創薬インフォマティクス

Department of PharmacoInformatics

奥野恭史

医薬品開発の成功確率

化合物数	499,915	197	97	62	32
移行確率		1/2,538	1/2.03	1/1.56	1/1.94
累積成功率		1/2,538	1/5,154	1/8,063	1/15,622

ボトルネック

（製薬協 Data Book (2001-2005)）

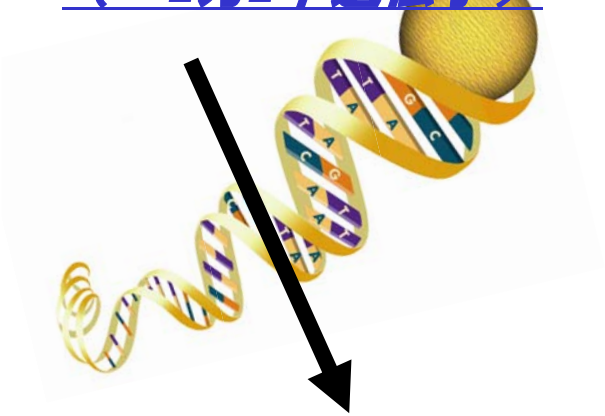
計算機を用いた超効率的な薬物候補探索

⇒ 医薬品の開発期間とコストの短縮

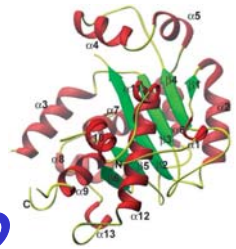
医薬品開発プロセス

ゲノム情報 → ゲノム基盤研究 → ターゲット探索 → ターゲットバリデーション → 創薬リード探索 → 創薬リード最適化 → 前臨床研究 → 臨床研究 → 創 薬

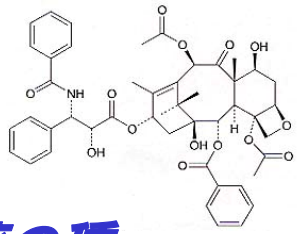
ゲノム情報
（～2万2千遺伝子）



化合物ライブラリー
（10^60 化合物）



疾患の
原因遺伝子の同定



薬の種
リード化合物の選択

医薬品最適化
＆
臨床試験

創薬におけるインフォマティクス

ゲノム情報 → ゲノム基盤研究 → ターゲット探索 → ターゲットバリデーション → 創薬リード探索 → 創薬リード最適化 → 前臨床研究 → 臨床研究 → 創 薬

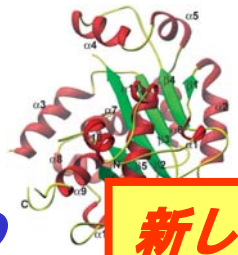
ゲノム情報
（～2万2千遺伝子）

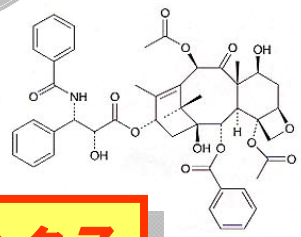
バイオインフォマティクス

化合物ライブラリー
（10^60 化合物）

ケモインフォマティクス

ケミカル
ゲノミクス





新しいインフォマティクス

疾患の
原因遺伝子の同定

リード化合物の選択

医薬品最適化
&
臨床試験

ちょっと復習

多変量解析：クラスター解析

例えば、５科目のテスト結果から、能力別（理系、文系、優秀など）にクラス分けを行いたい場合、どうすれば良いのか？

	国語	社会	数学	理科	英語
a	29	33	55	79	74
b	71	68	72	64	97
c	74	91	79	76	100
d	52	56	58	60	85
e	77	92	96	88	98

人間的に

a～eさんの点数のパターンを眺める → パターンが似ている者どうしを同じグループにする

数学的に

a～eさんの変数をベクトル表現する → 似ているか似ていないかを距離という尺度で定義する

ベクトル表現から類似度定義

a～eさんの変数をベクトル表現する

$\vec{V}a$ = (29, 33, 55, 79, 74)

$\vec{V}b$ = (71, 68, 72, 64, 97)

$\vec{V}c$ = (74, 91, 79, 76, 100)

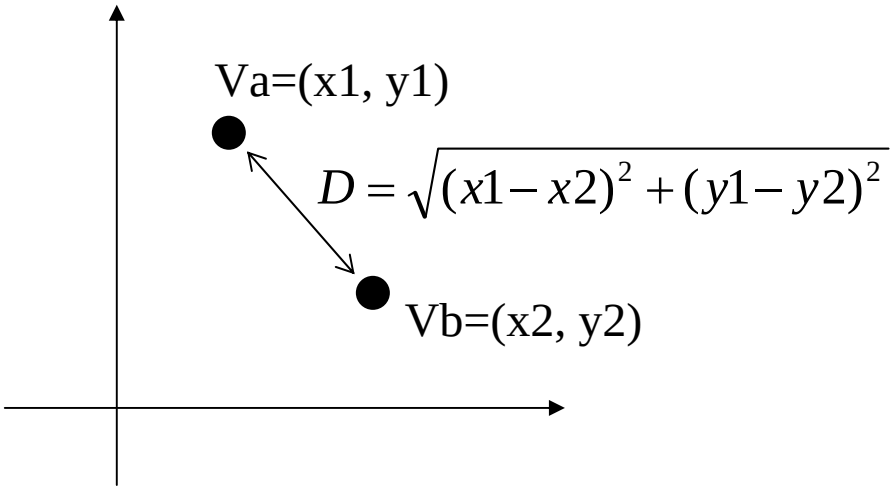
……

似ているか似ていないかを距離という尺度で定義する

ユークリッド距離で表現する
（似ているものは距離が小さい）

$$D = \sqrt{(\vec{V}a - \vec{V}b)^2}$$

簡単のため、２次元の場合



今の場合、５次元になる

$$Dab = \sqrt{(29-71)^2 + (33-68)^2 + ... + (74-97)^2}$$

$$Dac = \sqrt{\ .....}$$

$$Dbc = \sqrt{\ .....}$$

距離行列（類似度行列）

	a	b	c	d	e
a	0	63	81	39	90
b	63	0	27	29	42
c	81	27	0	51	21
d	39	29	51	0	65
e	90	42	21	65	0

a，e 間の距離

最も距離が近いものを一つにグループにまとめ、距離行列を作り直す

Single linkage clustering
小さい方を代表値にして、

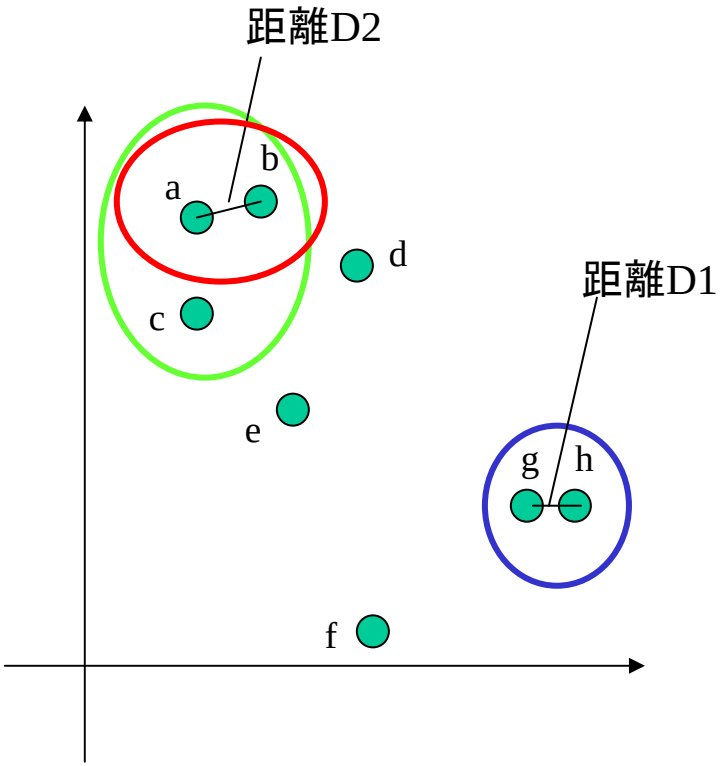
	a	b	d	c, e
a	0	63	39	81
b	63	0	29	27
d	39	29	0	51
c,e	81	27	51	0

Complete linkage clustering
大きい方を代表値にして、

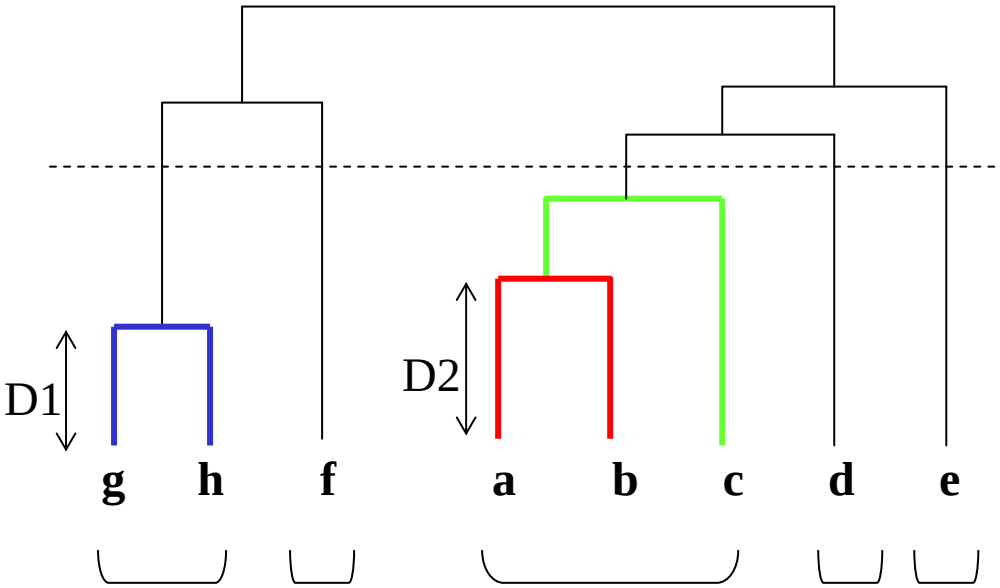
	a	b	d	c, e
a	0	63	39	90
b	63	0	29	42
d	39	29	0	65
c,e	90	42	65	0

階層型クラスタリング

距離の近いものから、グルーピングしていく。



クラスター表記：系統樹



簡単にするため２次元で表現している

創薬におけるインフォマティクス

ゲノム情報 → ゲノム基盤研究 → ターゲット探索 → ターゲットバリデーション → 創薬リード探索 → 創薬リード最適化 → 前臨床研究 → 臨床研究 → 創 薬

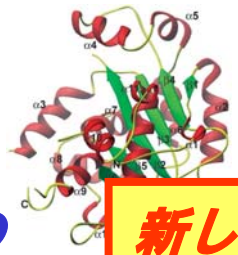
ゲノム情報
（～2万2千遺伝子）

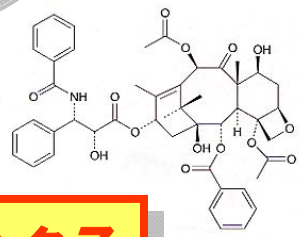
化合物ライブラリー
（10^60 化合物）

バイオインフォマティクス

ケモインフォマティクス

ケミカル
ゲノミクス





新しいインフォマティクス

疾患の
原因遺伝子の同定

リード化合物の選択

医薬品最適化
&
臨床試験

バイオインフォマティクス
配列解析

Sequences information

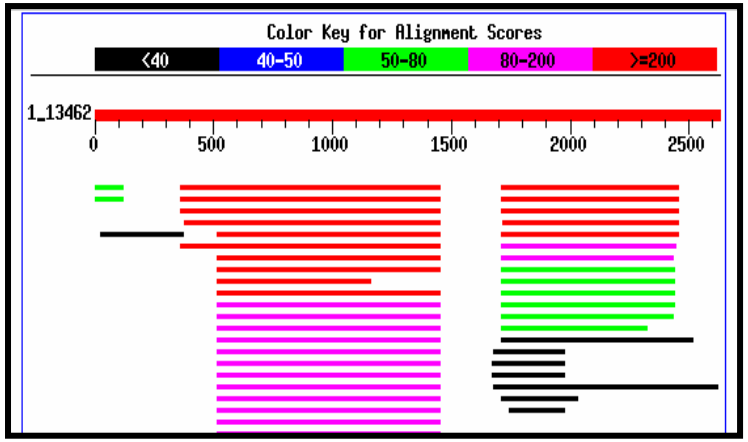
```
>gi|19548716|gb|AAL90755.1| adenosine deaminase [Mus musculus]
MAQTPAFNKPKVELHVHLDGAIKPETILYFGKKRGIALPADTVEELRNIIGMDKPLSLPGFLAKFDYYMP
VIAGCREAIKRIAYEFVEMKAKEGVVYVEVRYSPHLLANSKVDPMPWNQTEGDVTPDDVVDLVNQGLQEG
EQAFGIKVRSILCCMRHQPSWSLEVLELCKKYNQKTVVAMDLAGDETIEGSSLFPGHVEAYEGAVKNGIH
RTVHAGEVGSPEVVREAVDILKTERVGHGVHTIEDEALYNRLLKENMHFEVCPWSSYLTGAWDPKTTHAV
VRFKNDKANYSLNTDDPLIFKSTLDTDYQMTKKDMGFTEEEFKRLNINAAKSSFLPEEEKKELLERLYRE
YQ

>gi|15831585|ref|N                    Fasta format                    nerichia coli O157:H7]
MIDTTLPLTDIHRHLDGA                                                    DLVSFLTKLDWGVKVL
ASLDACRRVAFENIEDAARNGLHYVELRFSPGYMAMAHQLPVAGVVEAVIDGVREGCRTFGVQAKLIGIM
SRTFGEAACQQELEAFLAHRDQITALDLAGDELGFPGSLFLSHFNRARDAGWHITVHAGEAAGPESIWQA
IRELGAERIGHGVKAIEDRALMDFLAEQQIGIESCLTSNIQTSTVAELAAHPLKTFLEHGIRASINTDDP
GVQGVDIIHEYTVAAPAAGLSREQIRQAQINGLEMAFLNAEEKRALREKVAAK
```

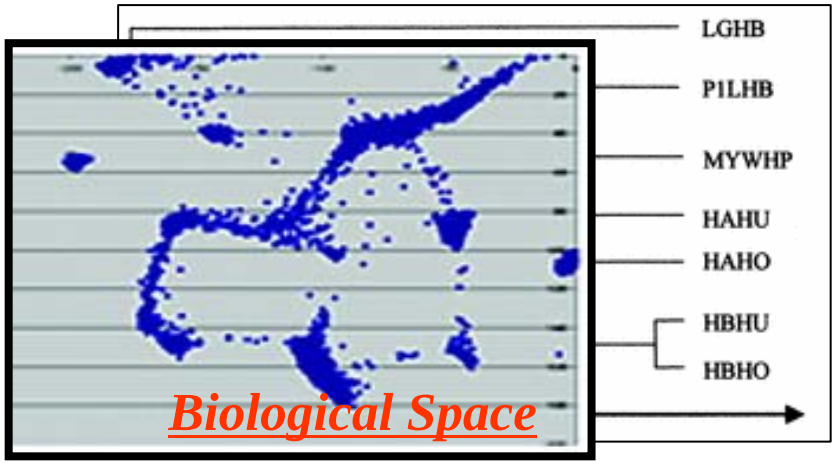
Similarity matrix

	HAHU	HBHU	HAHO	HBHO	MYWHP	P1LHB	LGHB
HAHU							
HBHU	21.1						
HAHO	32.9	19.7					
HBHO	20.7	39.0	20.4				
MYWHP	11.0	9.8	10.3	9.7			
P1LHB	9.3	8.6	9.6	8.4	7.0		
LGHB	7.1	7.3	7.5	7.4	7.3	4.3	

Alignment (ex. Blast…)



Classification



ケモインフォマティクス
構造解析

Structure

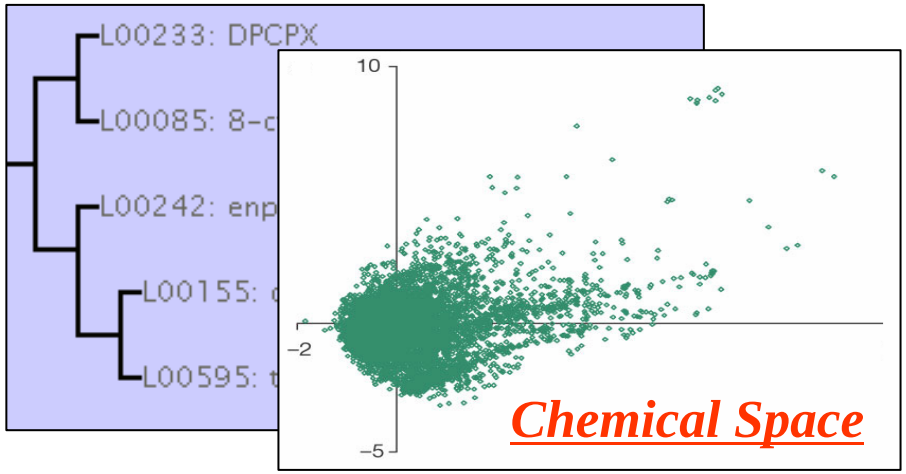
OC(=O)C(N)CC1=CC=C(O)C=C1

```
5  4  0     0  0            999 V2000
   -0.1276    0.2621    0.0000 C  0  0  0  0  0  0
    0.5552   -0.1862    0.0000 C  0  0  0  0  0  0
   -0.8552   -0.1483    0.0000 O  0  0  0  0  0  0
   -0.1552    1.0931    0.0000 O  0  0  0  0  0  0
    0.5793   -1.0207    0.0000 N  0  0  0  0  0  0
  1  2  1  0     0  0
  1  3  1  0     0  0
  1  4  2  0     0  0
  2  5  1  0     0  0
M  END
```

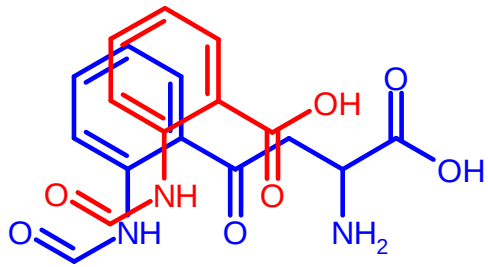
Distance matrix

	I1	I2	I3	I4	I5	I6
I2	1					
I3	7	5				
I4	13	6	16			
I5	10	0	7	5		
I6	9	9	12	13	12	
I7	11	20	10	9	14	8

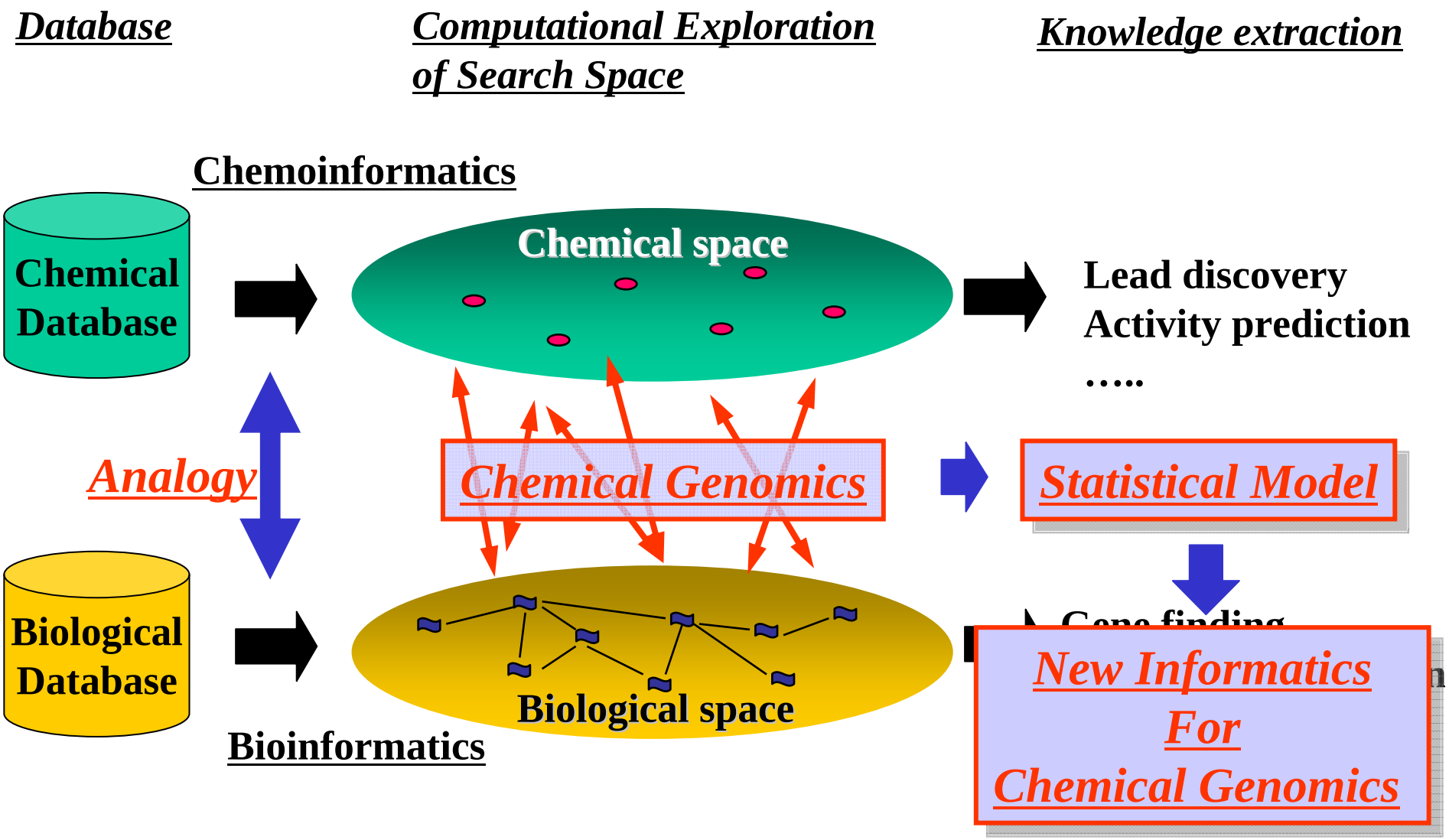
Classification



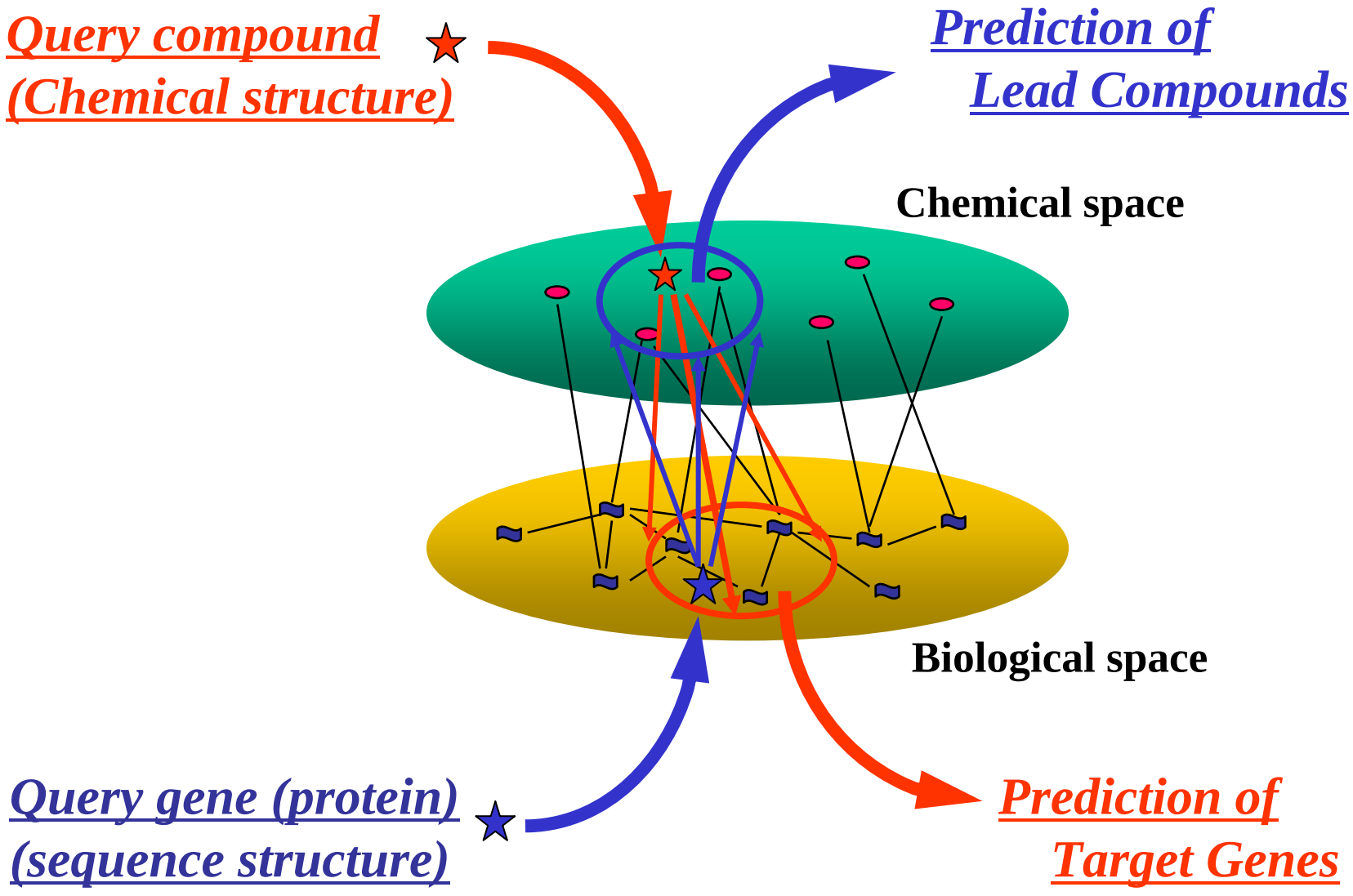
Structure comparison



ケミカルゲノミクスとインフォマティクス

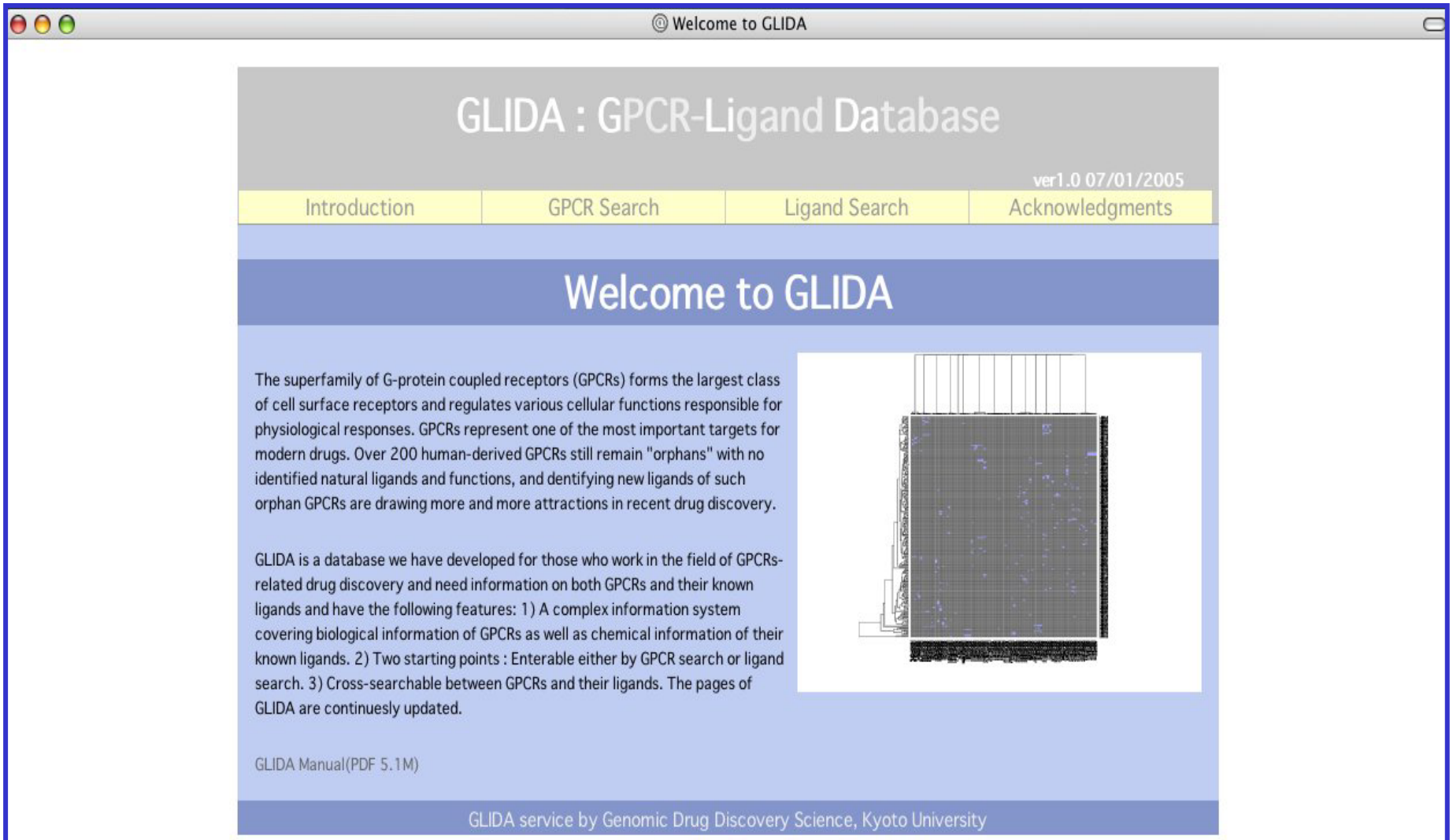


In silico スクリーニング



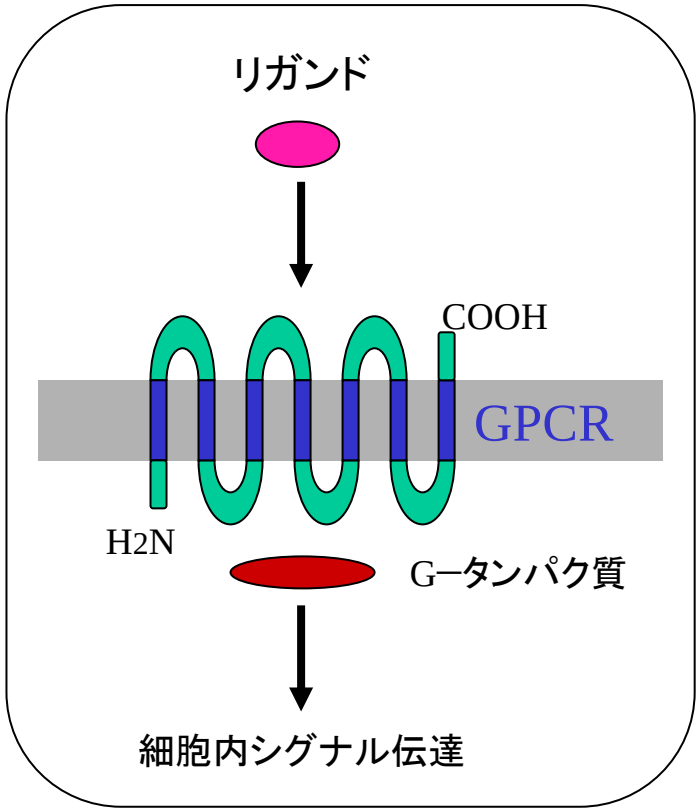
GLIDA: GPCR-Ligand Database

http://pharminfo.pharm.kyoto-u.ac.jp/services/glida

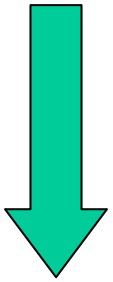


***Nucleic Acids Research*, 2006 Database issue, D673-677**

Motivation (Why GPCRs?)



- 医薬品の約50%がGPCRを標的としている
- ヒトでは約1000のGPCRが予測されており、約700がリガンド未知のオーファンGPCR
- GPCRとリガンドとの相互作用研究は、創薬において非常に重要



・公共のGPCR－リガンド相互作用データベースの開発
・ケモゲノミクスデータのマイニング手法（In silico screeing手法）の開発

GPCRとリガンドの相互作用情報

GLIDA : GPCR-Ligand Database

ver1.01 09/16/2005

Introduction | GPCR Search | Ligand Search | Acknowledgments

PE2R2_HUMAN

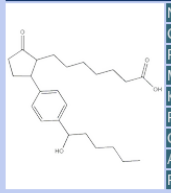
1. General information

PE2R2_HUMAN	
Gene Name	PTGER2
Family	Prostaglandin E2/D2 subtype EP2
GPCRDB ID	PE2R2_HUMAN
UniProt ID	P43116
IUPHAR ID	2.1:PG:3:EP2
Entrez Gene ID	5732
KEGG ID	hsa:5732
Sequence	PE2R2_HUMAN.fasta

Similarity Search | ALL

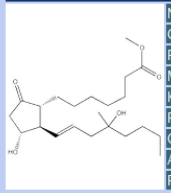
2. Ligand information

L00098



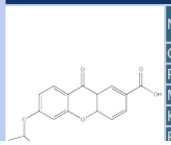
Name(s)	AH13205
Cas No	
Formula	C24 H36 O4
Mol. Wt.	388.544
KEGG ID	
PubChem SID	
ChEBI ID	
Activity	Agonist
Reference	PMID:10376989 PMID:3006854 PMID:8016385

L00406



Name(s)	misoprostol
Cas No	59122-46-2
Formula	C22 H38 O5
Mol. Wt.	382.537
KEGG ID	C07227
PubChem SID	9436 65384 183429
ChEBI ID	
Activity	Agonist
Reference	IUPHAR:2418

L00101



Name(s)	AH6809 6-isopropoxy-9-oxoxanthene-2-carboxylic acid
Cas No	33458-93-4
Formula	
Mol. Wt.	
KEGG ID	C07227
PubChem SID	696150

GLIDA : GPCR-Ligand Database

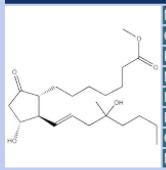
ver1.01 09/16/2005

Introduction | GPCR Search | Ligand Search | Acknowledgments

L00406

1. General information

L00406



Name(s)	misoprostol
Cas No	59122-46-2
Formula	C22 H38 O5
Mol. Wt.	382.537
KEGG ID	C07227
PubChem SID	9436 65384 183429
ChEBI ID	
MOLfile	L00406.mol

Similarity Search | ALL

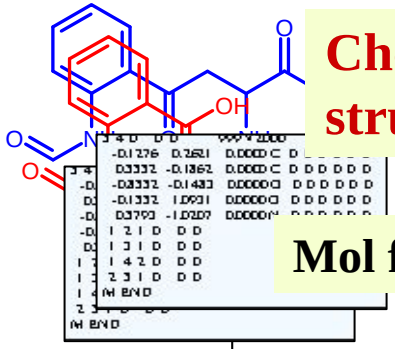
2. GPCR information

PE2R2_HUMAN	
Gene Name	PTGER2
Family	Prostaglandin E2/D2 subtype EP2
GPCRDB ID	PE2R2_HUMAN
UniProt ID	P43116
IUPHAR ID	2.1:PG:3:EP2
Entrez Gene ID	5732
KEGG ID	hsa:5732
Activity	Agonist
Reference	IUPHAR:2418

PE2R2_MOUSE	
Gene Name	Ptger2,Ptgerep2
Family	Prostaglandin E2/D2 subtype EP2
GPCRDB ID	PE2R2_MOUSE
UniProt ID	Q62053
IUPHAR ID	2.1:PG:3:EP2
Entrez Gene ID	19217
KEGG ID	mmu:19217
Activity	Agonist
Reference	IUPHAR:2418

PE2R3_HUMAN	
Gene Name	PTGER3
Family	Prostaglandin E2 subtype EP3
GPCRDB ID	PE2R3_HUMAN
UniProt ID	P43115
IUPHAR ID	2.1:PG:4:EP3
Entrez Gene ID	5733
KEGG ID	hsa:5733
Activity	Agonist

GPCR/リガンドのクラスタリング

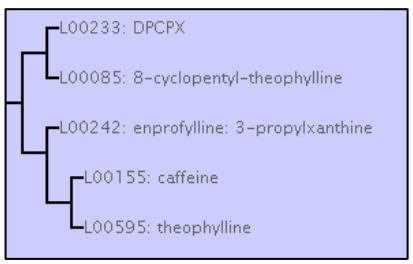


Chemical structures

Mol files

Profiles based on KEGG atom types

Similarity matrix

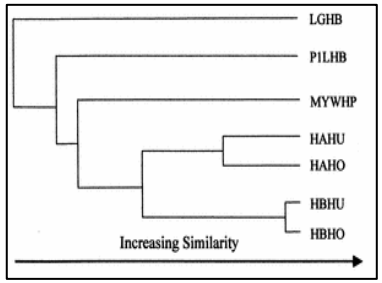


Classification with tree representation

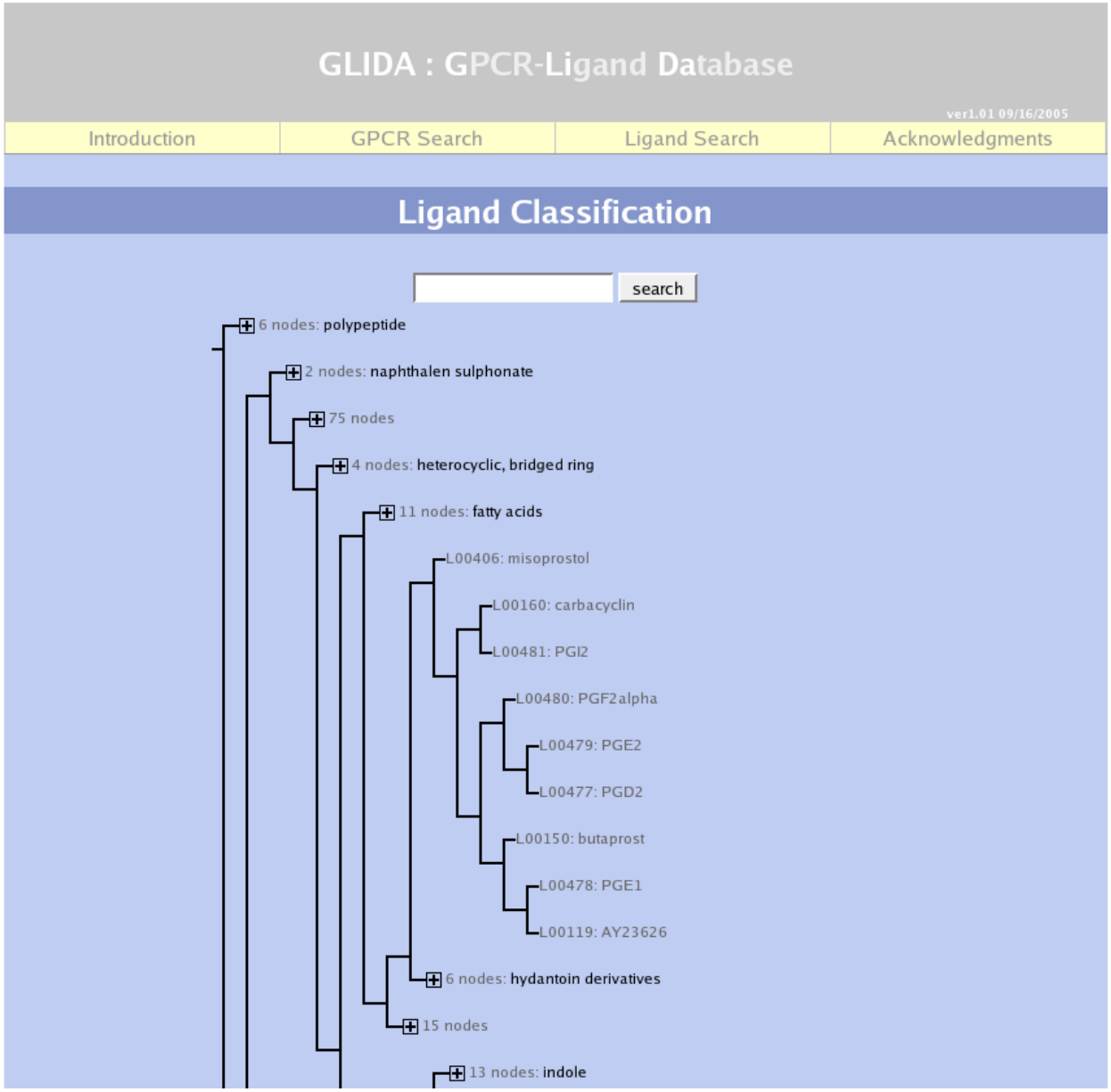
GPCR sequence

Profiles based on (k,m)-spectrum method

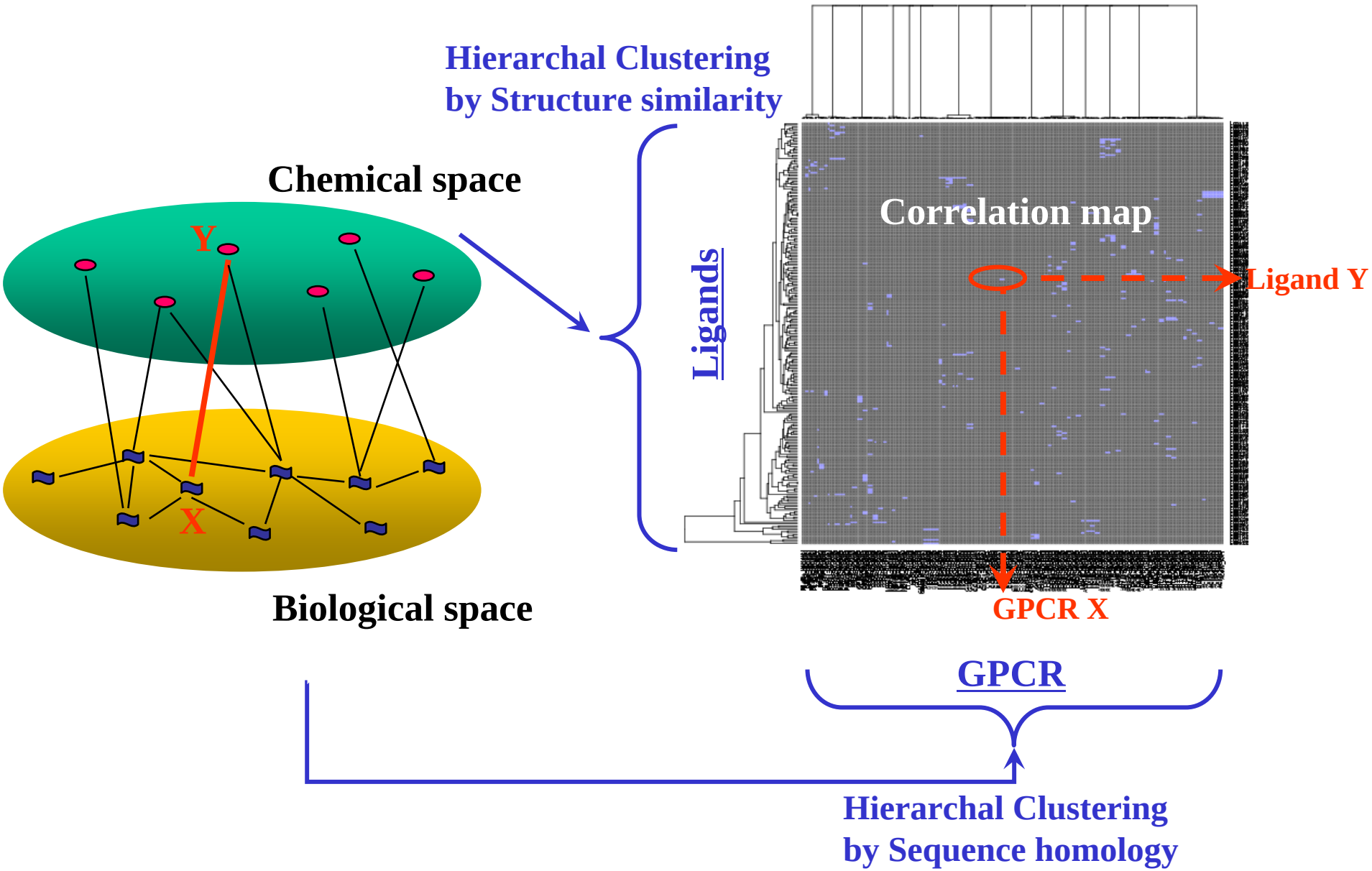
Similarity matrix



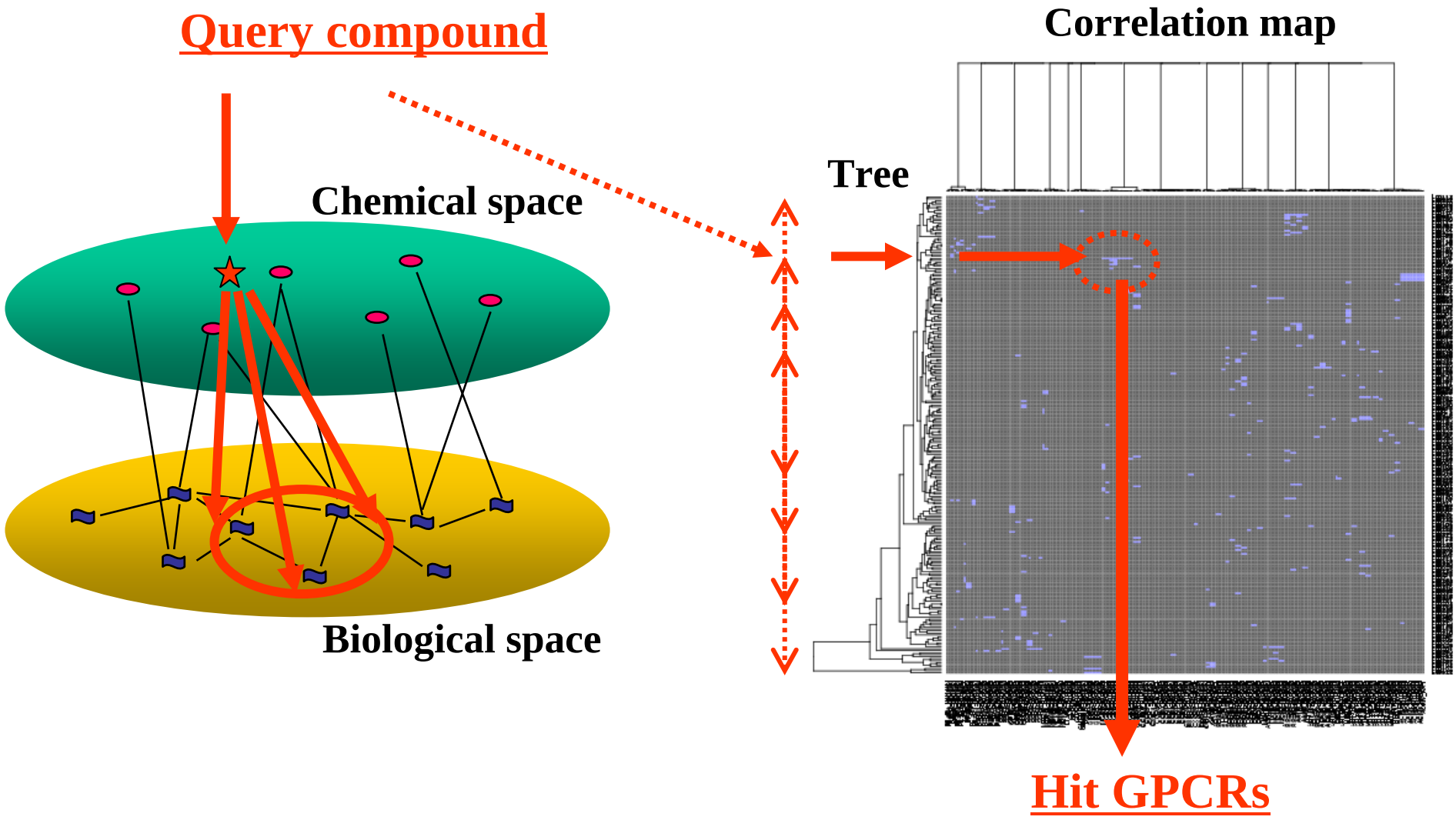
例）リガンド分類


GLIDA : GPCR-Ligand Database
ver1.01 09/16/2005
Introduction
GPCR Search
Ligand Search
Acknowledgments
Ligand Classification
search
6 nodes: polypeptide
2 nodes: naphthalen sulphonate
75 nodes
4 nodes: heterocyclic, bridged ring
11 nodes: fatty acids
L00406: misoprostol
L00160: carbacyclin
L00481: PGI2
L00480: PGF2alpha
L00479: PGE2
L00477: PGD2
L00150: butaprost
L00478: PGE1
L00119: AY23626
6 nodes: hydantoin derivatives
15 nodes
13 nodes: indole

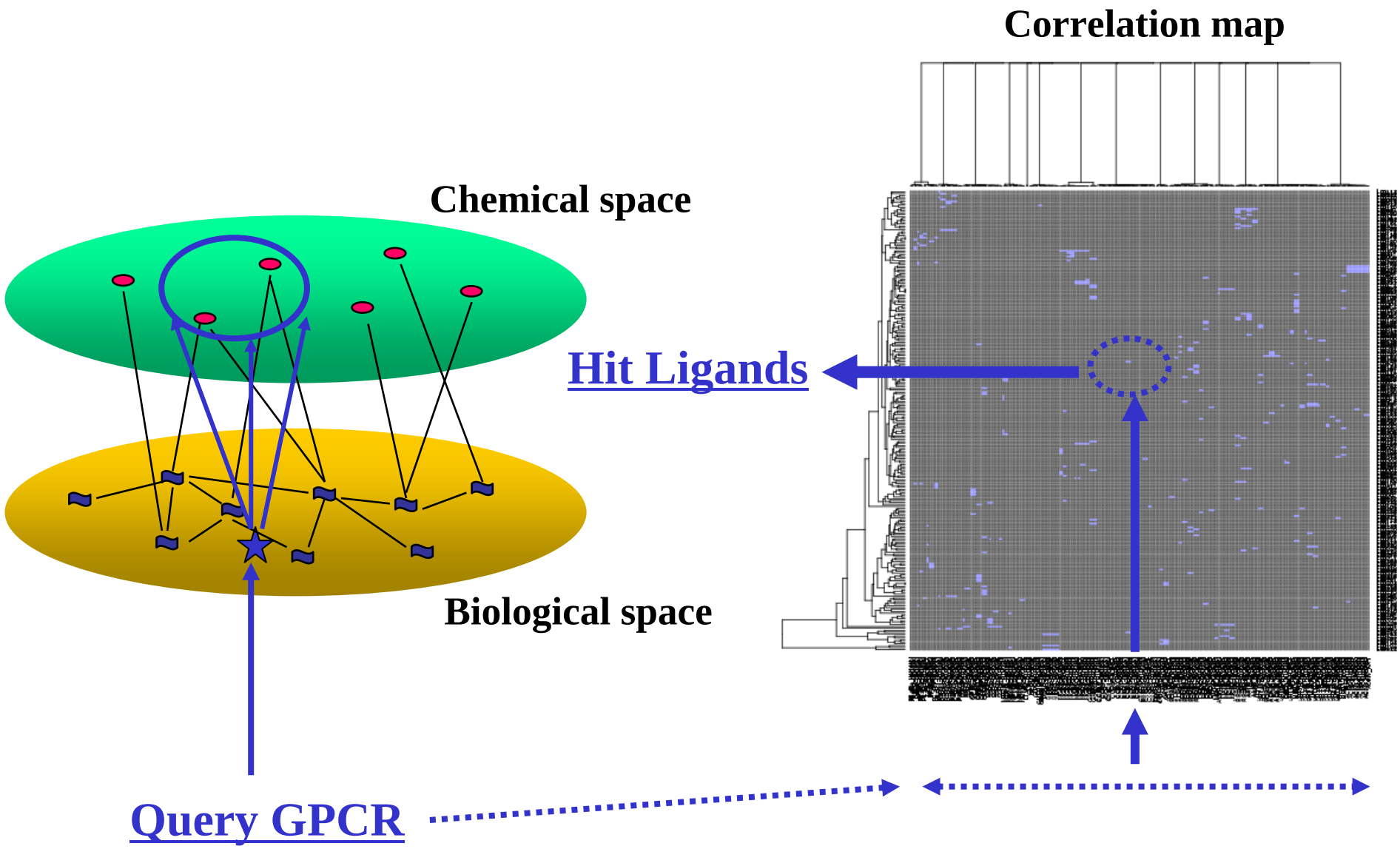
GPCR-Ligand Space of GLIDA



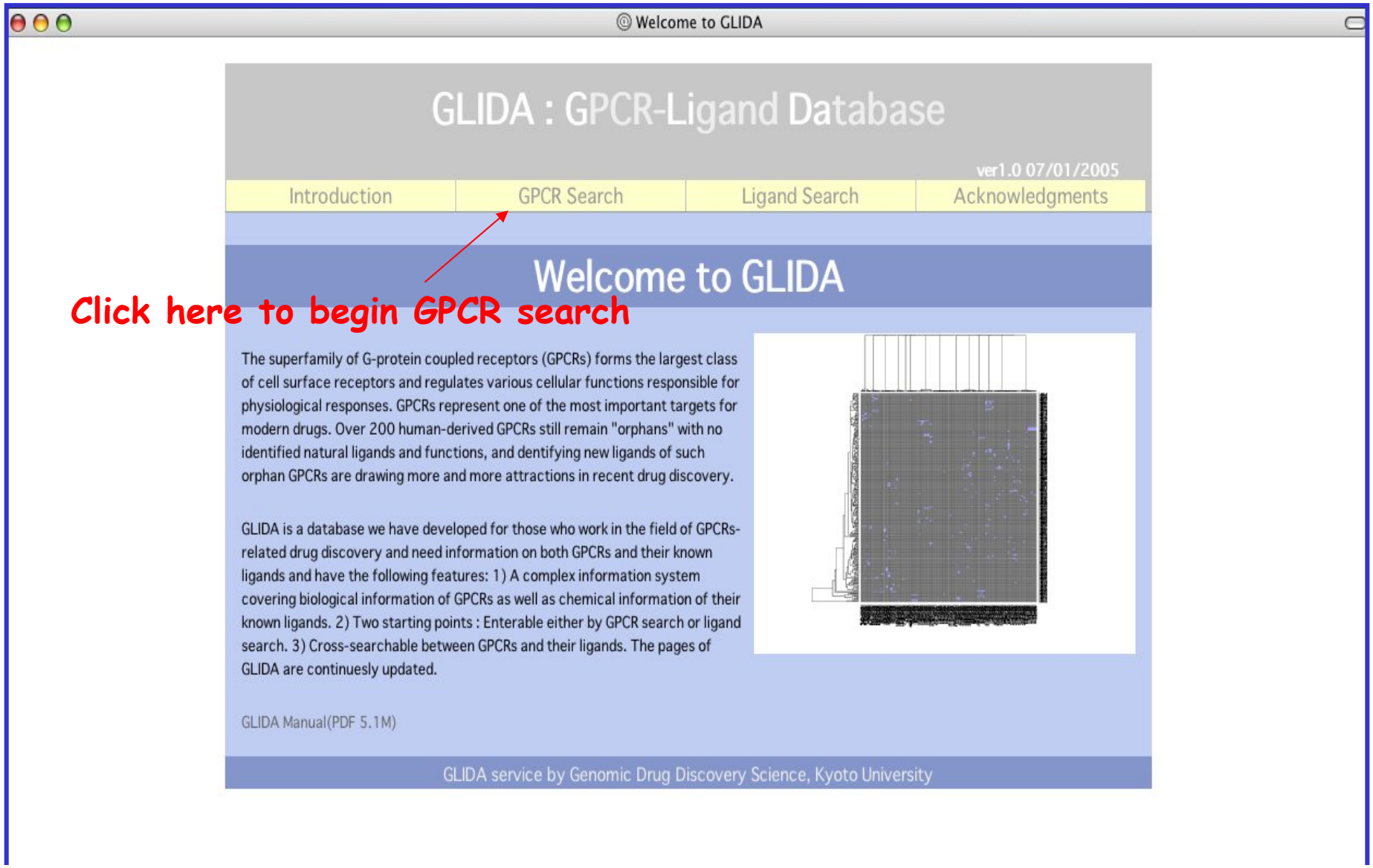
In silico Screening by GLIDA



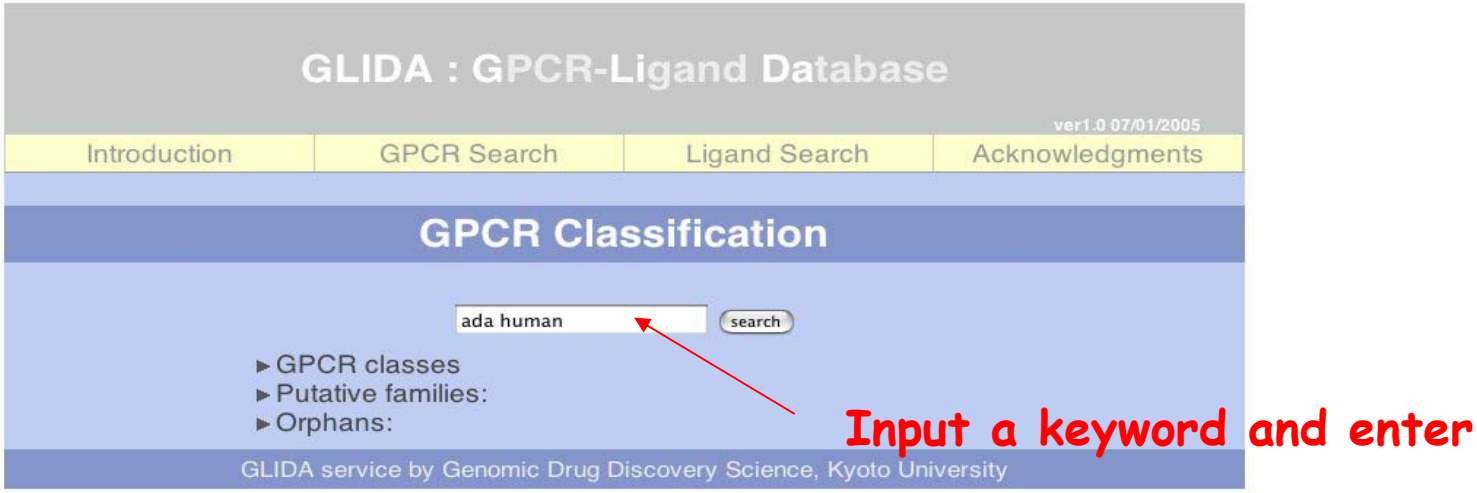
In silico Screening by GLIDA



Execution of GLIDA
(From a query GPCR to target Ligands)



Keyword search of GPCR

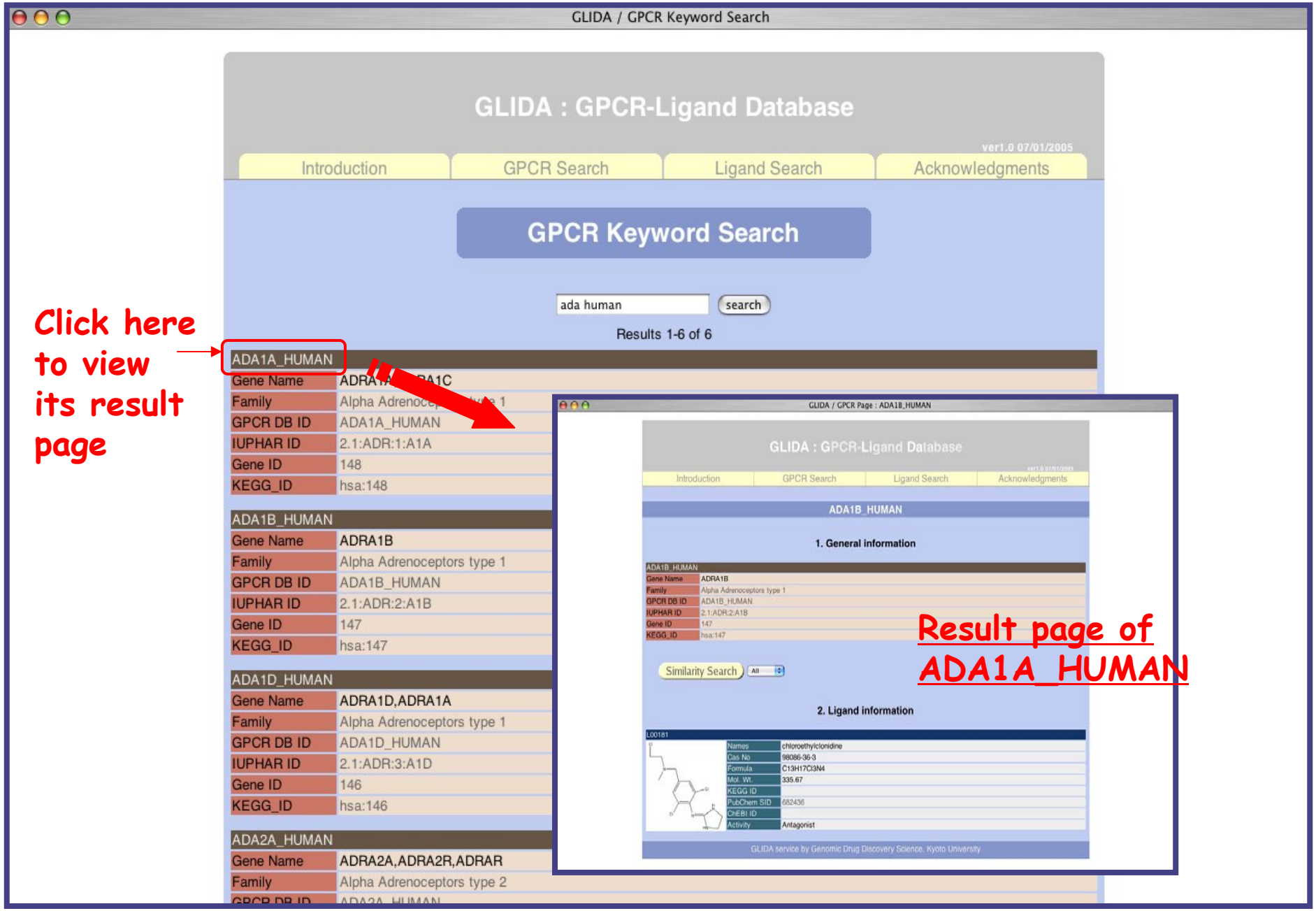


Input a keyword and enter

* Examples of search

HTR1A, DRD2_HUMAN(gene names), P14416 (Swiss-Prot ID),
Angiotensin (GPCRDB Family names),
oncogene, smell, hormone (gene ontology term or SwissProt functional annotation)

Result of keyword search



Click here to view its result page

Result page of ADA1A_HUMAN

Similarity search & binding prediction

GLIDA / GPCR Page : ADA1A_HUMAN

GLIDA : GPCR-Ligand Database

ver1.0 07/01/2005

Introduction | GPCR Search | Ligand Search | Acknowledgments

ADA1A_HUMAN

1. General information

ADA1A_HUMAN	
Gene Name	ADRA1A,ADRA1C
Family	Alpha Adrenoceptors type 1
GPCR DB ID	ADA1A_HUMAN
IUPHAR ID	2.1:ADR:1:A1A
Gene ID	148
KEGG_ID	hsa:148

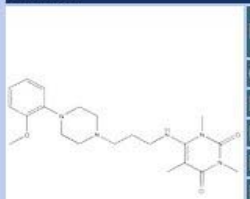
Click here to start calculation

Similarity Search | All

Analytical report

2. Ligand information

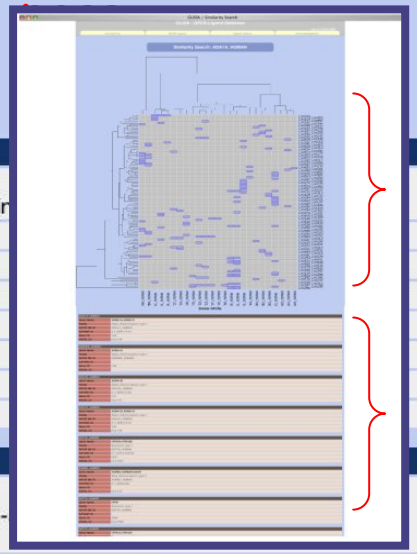
L00080



Names	5-methylurapidil 5-methyl-6[[3-[4-(2-methoxyphenyl)-1-piperazin
Cas No	34661-85-3
Formula	C21 H31 N5 O3
Mol. Wt.	401.508
KEGG ID	
PubChem SID	691790
ChEBI ID	
Activity	Antagonist

L00090

Names	A61603 N-[5-(4,5-dihydro-1H-imidazol-2-yl)-2-hydroxy- yl]methane sulfonamide

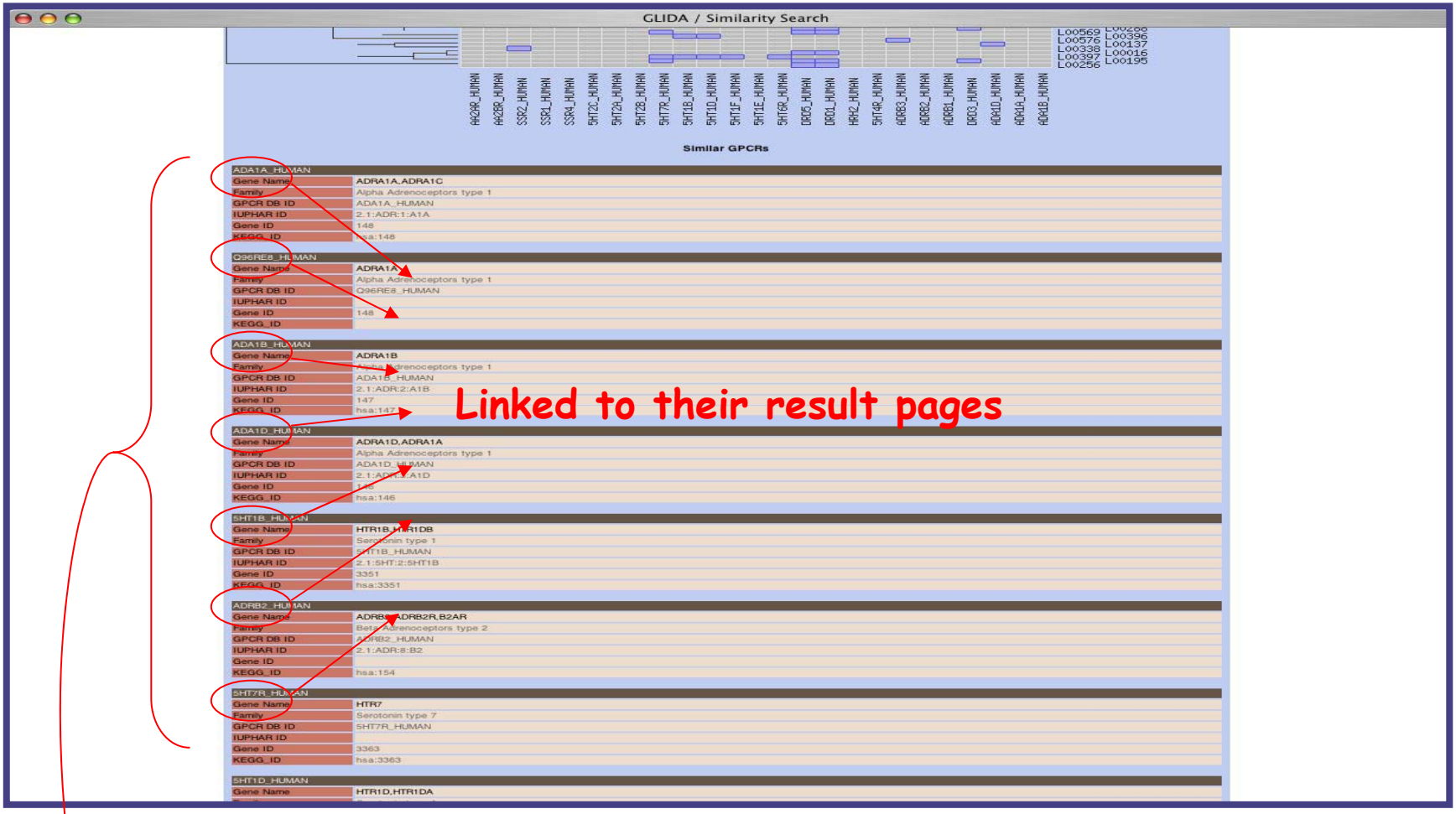


Information of the ligand of the GPCR

GPCR-ligand correlation map

Similar entries list

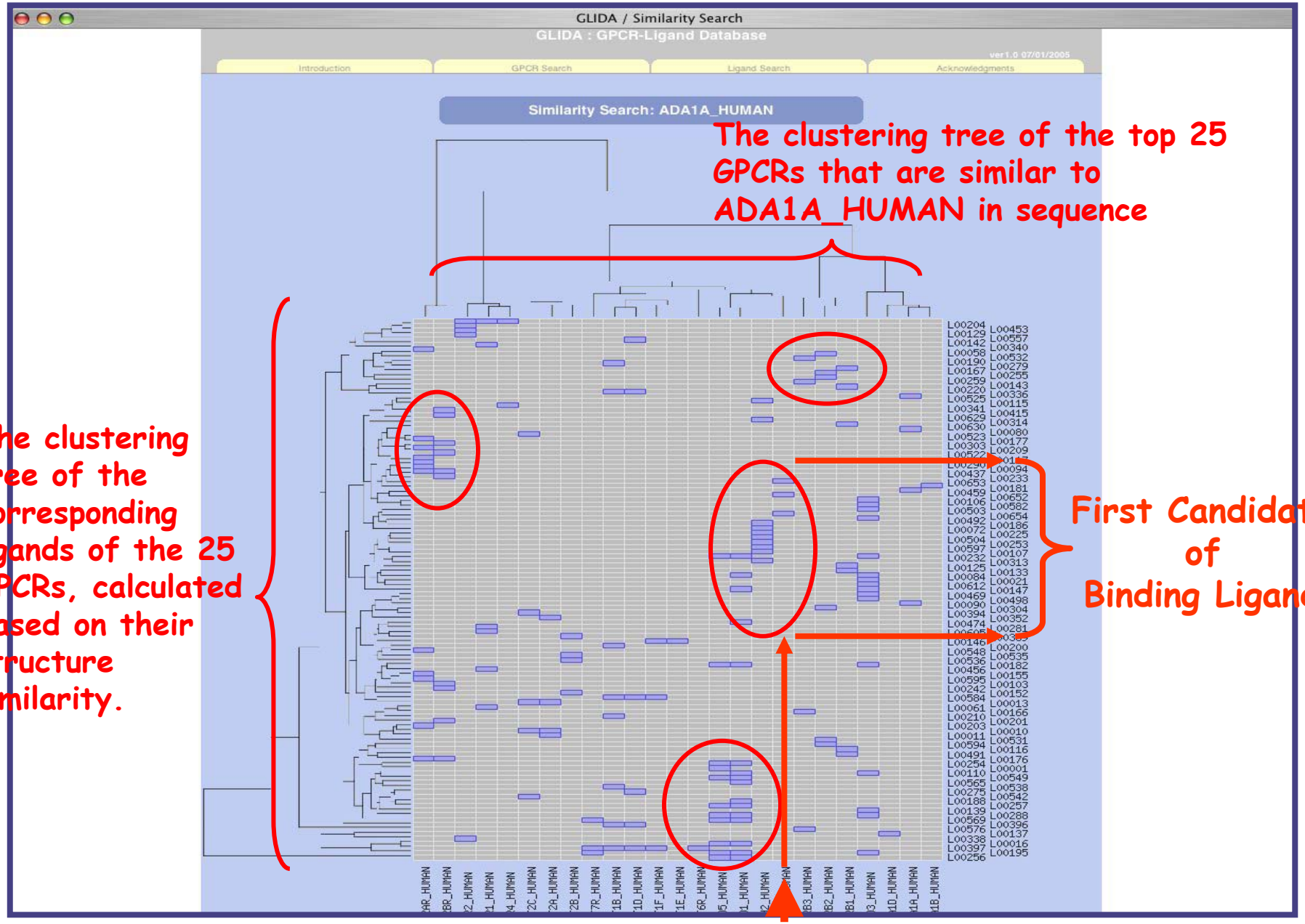
Result of Similarity search



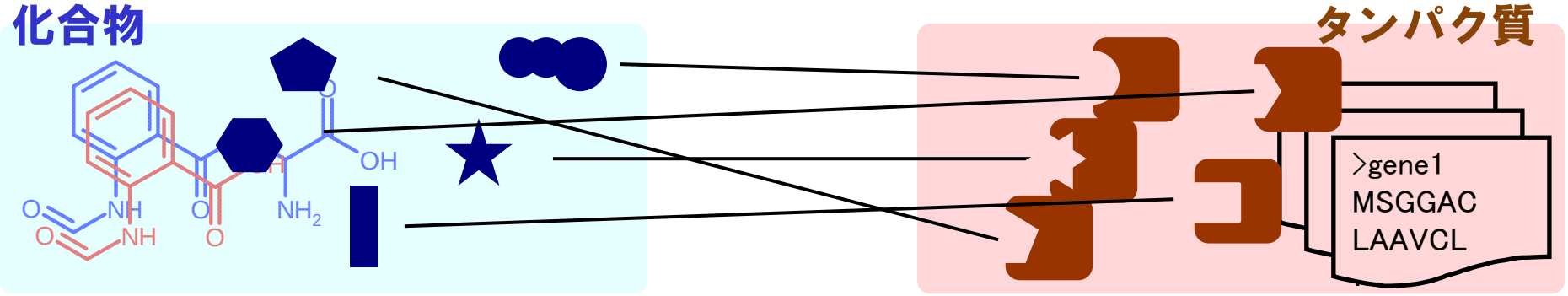
Linked to their result pages

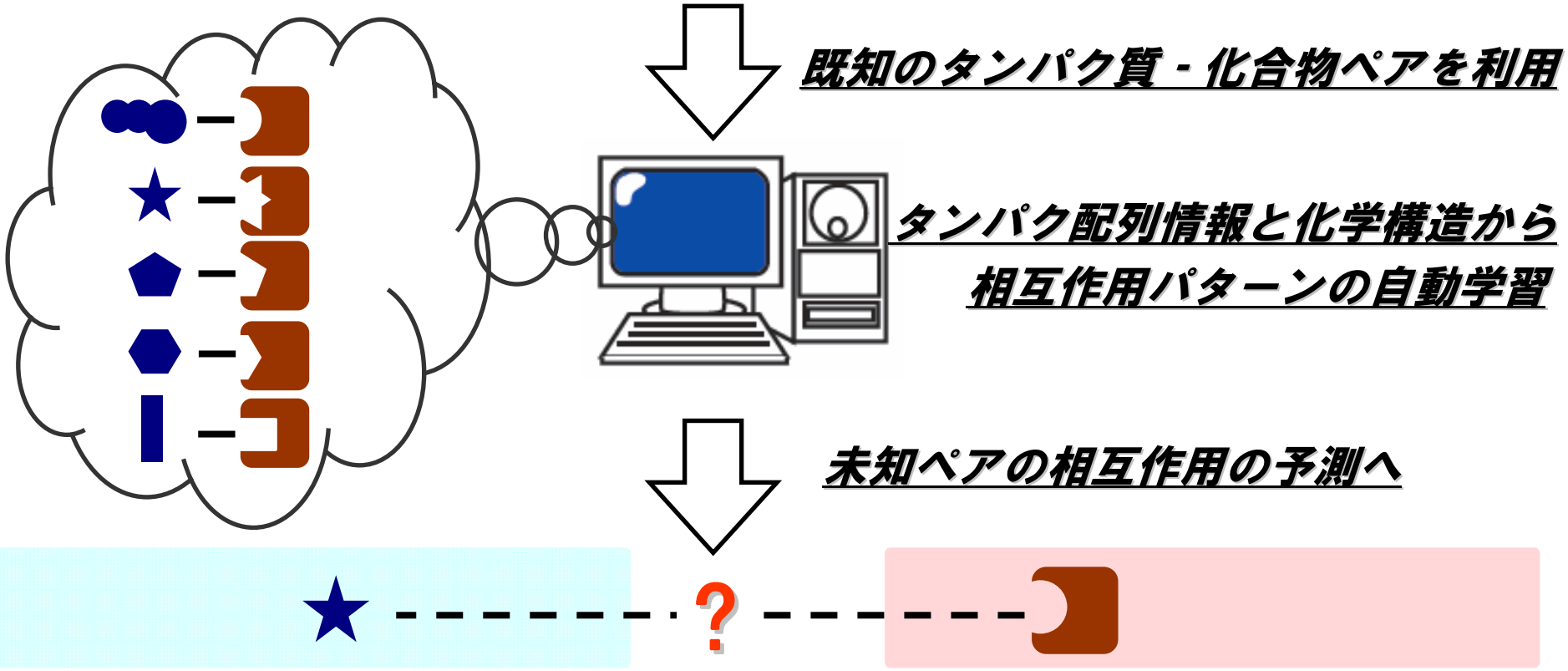
The top 25 GPCRs which are most similar to the selected GPCR (i.e.,ADA1A_HUMAN in this example) are displayed in this report page

Result of binding prediction: GPCR-Ligand correlation map

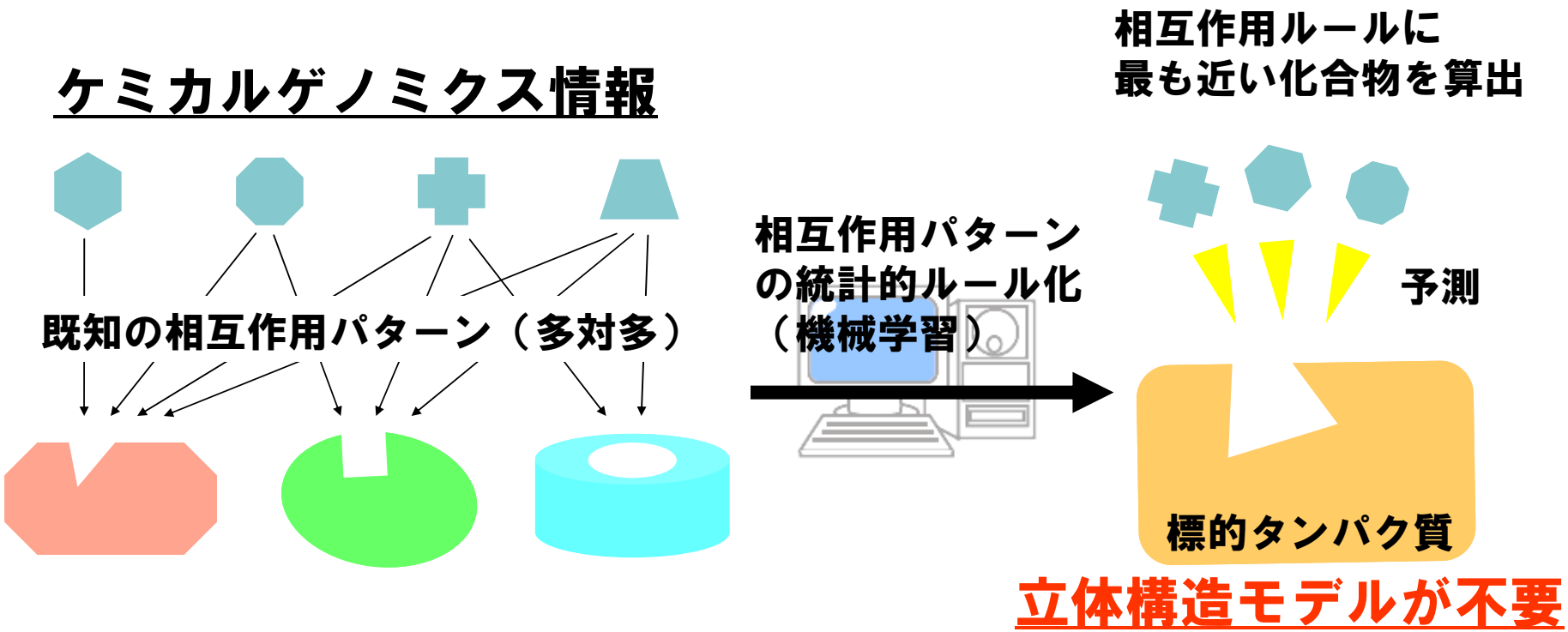


機械学習による
タンパク質 - 化合物相互作用予測





相互作用マシンラーニング法



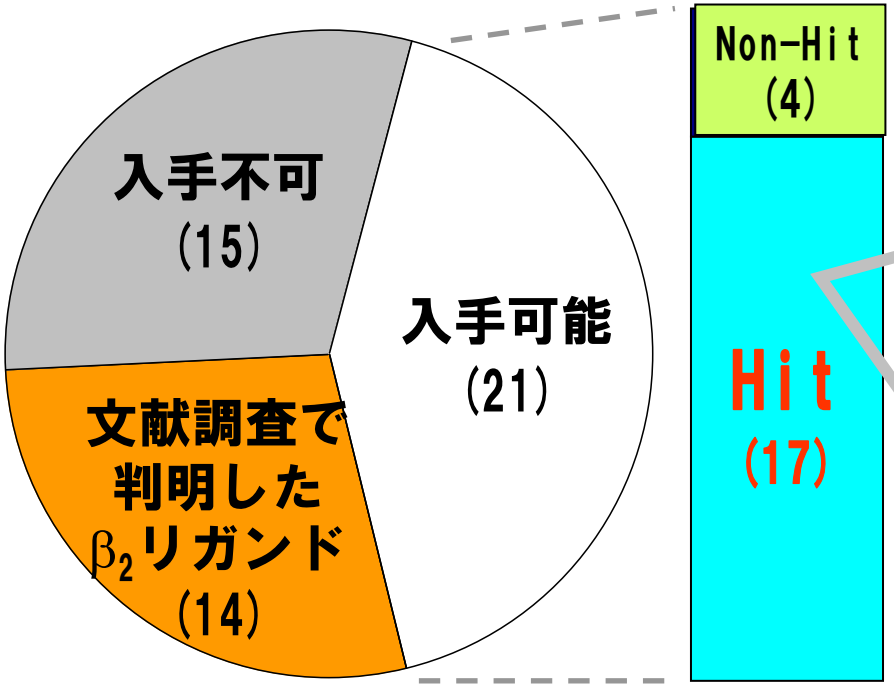
- 標的タンパク質の立体構造は不要（膜タンパクの場合、一次配列のみでの高精度予測が検証済み）
- 相互作用関係を優先し、化学構造の自由度を許容するため、新規骨格の発見の可能性が高い
- 計算時間が短時間で済み、計算コストが非常に良い

相互作用マシンラーニング法による β2-アドレナリン受容体リガンド予測の結果

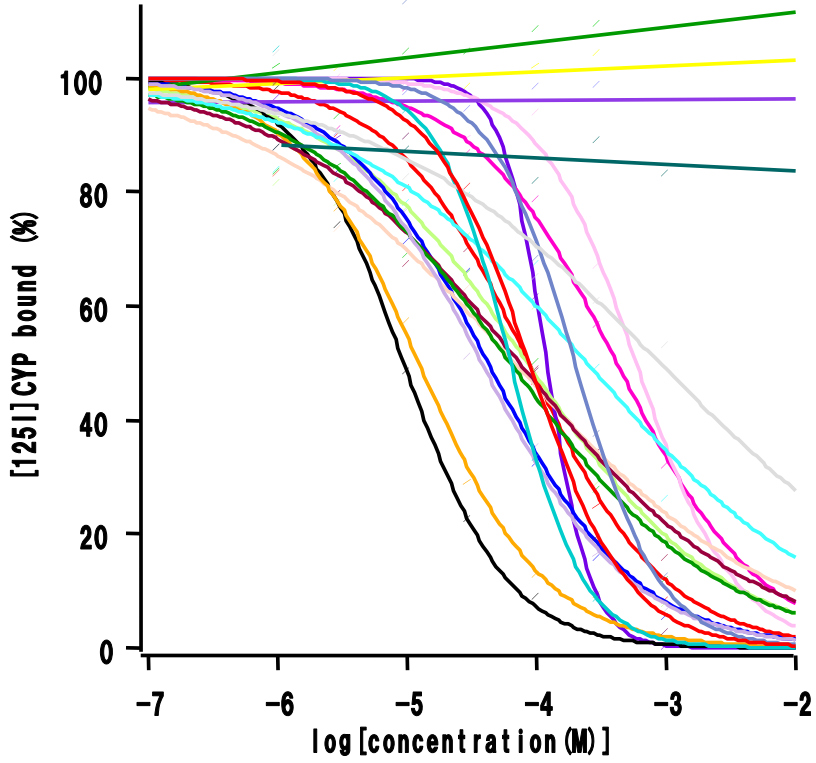
843種類の化合物
との相互作用の有無を予測

⇩

予測スコアTop50の化合物



In vitro 結合阻害実験



ヒット率 ： 81.0% (17/21)

トータルヒット率（実験＋文献調査） ： 89 % (31/35)

相互作用マシンラーニング法の予測実績

日刊工業新聞
2007. 3. 26

バイオ関連企業の動向

バイオグリッドセンター関西

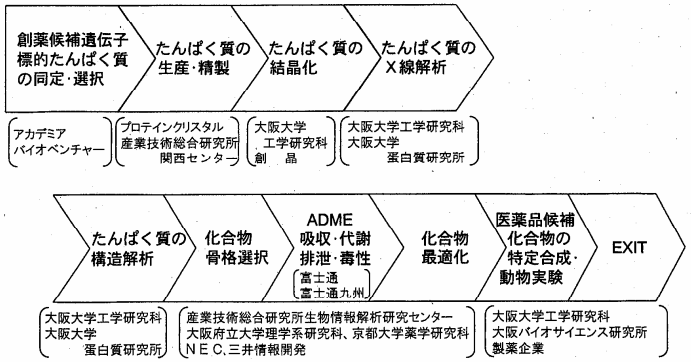
バイオグリッドセンター関西（大阪府豊中市、下條真司理事長、06・6873・2156）は「創薬バリューチェイン」プロジェクト（代表＝坂田恒昭理事）を進めている。同プロの提案「インシリコでの創薬手法の確立とその実証研究」は文科省の大阪北部（彩都）地域知的クラスター創成事業に採択された。コンピューター技術を用いて、医薬品候補化合物の発見（探索）および候補化合物の詳細設計（最適化）がそれぞれ可能かというテーマを掲げている。これらのテーマについて実験系とコンピューター系のチームにより実証研究を進めてきた。

コンピューターによる

創薬バリューチェインプロジェクトを進行

医薬品候補化合物探索の道探る

創薬バリューチェインの概念



候補化合物の発見の取り組みとしては、京都大学大学院薬学研究科の奥野恭史助教授が開発したコンピューターを用いた構造未知のたんぱく質に対する薬剤候補化合物の探索手法「相互マシンラーニング法」がある。

熱力学的シミュレーションにより結合モードを探索するドッキングシミュレーションはたんぱく質の構造情報が必要になる。構造の動的な変動を正確にシミュレーションすることは難しく、予測結果の信頼性、再現性などに課題がある。特に膜たんぱく質は新薬開発の標的となるが、結晶化が難しく構造はほとんど分かっていない。このため構造が分からなくても、薬を作ることも求められている。

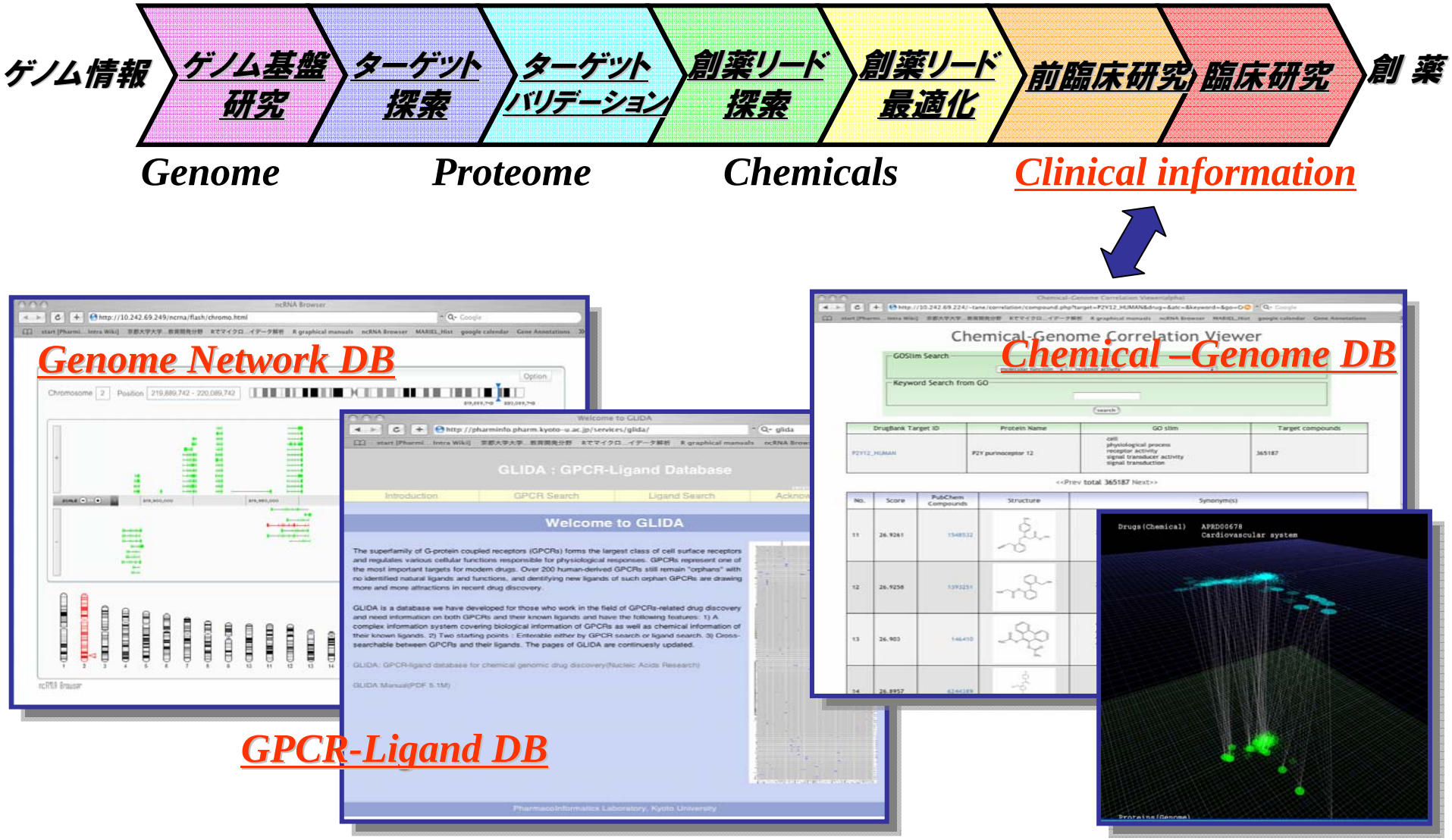
相互作用マシンラーニング法は化合物空間とたんぱく質空間を、それぞれ類似性に基づいて定義する。既知の多対多の相互作用パターンを統計的にルール化し、相互作用ルールに最も近い化合物を算出、予測する。同法では標的たんぱく質の立体構造は不要。膜たんぱく質の場合、一次配列のみでの高精度予測できることは検証済みだ。相互作用関係を優先し、化学構造の自由度を許容するため、新規骨格の発見の可能性が高い。計算時間が短時間で済むという利点がある。

実験では構造が未知の膜たんぱく質を用いた。同法を用い論文などの既知化合物の情報を用い、計算機で解析し約130の化合物を購入して生物活性実験した。すると、既に論文で発表されている化合物活性より1000倍の作用と10倍の作用を示す化合物を見つけた。1000倍の活性を示す化合物は、すでに薬として使えるナノモルオーダーの活性を有し、新薬開発に近づくことになる。また既知化合物と同等の活性を示す26個の化合物を探索することができている。

国際級クラスターの形成でわが国をけん引

1. 他のGPCR （１０μMオーダー化合物がヒット）
2. TRPタンパク （ナノモルオーダー化合物（1000倍の活性）がヒット）
3. マラリア標的ピリミジン合成酵素（１０％のヒット率）

Kyoto-Univ
Pharmaco-Informatics Navigation System



特願2006-147433
国際出願番号PCT/JP2006/312858

Zhu, S., Okuno, Y., et al., **Bioinformatics, 21(s2),** ii245-ii251, 2005
Okuno, Y. et al., **Nucleic Acids Research, Database issue,** D673-677 2006

統合薬学フロンティア教育センター
統合薬学教育開発分野
Department of PharmacoInformatics

http://pharminfo.pharm.kyoto-u.ac.jp/

（メリット）

- 世界一戦級の研究ができる
- 計算に強くなる（とりあえず、賢くみえる）
- 実験が肌に会わない人に最適
- 生き物の命を大切にする人に最適
- 常に人材不足であり、世界をリードする人材となれる
- 私の指導が受けられる

（デメリット）

- 特にないと思いますが、、
- 本格的なWet実験が出来ない
- デスクワークが続き不健康気味
- 秋葉系に間違えられる可能性があるかも