 1 SKE-DOCK サーバーと、Human チームのフローチャート

### 3.結果

T20-28 の 9 つのターゲット中、T25 で SKE-DOCK (図 2), Human チーム (図 3) 共に精度の高い予測をすることができた。ドッキングの精度を表す指標 Ligand RMS ( $L_{rms}$ )とは、予測構造と実験構造をレセプター分子で重ね合わせたとき、リガンド分子間の rmsd 値を求めたものである。T25 における  $L_{rms}$  値は、共に 3.3Å であった。このときの、CIRCLE のスコアと  $L_{rms}$  の分布図を図 4 で示した。

図 4 にあるように、T25 において CIRCLE のスコアが高い候補構造は  $L_{rms}$  が低い傾向にある。このように CIRCLE による評価が正しい場合には精度の良いドッキング構造をよそくすることが

できる。



図2 Human チームの予測構造 (緑、シアン) と実験構造 (マゼンダ、黄) の比較



図3 SKE-DOCK サーバーの予測構造 (緑、シアン) と実験構造 (マゼンダ、黄) の比較

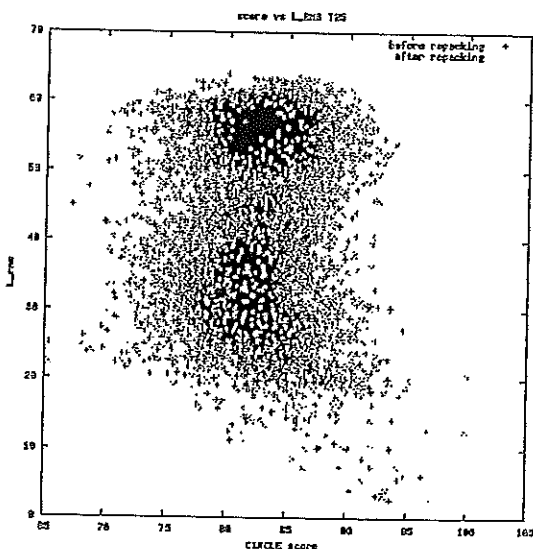


図4 T25 の候補構造における CIRCLE スコアと  $L_{rms}$  の分布図。赤：側鎖の再構築前。緑：側鎖の再構築後。

逆に、T26 では CIRCLE の評価で失敗している事例も観察された (図5)。相補性によるドッキングでは、 $L_{rms}$  の低い候補構造が得られていたが、CIRCLE でそれらを選ぶことができなかった例で

ある。

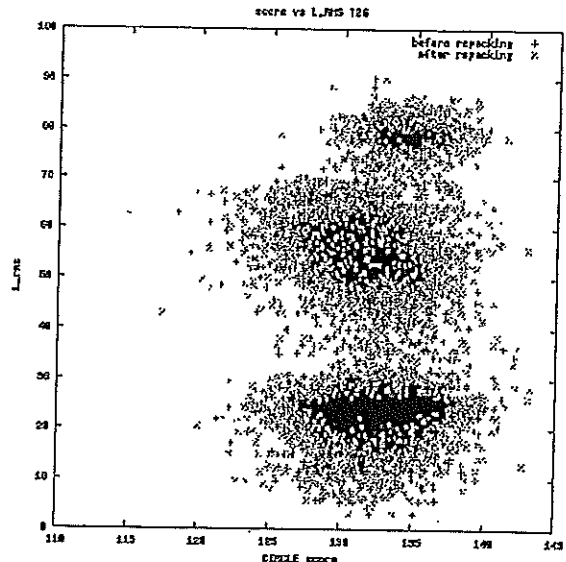


図5 T26 の候補構造における CIRCLE スコアと  $L_{rms}$  の分布図

#### 4. 今後の展望

相補性を考慮したドッキングと CIRCLE による評価を組み合わせることで、T25 では精度の高い予測を行うことができた。だが、他の例では正しく評価を行うことができなかった。このことから、CIRCLE を使用した評価にはまだ改善の余地があると思われる。

また、ドッキングにはコンフォメーションの変化を考慮する必要がある。コンフォメーションの変化が小さいものについては、相補性によるドッキングで対応できるが、コンフォメーションの変化が大きいものでは対応できない。これらの問題を解決する新しいドッキング方法を今後取り入れていく予定である。

#### 参考文献

1. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, and Hideaki Umeyama. The SKE-DOCK server and human teams based on a combined method of shape complementarity and free energy estimation. *Proteins*, accepted.
2. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, Akio Hosoi, Kazuhiro Ohta, and Hideaki Umeyama. Fams-ace: a combined method to select the best model after remodeling all server models. *Proteins*, accepted.

# KP23

## fams-ace: CASP7 におけるタンパク質立体構造の再構築と、最適構造の評価法の開発

(北里大学・薬) ○竹田一志鷹 真由子, 寺師 玄記, 加納和彦, 岩館 満雄, 高谷 大輔, 酒井 博子, 梅山 秀明

### 1 はじめに

タンパク質立体構造予測コンテストである Critical Assessment of Protein Structure Prediction (CASP7, Pacific Grove, CA, 2006)<sup>1</sup>とは、実験解析による立体構造の情報がホールドされた状態で、アミノ酸配列情報のみが与えられ、その三次元構造の予測精度を競うものである。参加するチームは主に二つに分けられる。一つは全自動で予測を行う Server部門。もう一つは、サーバーが予測した結果を考慮して人間による手動操作による修正などを行う Human expert部門である。

2006年5月から開催されたタンパク質立体構造予測コンテストであるCAS7において、本研究室は立体構造予測部門(TS)で4つの立体構造予測サーバーチーム (FAMS, FAMSD, CIRCLE, FUNCTION) と5つのHuman expertチーム (CHIMERA, fams-ace, CIRCLE-FAMS, fams-multi, Ligand-CIRCLE) で参加した。特に、fams-aceにおいては、当研究室の梅山秀明教授がCAS7国際会議でスコア上位者として招待講演を行った。<sup>2,3</sup>

ここでは、fams-aceのアルゴリズムと、更なる改良を試みた結果を報告する。

### 2 fams-ace

fams-aceの処理プロセスは主に以下の3つに要約される。

- ① サーバー部門で提出されたサーバーの予測構造を収集し、FAMSプログラム<sup>3</sup>を使用したホモロジーモデリング法によって三次元構造の再構築と最適化を行う。
- ② サーバー部門に参加した各サーバーは、それぞれのターゲットに5つの予測構造を提出している。①の構造再構築後、各サーバーの代表構造を選ぶ。代表構造の決定には、本研究室で開発されたタンパク質立体構造評価プログラムのCIRCLEを使用する。CIRCLEはアミノ酸残基の側鎖の環境を評価するプログラムである。<sup>4</sup>

- ③ 代表構造から最も多数意見を反映している構造をコンセンサス法によって5つ選び、fams-aceの予測構造として提出する。

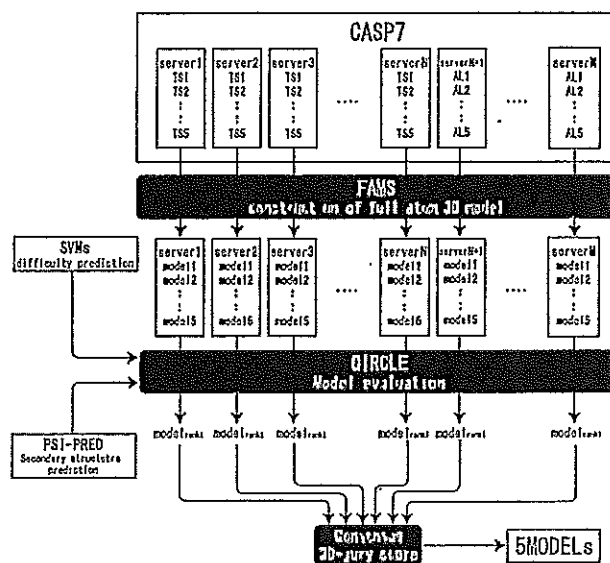


図1 fams-ace のフローチャート

### 3 結果

fams-ace は全自動でありながら GDT\_TS (実験構造とモデル構造の類似度の指標) の高い優れた構造を提出することができた。Zhang らが公表した単純な GDT\_TS の合計値の比較においては、全参加チーム中4位であった。これは fams-ace が、多くの Human expert チーム、つまり人間の手作業を取り入れているチームと比較しても上位の成績を収めることができた事を意味している。このような成績を収めることができたのは以下の3つの要因が考えられる。

- (1) FAMS による構造の再構築により、側鎖の構造の精度 (原子の衝突は90%以上、 $\alpha 1$ ,  $\alpha 1+2$ の精度はそれぞれ5.3%, 4.2%改善) を高めることができた。
- (2) CIRCLE により間違った構造のフィルタリングが成功した。
- (3) コンセンサス法によって多数意見を反

映した構造を選ぶことで、大きな間違いの無い構造を選ぶことができた。

例として T0371 のドメイン2 (図2) では、CASP7 に参加した全チームの中で最も GDT\_TS が高い構造を提出することができた。

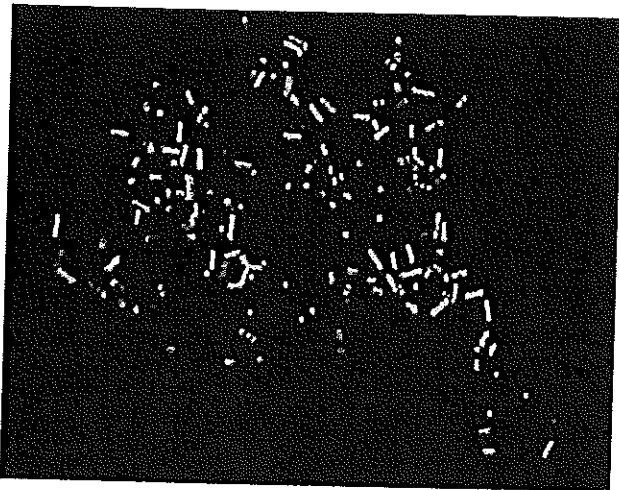


図2 T0371 のドメイン2 magenta: fams-ace, green: 実験構造, cyan:再構築前の構造

#### 4 今後の展望

fams-ace の方法では、難易度が高いターゲットで精度の高いモデルを提出することができないという問題点があった。コンセンサス法によって最終構造を決定する方法は、サーバー構造の多様性が大きいターゲットでは効果を発揮しないためである。そこで、コンセンサス法を使用してサーバーの構造のフィルタリングを行い、最終的な提出構造は CIRCLE で選ぶという方法を行い検証した(fams-ace(improved))。

図3は、fams-ace, fams-ace(improved), コンセンサス法のみで提出構造を選ぶ 3D-Jury<sup>5</sup> の3つの方法を比較したものである。サーバーの構造の内、GDT\_TS の最高値(bestGDT\_TS)と、各方法が提出したモデル構造の GDT\_TS の比較を行った。図3にあるように、fams-ace(improved)は fams-ace よりも良いモデル構造を選び出し提出することができることが分かる。fams-ace(improved)は、側鎖の精度( $\alpha 1, \alpha 1+2$ )においても、5.5%, 7.3%の改善が観察することができた。

fams-ace は全自動で精度の高い構造の構築ができるシステムを構築することができるシステムであるが、多くの改善の余地があることも分かっている。今後、fams-ace を改良していくことで、より精度の高い構造を自動的に計算し、生物学的

な情報、知識を人間の手で加えることができるシステムを構築する予定である。

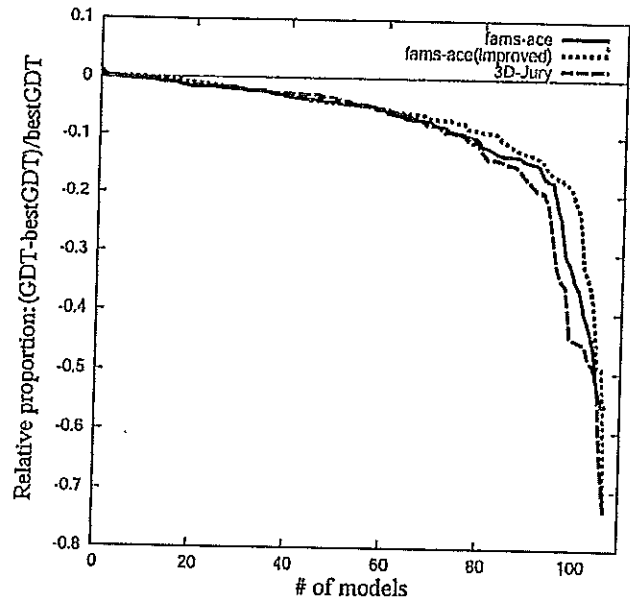


図3 各ターゲットにおけるサーバーの提出構造の中での最も高い GDT\_TS と fams-ace, fams-ace(improved), 3D-Jury が提出した構造の比較。

#### 参考文献

1. <http://predictioncenter.org/casp7/>
2. [http://predictioncenter.org/casp7/meeting\\_docs/meeting\\_program.html](http://predictioncenter.org/casp7/meeting_docs/meeting_program.html)
3. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. *J Mol Graph Model* 2000;18:258-272, 305-256.
4. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, Akio Hosoi, Kazuhiro Ohta, and Hideaki Umeyama. Fams-ace: a combined method to select the best model after remodeling all server models. *Proteins*, accepted.
5. Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, and Hideaki Umeyama. The SKE-DOCK server and human teams based on a combined method of shape complementarity and free energy estimation. *Proteins*, accepted.
6. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 2003;19:1015-1018.

## KP24

### Membrane-associated Proteins in Eicosanoid and Glutathione Metabolism (MAPEG) superfamily complex modeling

(<sup>1</sup> - School of Pharmaceutical Sciences, Kitasato University)

○ Kazuhiko Kanou<sup>1</sup>, Mayuko Takeda-Shitaka<sup>1</sup>, Mitsuo Iwadate<sup>1</sup>, Genki Terashi<sup>1</sup>, Daisuke Takaya<sup>1</sup>, Hiroko Sakai<sup>1</sup> and Hideaki Umeyama<sup>1</sup>

#### INTRODUCTION:

MAPEG (membrane associated proteins in eicosanoid and glutathione metabolism) is a widespread superfamily which has been defined according to sequence similarity, enzymatic activities and structural properties. This family consists of some human proteins including microsomal prostaglandin E<sub>2</sub> synthase (mPGES), 5-lipoxygenase-activating protein (FLAP), leukotriene C<sub>4</sub> (LTC<sub>4</sub>) synthase, microsomal glutathione S-transferase 1 (MGST1), MGST2, MGST3, and MGST1-like 1 (MGST-L1). mPGES catalyses the synthesis of PGE<sub>2</sub> from PGH<sub>2</sub> (produced by cyclooxygenase from arachidonic acid). Because of structural similarities in the active sites of FLAP, LTC<sub>4</sub> synthase and PGE synthase, substrates for each enzyme can compete with one another and modulate synthetic activity.

Recently the crystal structure of human LTC<sub>4</sub> synthase (PDB-ID=2UUH, 2UUI, 2PNO)[1] and human FLAP (PDB-ID=2Q7M, 2Q7R) had been determined. These structures form homo-trimer that each subunit is composed five long  $\alpha$ -helices and first four helices form the transmembrane segments. We performed automatic protein structure prediction of MAPEG superfamily, and these crystal structures were used as a template structure of homology modeling method. 104 protein structures were predicted using structure of LTC<sub>4</sub> synthase and 234 protein structures were predicted using structure of FLAP. All of models were constructed as three subunit complex by using homology modeling program FAMS Ligand&Complex[2,3]. These predicted models are

now available on the following URL.

<http://famshelp.gsc.riken.jp/famsbase/MAPEG/MAPEG.htm>

Furthermore we predicted human mPGES models using non-automatic prediction method.

#### METHODS:

##### High-throuput modeling of MAPEG family

Target proteins for high-throuput homology modeling were collected using PSI-BLAST for NCBI-nr sequence database. The homologous proteins of each crystal structure were selected with criteria of E-value < 1e-10. For each target sequence, three homology search programs (PSI-BLAST, RPS-BLAST and IMPALA) were performed and three alignments between target protein and template structure were obtained. These alignments were used to construct trimer model by using homology modeling program FAMS Ligand&Complex. Then ligands which are complexed with each template protein were involved.

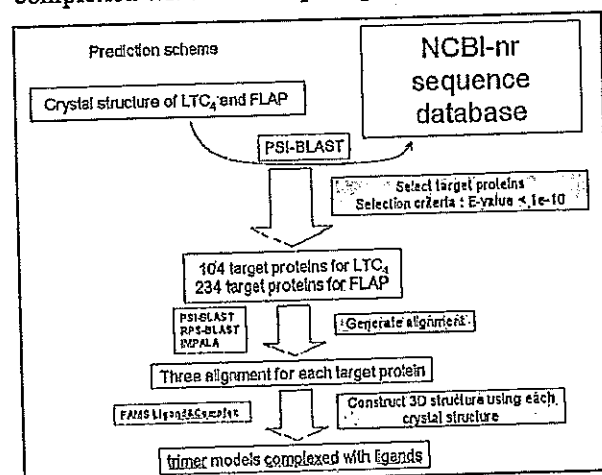


Figure1 Prediction scheme.

Crystal structures and ligands which were used as a template structure are shown as following table.

PDB-ID	Protein name	Ligands
2UUH	Human LTC <sub>4</sub>	GLUTATHIONE, MALTOSE
2PNO	Human LTC <sub>4</sub>	GLUTATHIONE, MALTOSE
2Q7M	Human FLAP	MK-591 (FLAG inhibitor)
2Q7R	Human FLAP	IODINATED ANALOG OF MK-591
2H8A	Rat MGST1	GLUTATHIONE

Table1 templates and ligands

### Human mPGES-1 modeling

MPGES-1 may be a critical determinant of postischemic neurological dysfunctions and a valuable therapeutic target for treatment of human stroke [4]. Therefore we predicted structure of human mPGES-1 manually. The automatic alignment between mPGES-1 and LTC<sub>4</sub> was modified using SKE-CHIMERA [5] and two trimer models with ligands were constructed based on this alignment using 2UUH and 2PNO as a template. Because the alignment between mPGES-1 and FLAP couldn't be obtained by automatic procedure, we performed intermediate sequence search (ISS) method. Alignment between mPGES-1 and FLAP was made from SKE-CHIMERA alignment between mPGES-1 & LTC<sub>4</sub> and automatic alignment between LTC<sub>4</sub> & FLAP by using sequence of LTC<sub>4</sub> as "intermediate sequence", and two trimer models with ligands were constructed based on this alignment using 2Q7M and 2Q7R as a template. Models using 2UUH and 2PNO as a template involve three glutathiones in each active site. Models using 2Q7M and 2Q7R as a template involved MK-591 (FLAP inhibitor) and iodinated analog of MK-591, respectively.

### RESULTS & DISCUSSION:

After the all of high-throughput modeling was complete, the quality of the stereochemistry of the models of MAPEG superfamily was verified. In the Ramachandran plot of the mainchain  $\phi$ - $\psi$  angles rendered by the program PROCHECK (Laskowski et al., 1993), almost all of the non-glycine residues were in the most favoured or allowed regions. Average percentage of residues in disallowed regions of all models is only 1.45%. Moreover, all of the  $\omega$  angles were trans-planar. The average of percentage of

residues in most favoured regions, in additional allowed regions, in generously allowed regions and in disallowed regions are shown as table2.

Template structure	Residues in most favoured regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
2UUH	88.89%	8.16%	2.01%	0.94%
2PNO	89.99%	7.62%	1.51%	0.89%
2Q7M	82.33%	12.87%	3.05%	1.75%
2Q7R	83.03%	11.79%	3.01%	2.18%
2H8A	74.03%	19.84%	4.14%	1.99%
Total	86.12%	10.10%	2.33%	1.45%

Table2 PROCHEK result. Average of all models from each template structure.

We show superposition of two human mPGES-1 models based on manual alignments (Figure2). One of them is a model using 2UUH as a template (green), and another is a model using 2Q7M as a template (cyan). Two models are similar, but binding site of each ligand is different.



Orange molecule is glutathione, and magenta molecule is MK-591 (FLAP inhibitor). We suggest the possibility that MK binding site of this model is substrate binding site of mPGES, and now we attempt docking simulation for this site using original protein-ligand docking program. We expect that the candidate compounds for treatment of human stroke are found as a result of this docking simulation.

### REFERENCE:

- [1] Wetterholm A, Kohl A, McCarthy AA, Niegowski D, Ohlson B, Hammarberg T, Eshaghi S, Haeggström JZ, Nordlund P, Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase. *Molina DM, Nature.* 2007 Aug 2;448(7153):613-6. Epub 2007 Jul 15.
- [2] Ogata, K. and Umeyama, H. (2000). An automatic homology modeling method consisting of database searches and simulated annealing *J Mol Graph Model/J Mol Graph Model* 18, 258-272, 305-256.
- [3] Takeda-Shitaka M, Terashi G, Chiba C, Takaya D, Umeyama H. FAMS complex: a fully automated homology modeling system for protein complex structures. *Med Chem.* 2006 Mar;2(2):191-201.
- [4] Ikeda-Matsuo Y, Ota A, Fukada T, Uematsu S, Akira S, Sasaki Y., Microsomal prostaglandin E synthase-1 is a critical factor of stroke-reperfusion injury., *Proc Natl Acad Sci U S A.* 2006 Aug 1;103(31):11790-5. Epub 2006 Jul 24.
- [5] Takeda-Shitaka M, Terashi G, Takaya D, Kanou K, Iwadate M, Umeyama H., Protein structure prediction in CASP6 using CHIMERA and FAMS., *Proteins.* 2005;61 Suppl 7:122-7.

**K09: fams-ace: Refinement of 3D models and development of assessment method for selecting the best model in CASP7.**

(School of Pharmacy, Kitasato University) ○Mayuko Takeda-Shitaka, Genki Terashi, Kazuhiro Kanou, Mitsuo Iwadate, Daisuke Takaya, Hiroko Sakai and Hideaki Umeyama.

fams-ace was entered in the 3D coordinate prediction category as a human expert group in CASP7. The procedure of fams-ace can be summarized as the following 3 steps. (1) All the server models were refined and rebuilt utilizing our homology modeling method by using FAMS. (2) Representative structures were selected from each server, according to a model quality evaluation based on a 3D1D profile score. (3) The top 5 models were selected and submitted in the order of their consensus-based score. Fams-ace is a fully automated server and does not require human intervention. We introduce the methodology of fams-ace and discuss the successes and failures of this approach during CASP7. In addition, we discuss possible improvements of fams-ace.

**K10: Calculation of the space between subunits of protein**

(\* National Institute of Agrobiological Sciences, \*\* Mitsubishi Space Software Co. Ltd.) ○Masahiko Maeda\* and Hiroya Nobori\*\*

A new method to calculate the space between subunits of a protein is proposed. It consists of three steps as follows. (1) The interface area is defined from the decreasing accessible surface area ( $\Delta$ ASA) of each subunit. (2) The assembling region is defined as a polyhedron gathering tetrahedra. (3) Grid points are produced in the tetrahedra and the points outside of the atoms are picked up. The chosen points are counted and the volume of the empty space is calculated. The volume (assembling space volume, ASV) is well correlated to the  $\Delta$ ASA. The average of the ASV/ $\Delta$ ASA value (ASV index) is about 1.4 Å. The ASV values of 68 homo-dimer proteins were compared to the gap volumes of SURFNET by Laskowski. Generally the gap volume was larger than the corresponding ASV value. We aim to use the ASV/ $\Delta$ ASA value (ASV index) as an index of the shape complementarity. Thus we also analyze the ASV index to the SC of Lawrence et al., but no correlations were detected between two parameters. We also tried to detect atom occupancy in the defined assembling region. As the result, the average of the atom occupancy of protein is approximately 50% in homo-dimer proteins.



### **KP20: Identification of novel inhibitors for ubiquitin C-terminal hydrolase-L3 by virtual screening**

<sup>1</sup>Dept. of Electr. Eng. & Bioscience, Grad. Sch. of Advanced Sci. and Eng., Waseda Univ., Japan

<sup>2</sup>Dept. of Degenerative Neurological Diseases, Nat. Inst. of Neuroscience, NCNP, Japan)

○Kazunori Hirayama<sup>1,2</sup>, Shunsuke Aoki<sup>2</sup>, Kaori Nishikawa<sup>2</sup>, Takashi Matsumoto<sup>1</sup> and Keiji Wada<sup>2</sup>

UCH-L3 (ubiquitin C-terminal hydrolase-L3) is a de-ubiquitinating enzyme that is a component of the ubiquitin-proteasome system and known to be involved in programmed cell death. A previous study of high-throughput drug screening identified an isatin derivative as a UCH-L3 inhibitor. In this study, we attempted to identify a novel inhibitor with a different structural basis. We performed in silico structure-based drug design (SBDD) using human UCH-L3 crystal structure data (PDB code; 1XD3) and the virtual compound library (ChemBridge CNS-Set), which includes 32,799 chemicals. By a two-step virtual screening method using DOCK software (first screening) and GOLD software (second screening), we identified 10 compounds with GOLD scores of over 60. To address whether these compounds exhibit an inhibitory effect on the de-ubiquitinating activity of UCH-L3, we performed an enzymatic assay using ubiquitin-7-amido-4-methylcoumarin (Ub-AMC) as the substrate. As a result, we identified three compounds with similar basic dihydro-pyrrole skeletons as UCH-L3 inhibitors. These novel compounds may be useful for the research of UCH-L3 function, and in drug development for UCH-L3-associated diseases.

### **KP21: FAMSmulti: Automated homology modeling using multiple reference proteins**

<sup>1</sup> - School of Pharmaceutical Sciences, Kitasato University) ○Kazuhiko Kanou<sup>1</sup>, Mitsuo Iwadata<sup>1</sup>, Genki Terashi<sup>1</sup>, Mayuko Takeda-Shitaka<sup>1</sup>, Daisuke Takaya<sup>1</sup>, Hiroko Sakai<sup>1</sup> and Hideaki Umeyama<sup>1</sup>

We had developed an automated method of the protein structure prediction called FAMS (Full Automatic Modeling System). FAMS is a homology modeling program, and can construct high accuracy model when appropriate reference protein was detected. For predicting more accurate model (especially loop structure and side-chain torsion angles), we developed new version of FAMS, called FAMS-multi, which use multiple reference proteins.

For the purpose of assessment of this method, we participated in CASP7 (7th Critical Assessment of Techniques for Protein Structure Prediction) experiment (our team name is 'fams-multi'). CASP is a world-wide experiment for protein structure prediction held every two years since 1994. All procedure of 'fams-multi' had been performed full automatically. Models which were predicted by other automatic predictors were used to generate the better alignments, and we rebuilt models by FAMS-multi program which use multiple reference proteins. Results of CASP7 experiment is available on following URL. <http://predictioncenter.org/casp7/>

**KP22: SKE-DOCK server and human teams: protein-protein docking based on the shape complementarity and free energy estimation.**

(School of Pharmacy, Kitasato University) ○Genki Terashi, Mayuko Takeda-Shitaka, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, Hiroko Sakai and Hideaki Umeyama.

We participated in rounds 6–12 of the critical assessment of predicted interaction (CAPRI) contest as the SKE-DOCK server and human teams. The SKEDOCK server is based on simple geometry docking and a knowledge base scoring function. The procedure is summarized in the following three steps: (1) protein docking according to shape complementarity, (2) evaluating complex models, and (3) repacking side-chain of models. Good predicted models were obtained for target 25 by both the SKEDOCK server and human teams. The smallest Ligand-rmsd values corresponding to the rmsd between the model and experimental structures were 3.307 and 3.324Å, respectively.

**KP23: fams-ace: Refinement of 3D models and development of assessment method for selecting the best model in CASP7.**

(School of Pharmacy, Kitasato University) ○Mayuko Takeda-Shitaka, Genki Terashi, Kazuhiko Kanou, Mitsuo Iwadate, Daisuke Takaya, Hiroko Sakai and Hideaki Umeyama.

fams-ace was entered in the 3D coordinate prediction category as a human expert group in CASP7. The procedure of fams-ace can be summarized as the following 3 steps. (1) All the server models were refined and rebuilt utilizing our homology modeling method by using FAMS. (2) Representative structures were selected from each server, according to a model quality evaluation based on a 3D1D profile score. (3) The top 5 models were selected and submitted in the order of the consensus-based score. Fams-ace is a fully automated server and does not require human intervention. We introduce the methodology of fams-ace and discuss the successes and failures of this approach during CASP7. In addition, we discuss possible improvements of fams-ace.

## **KP24: Membrane-associated Proteins in Eicosanoid and Glutathione Metabolism (MAPEG) superfamily complex modeling**

(<sup>1</sup> - School of Pharmaceutical Sciences, Kitasato University) ○Kazuhiko Kanou<sup>1</sup>, Mayuko Takeda-Shitaka<sup>1</sup>, Mitsuo Iwadate<sup>1</sup>, Genki Terashi<sup>1</sup>, Daisuke Takaya<sup>1</sup>, Hiroko Sakai<sup>1</sup> and Hideaki Umeyama<sup>1</sup>

### **INTRODUCTION:**

MAPEG (membrane associated proteins in eicosanoid and glutathione metabolism) is a widespread superfamily which has been defined according to sequence similarity, enzymatic activities and structural properties. This family consists of some human proteins including microsomal prostaglandin E2 synthase (mPGES), 5-lipoxygenase-activating protein (FLAP), leukotriene C4 (LTC4) synthase, microsomal glutathione S-transferase 1 (MGST1), MGST2, MGST3, and MGST1-like 1 (MGST-L1).

Recently the crystal structure of human LTC4 synthase and human FLAP had been determined. We performed automatic protein structure prediction of MAPEG superfamily, and these crystal structures were used as a template structure of homology modeling method. These predicted models are now available on the following URL,

<http://famshelp.gsc.riken.jp/famsbase/MAPEG/MAPEG.htm>

Furthermore we predicted human mPGES models using non-automatic prediction method.

## **KP25: Analysis of carcinogens classified by IARC and prediction of human carcinogenicity using Support Vector Machines.**

(Graduate School of Pharmaceutical Sciences, The University of Tokushima) ○Toshiaki Takezaki, Shinichi Kobayashi, Kumiko Sakamoto, Masaru Kihara, Aiko Yamauchi

In the drug development process, carcinogenicity test using animals takes huge cost and period. Using an *in silico* carcinogenicity prediction system which shows higher accuracy, carcinogenicity-positive compounds can be removed at the beginning of the process and new valuable compounds would be found efficiently. In this study, we predicted carcinogenicity of 1,072 compounds by Support Vector Machine (SVM), a statistical-learning method. The compounds include both medicines and those evaluated on human carcinogenicity by IARC. The average accuracy of the SVM prediction for carcinogenicity positive and negative compounds was 91.6% and 78.0%, respectively. In order to create more efficient SVM model, we calculated Fscore and applied Mann-Whitney U test on each descriptors. Then 12 significant structural descriptors on the prediction were specified. Furthermore, we got knowledge about CYP isozymes which related the metabolism of carcinogenicity negative/positive compounds. After adding the CYPs information to descriptors, more refined and more accurate SVM analytical model could be created.