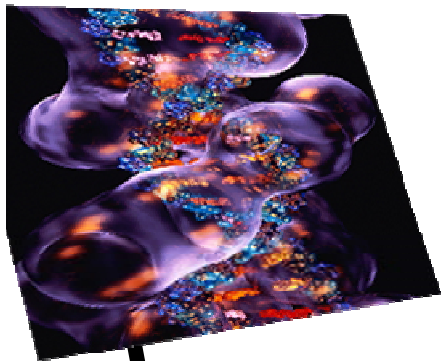




# GPGPU기반 AMBER-GPU 단백질 시뮬레이션 및 게놈 어셈블리 시스템

시스템바이오 대표기업  
(주)신테카바이오



데이터 쓰나미





IT-track

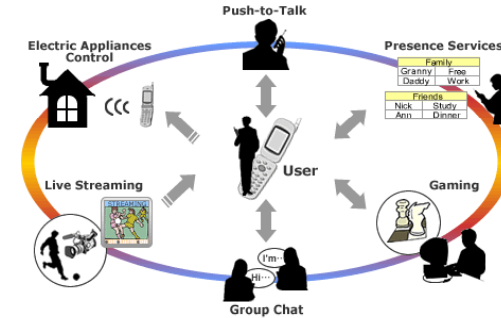
구글



클라우드

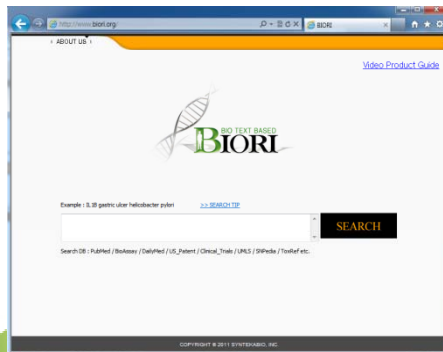


유비쿼터스

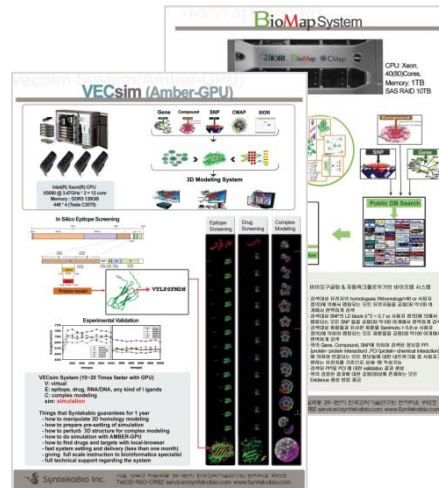


BIT-track

바이오구글



시스템바이오

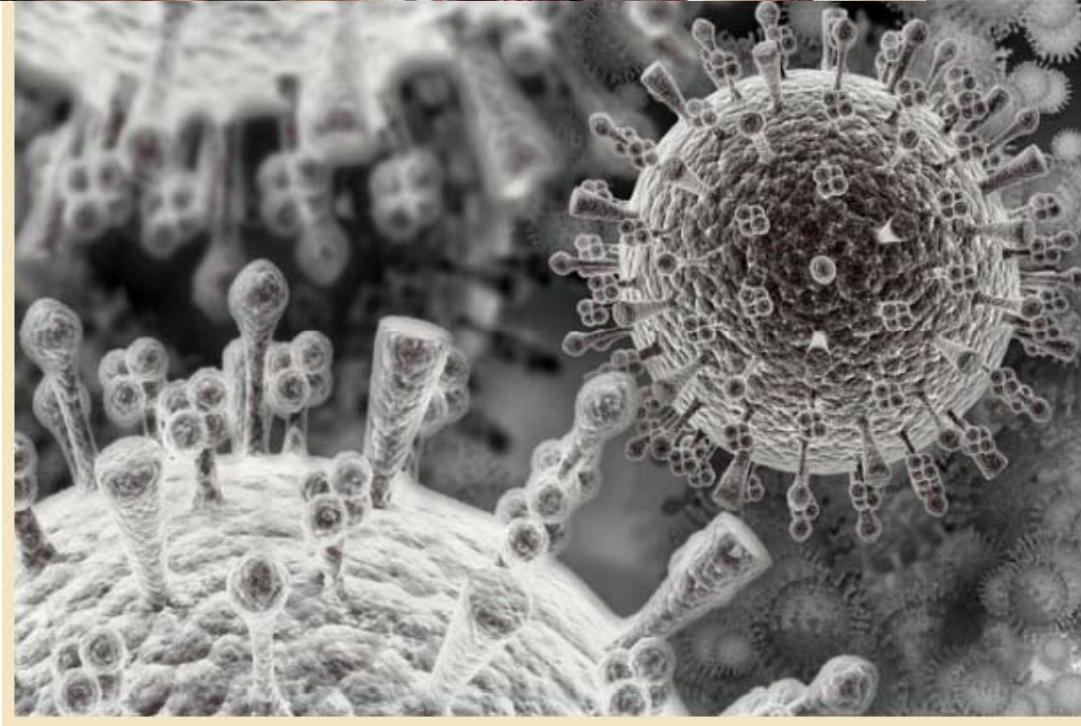
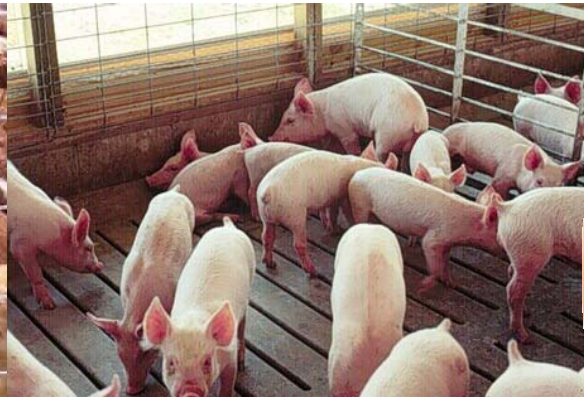


개인별 유전체





# 독감(FLU) ?





Korea Graphics World 2012

신종독감

백신  
바이오신약



글로벌 파마



차세대 시퀀서

2

**In Silico Epitope Screening**

**Experimental Validation**

**Calm System (10-20 Times faster with GPU)**  
 V: virtual  
 E: epitope, drug, RNA/DNA, any kind of ligands  
 C: complex modeling  
 sim: simulation

Things that Syntekbio guarantees for 1 year  
 - how to manipulate 3D homology modelling  
 - how to prepare pre-setting of simulation  
 - how to perturb 3D structure for complex modeling  
 - how to do simulation with AMBER-GPU  
 - how to find drugs and targets with local-browser  
 - fast system setting and delivery (less than one month)  
 - give full scale instruction to bioinformatics specialist  
 - full technical support regarding the system

Epitope Screening, Drug Screening, Complex Modeling

Intel(R) Xeon(R) CPU  
X5690 @ 3.47GHz \* 2 = 12 core  
Memory : DDR3 128GB  
448 \* 4 (Tesla C2075)

\*게놈어셈블러

1

Genome assembly software interface showing sequence alignment and assembly graphs.

\*바이오맵시스템

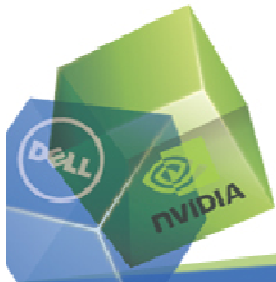
\*시스템바이오

\*Virus 게놈맵  
Genome Map

**BioMapSystem**

Genome Map

Web Programming, Info Integration, Public DB Search, User's Query



# NGS 게놈어셈블러 주요 툴

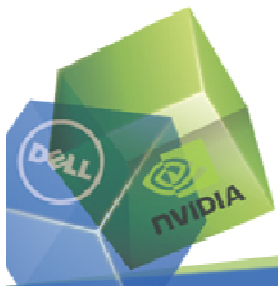
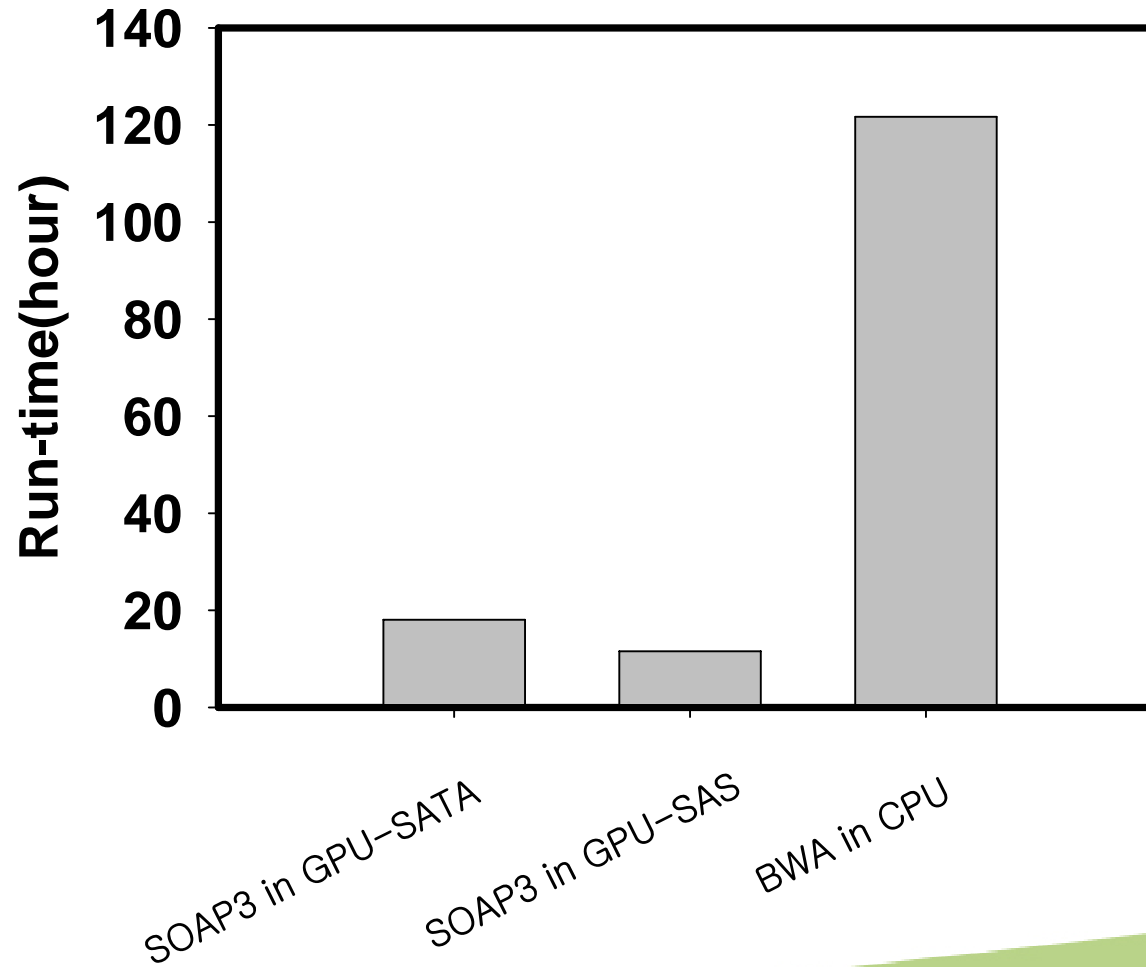
## < 해외 개발 현황 >

- 1) Mummer (NAR, 1999) 미국 동부 -> Genome vs. Genome alignment
- 2) SSAHA(Genome research, 2001) 유럽
- 3) Mummer(NAR, 2002) 미국 -> 논문 600편, utility별도
- 4) GMAP(Bioinformatics, 2005) , 미국 서부
- 5) Mummer-GPU (BMC bioinfo, 2007) 미국 -> Mummer와 같은 포맷\*
- 6) SOAP(Bioinformatics, 2008) 유럽/중국
- 7) \*MAQ (Genome research, 2008) 유럽/중국
- 8) Bowtie(Genome Biology,2009), 미국
- 9) BWA (Bioinformatics, 2009), 영국
- 10) \*BWA (Bioinformatics, 2010), 영국  
(Short Reads에 막강하지만, Long Reads에는 대안이 없음)
- 11) SOAP3-GPU 버전 공개됨. \*BWA와 동일 알고리즘 사용

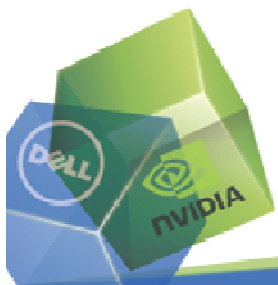
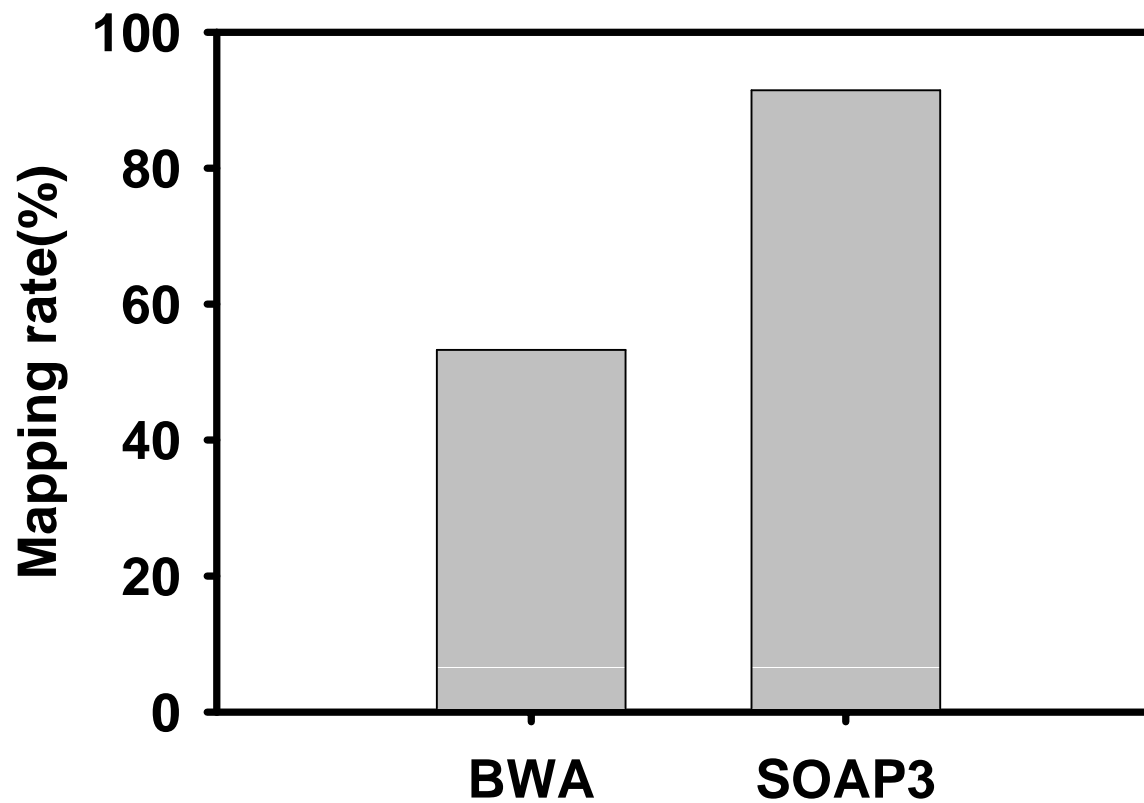
GPU 버전



# SOAP3 vs. BWA time test



## Mapping rate between SOAP3-GPU and BWA

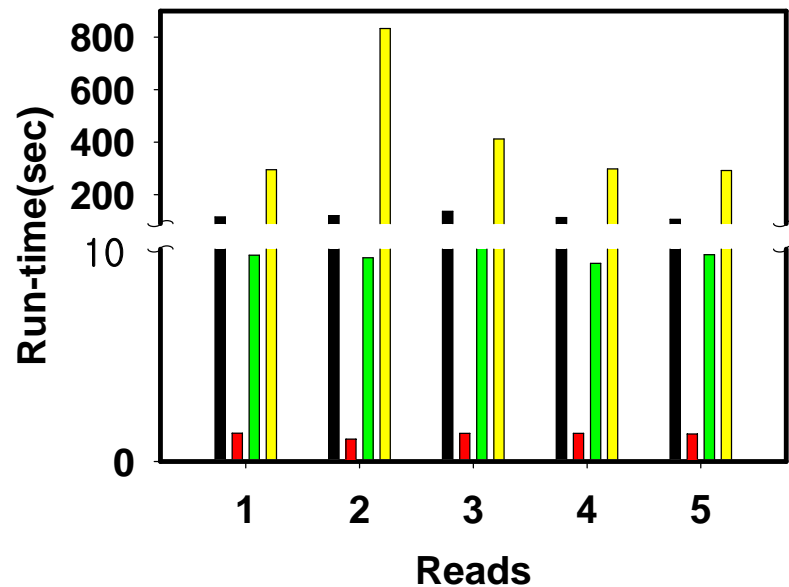


# SOAP3-GPU 테스트 결과

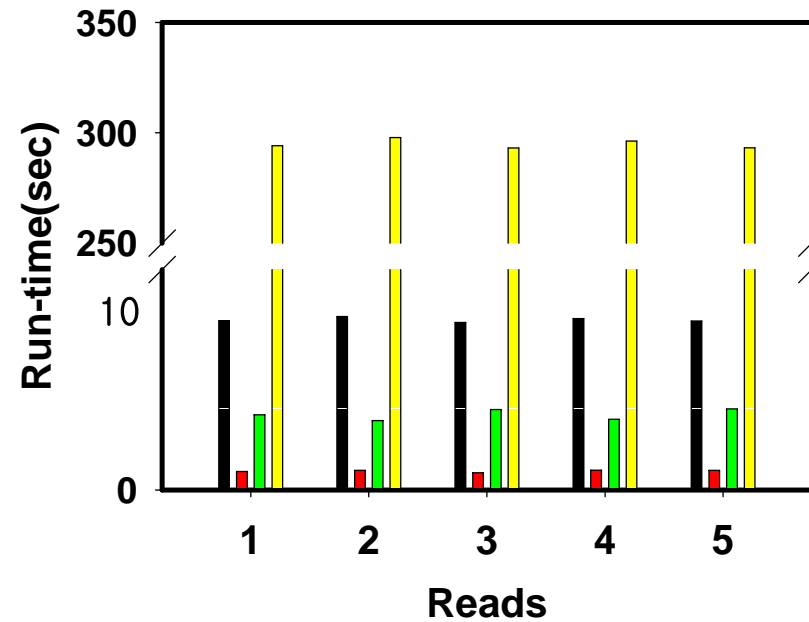
SOAP time test (II)

Time test at each step for short reads alignment

GPU-SATA



GPU-SAS



- Loading index into host
- Copying index into device(GPU)
- Alignment time
- Loading short reads from query file





## 신테카바이오 게놈어셈블러 프로그램 진행상황

순서	주요모듈	C	CUDA
1	Inverted Indexing of Reference sequence	완료	N/A
2	Query Record Indexing	완료	N/A
3	Mapping between Reference & Query	완료	N/A
4	Graph theory-based Clustering	완료	완료
5	Gapped alignment by Smith Waterman	완료	완료

N/A: CUDA 코드가 효율적이지 못함

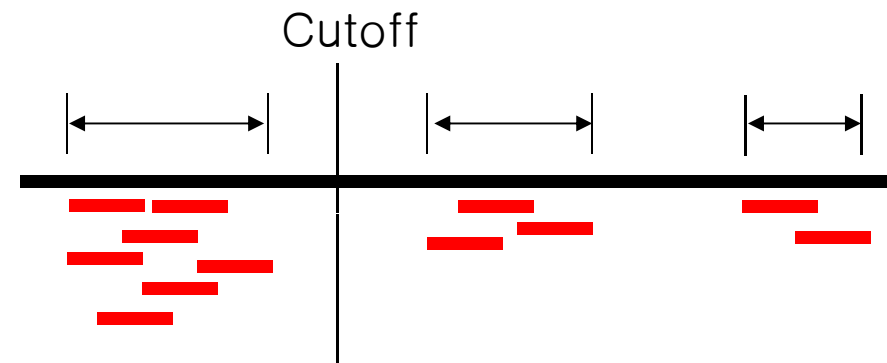


# Clustering by Graph theory (GPU version)

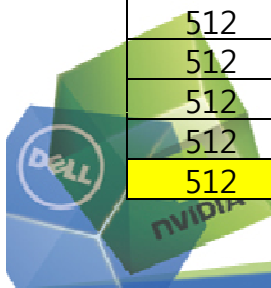
CUDA-GPU

nBlock	nThread	출력파일	sec(작동시간)
32	32	961M	2m33.702s
32	64	961M	2m2.097s
32	128	961M	1m47.007s
32	256	961M	1m48.129s
32	512	972M	1m43.742s
64	32	961M	2m1.741s
64	64	961M	1m47.360s
64	128	961M	1m46.301s
64	256	972M	1m46.070s
64	512	978M	1m42.180s
128	32	961M	1m54.738s
128	64	961M	1m47.348s
128	128	972M	1m48.987s
128	256	978M	1m45.121s
128	512	983M	1m41.387s
256	32	961M	1m48.786s
256	64	972M	1m43.246s
256	128	978M	1m45.195s
256	256	983M	1m43.539s
256	512	981M	1m39.424s
512	32	972M	1m45.553s
512	64	978M	1m41.079s
512	128	983M	1m45.405s
512	256	981M	1m42.975s
512	512	994M	1m39.523s

CPU  
Time: **4m1.299s**  
출력파일 크기 963M



- Clustering 용 데이터가 불규칙함.
- 데이터를 생성할때 필터링을 해야 하는 것을 확인하였음.

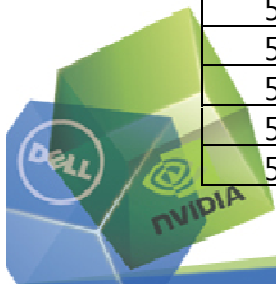


# Gapped Alignment by Smith Waterman

## CUDA-GPU

nBlock	nThread	출력파일	작동시간(sec)
32	32	644M	1m39.977s
32	64	644M	0m57.054s
32	128	644M	1m28.842s
32	256	644M	1m58.892s
32	512	644M	2m11.364s
64	32	644M	0m56.512s
64	64	644M	1m36.144s
64	128	644M	2m2.594s
64	256	643M	2m6.949s
64	512	247M	0m19.149s
128	32	644M	1m39.214s
128	64	644M	2m3.586s
128	128	643M	2m6.717s
128	256	247M	0m19.842s
128	512	20M	0m14.204s
256	32	475M	1m4.484s
256	64	490M	1m28.930s
256	128	20M	0m14.144s
256	256	145M	0m15.945s
256	512	0M	0m0.625s
512	32	288M	0m46.249s
512	64	20M	0m14.133s
512	128	145M	0m16.875s
512	256	0M	0m0.595s
512	512	0M	0m0.248s

CPU  
 Time: **10m19.835s**  
 출력파일 크기 644M





# Korea Graphics World 2012

## 신종독감

## 백신 바이오신약



## 글로벌 파마



### 차세대 시퀀서

2

**In Silico Epitope Screening**

Protein Model → VETLPEPDR → Experimental Validation

**Hardware Specifications:**  
 Intel(R) Xeon(R) CPU X5690 @ 3.47GHz \* 2 = 12 core  
 Memory : DDR3 128GB  
 448 \* 4 (Tesla C2075)

**System Capabilities:**  
 Calm System (10-20 Times faster with GPU)  
 V: virtual  
 E: epitope, drug, RNA/DNA, any kind of ligands  
 C: complex modeling  
 sim: simulation

**Things that Syntekbio guarantees for 1 year**  
 - how to manipulate 3D homology modelling  
 - how to prepare pre-setting of simulation  
 - how to perturb 3D structure for complex modeling  
 - how to do simulation with AMBER-GPU  
 - how to find drugs and targets with local-browser  
 - fast system setting and delivery (less than one month)  
 - give full scale instruction to bioinformatics specialist  
 - full technical support regarding the system

**Vertical Bar Charts:** Epitope Screening, Drug Screening, Complex Modeling

## \*게놈어셈블러

1

**Genome Assembly Software Interface**

Sequence alignment and assembly graphs are displayed on the screen.

**Hardware:** NVIDIA TESLA

## \*바이오맵시스템

## \*시스템바이오

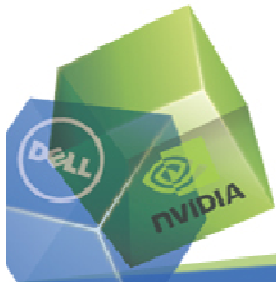
### \*Virus 게놈맵 Genome Map

**BioMap System**

Genome Map visualization with various data integration tools.

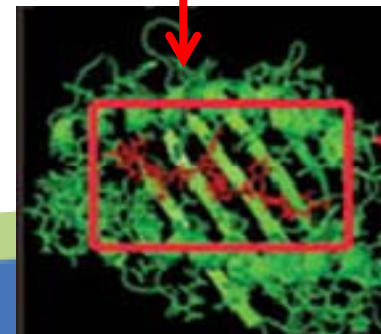
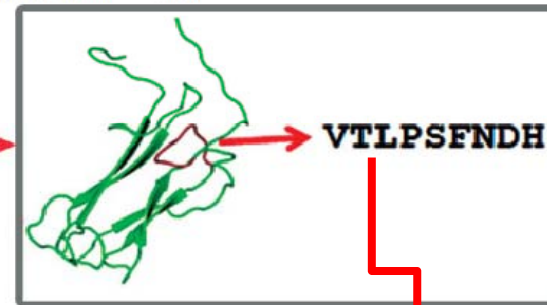
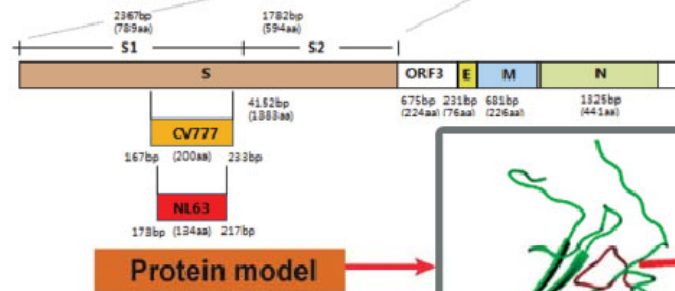
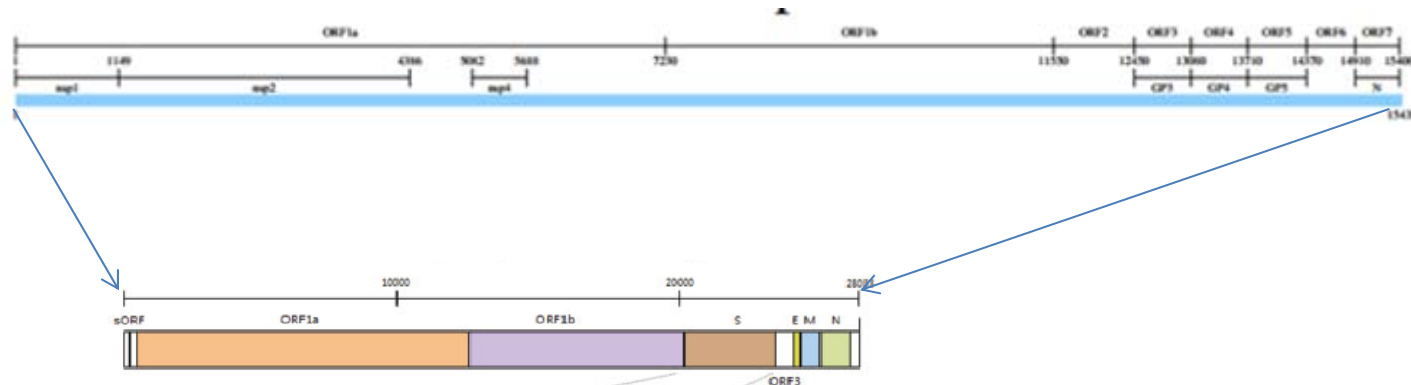
**Tools:** User's Query, Public DB Search, Web Programming, Info Integration

**Hardware:** BIORI BioMap CMap



# “전체 게놈에서 표적 항원까지”

초고속  
분석



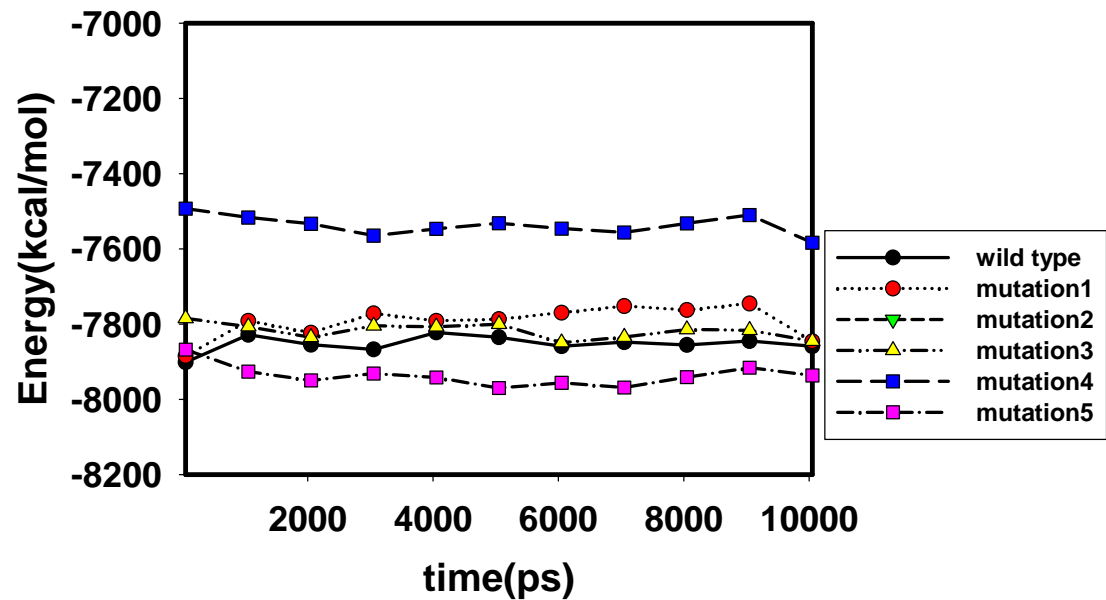


# AMBER-GPU 튜닝 테스트

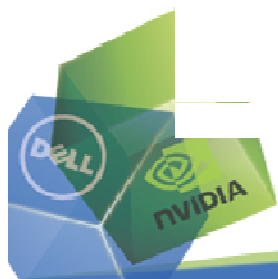
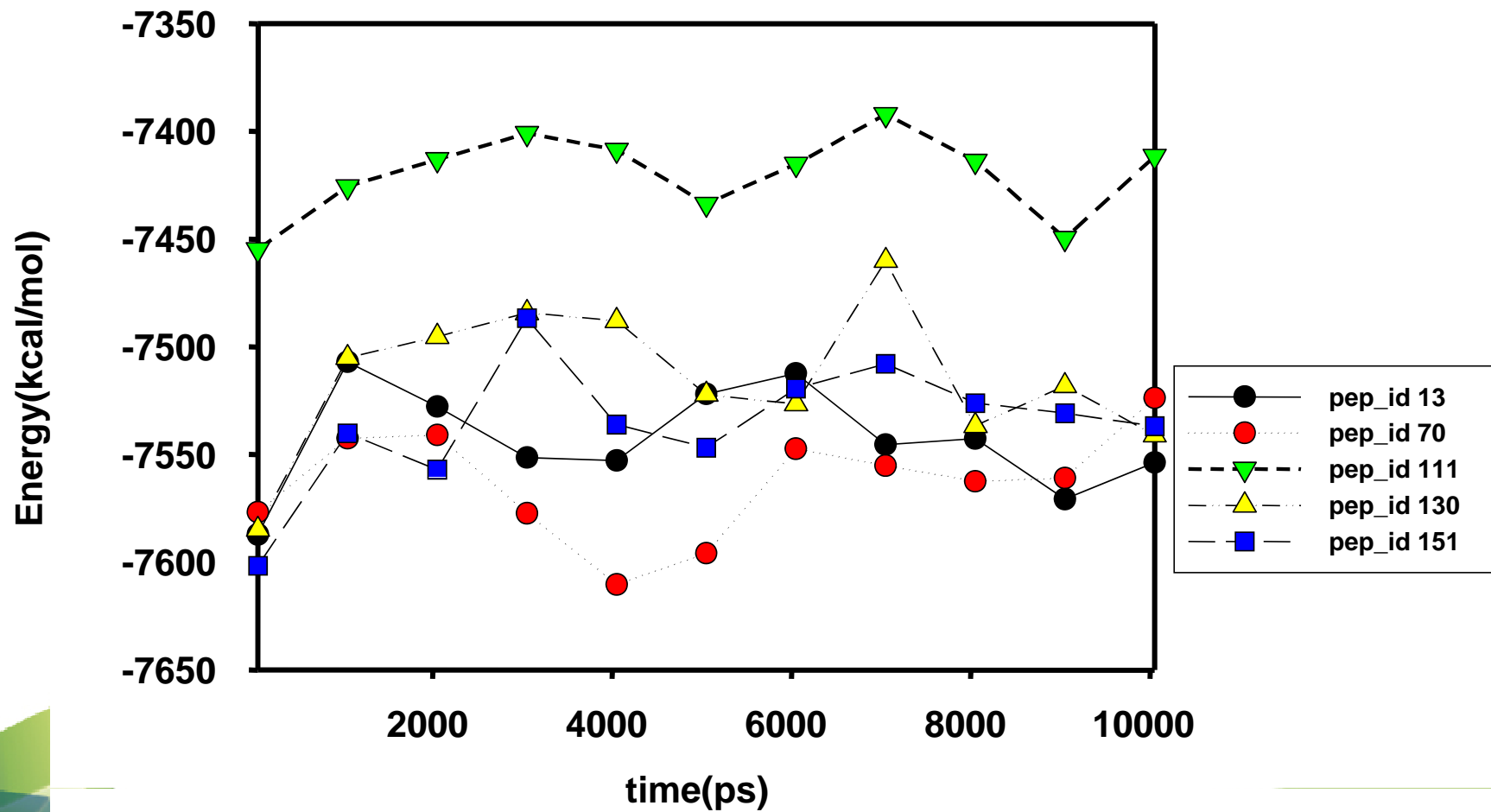
Test1. randomly mutated  
Peptide with swine MHC  
(PD:3LKR)

	1						10		
wt_pep	L	P	F	E	R	A	T	V	M
mutation1	L	A	F	E	R	W	T	V	M
mutation2	L	P	F	W	F	A	T	V	M
mutation3	L	A	A	A	R	A	T	V	M
mutation4	L	P	A	A	A	A	T	V	M
mutation5	L	P	F	E	R	A	A	A	A

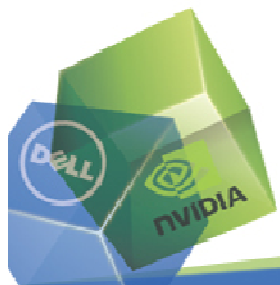
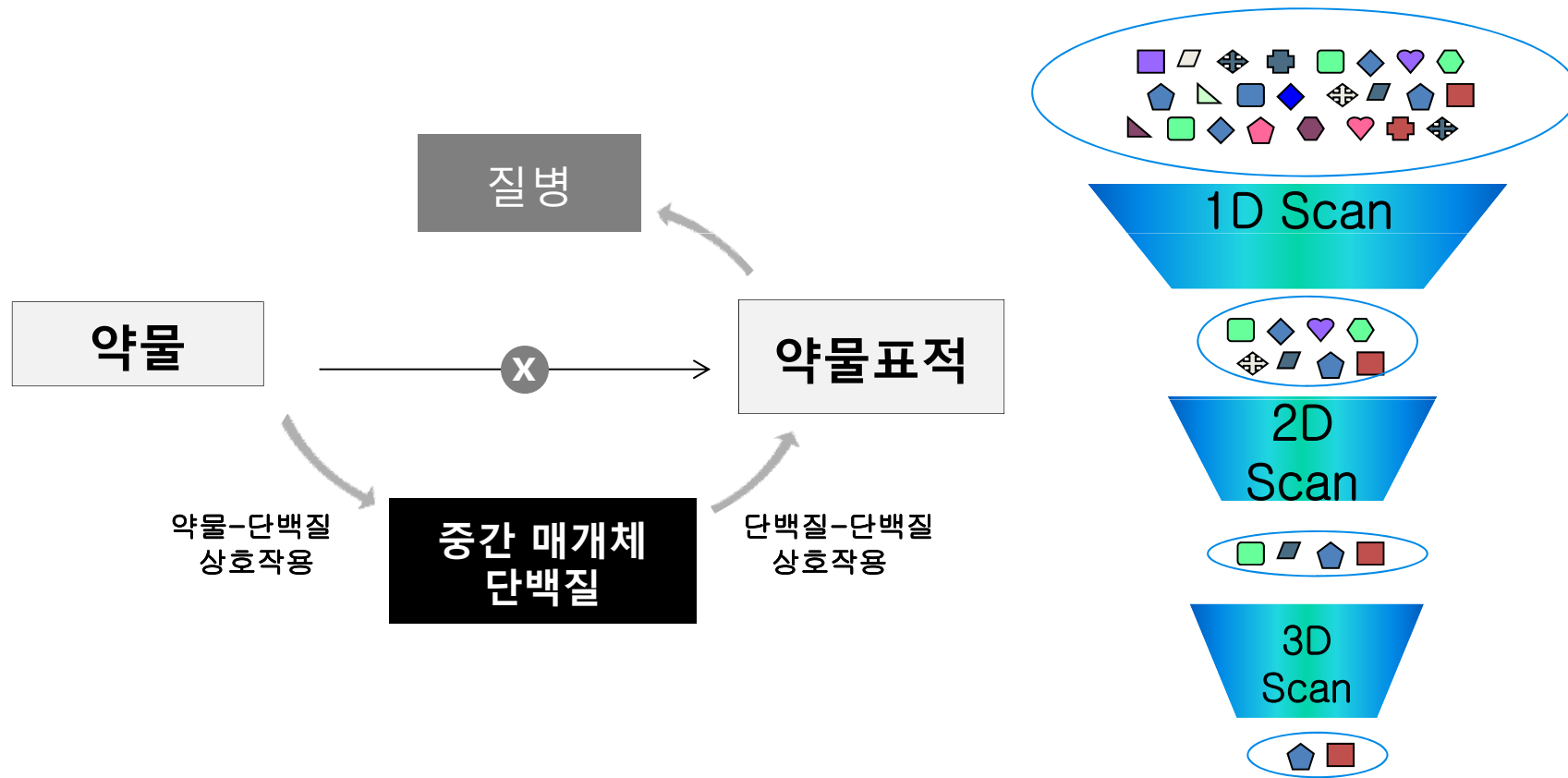
PEDV epitope simulation  
with swine MHC  
(PD:3LKR)



## Amber 기반 최적 epitope 예측



# 약물/상호작용 버추얼 스크리닝





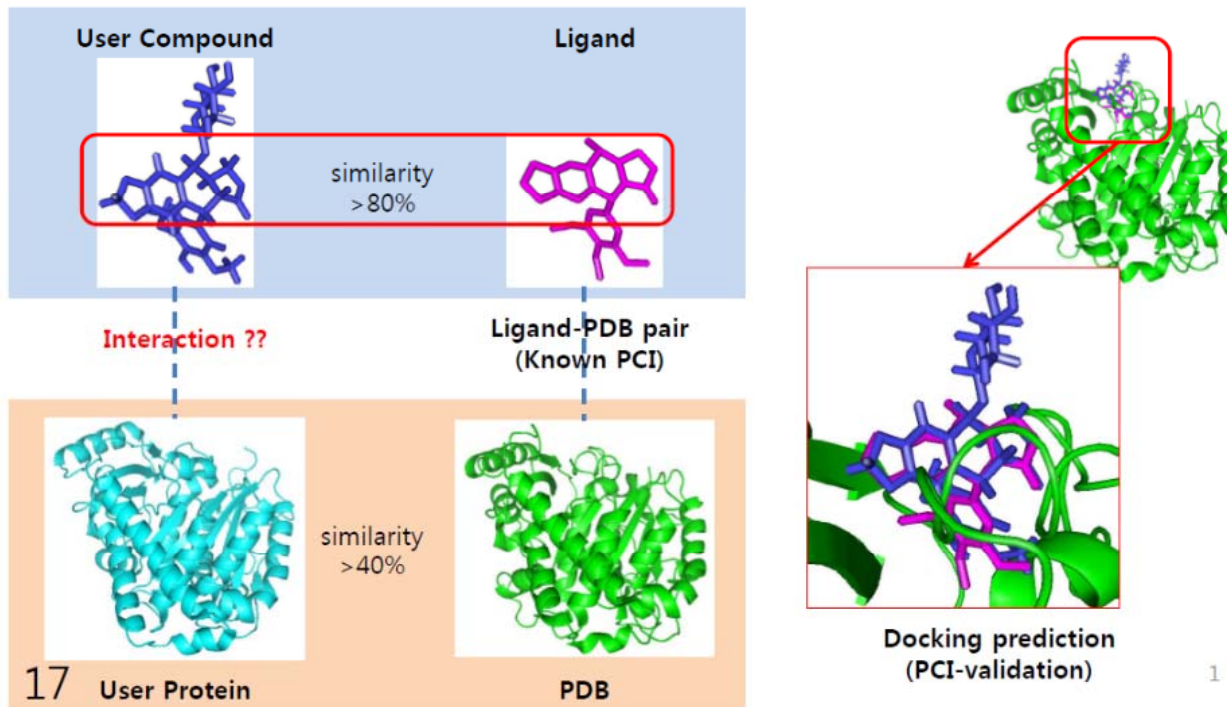
# 약물 표적 버추얼 스크리닝

## PCI validation \*Process Overview

## PPI / PCI Analysis

※ Based on the information about known Protein-Compound Interactions (PDB-Ligand pairs), PCI-validation predicts the customer's proteins' or compounds' possibility of binding. It looks up proteins with a sequence similarity of 40% or more, and compounds with a Tanimoto Coefficient of 0.8 or more.

Gene ID	No. of A.A	Gene Symbol	Chem ID	No. of Atoms	*P-G sim.	**T.C.	Score	Ref. PCI	***No.	Ligand-PDB pair
Gid_4810622	900	Cthe_1472	CID_11282830	65	100	1	94	PDB Ligand	1	ISX - 2bvd



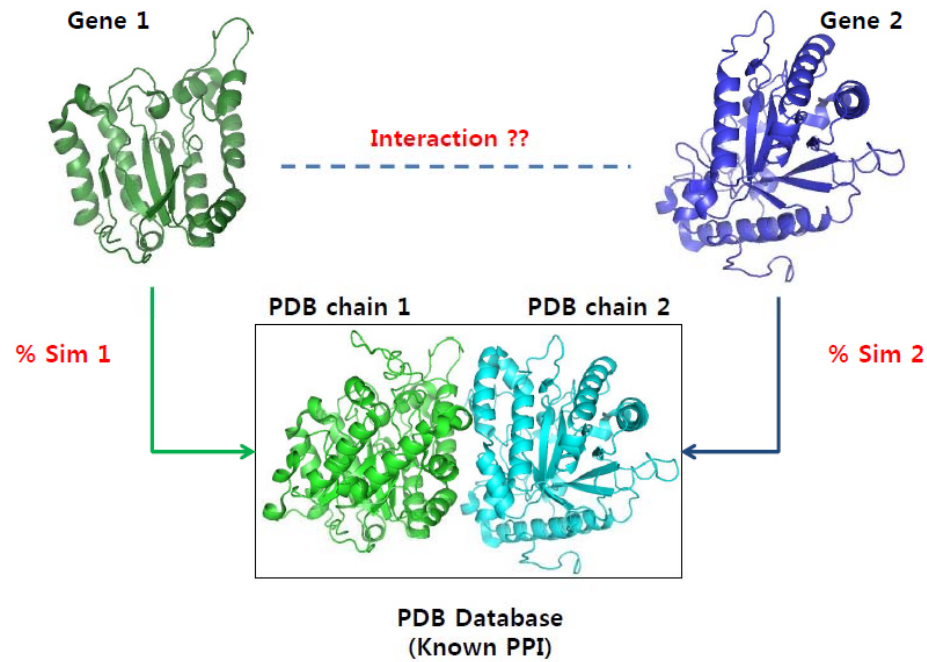
# 단백질 상호작용 버추얼 스크리닝

PPI validation **\*Process Overview**

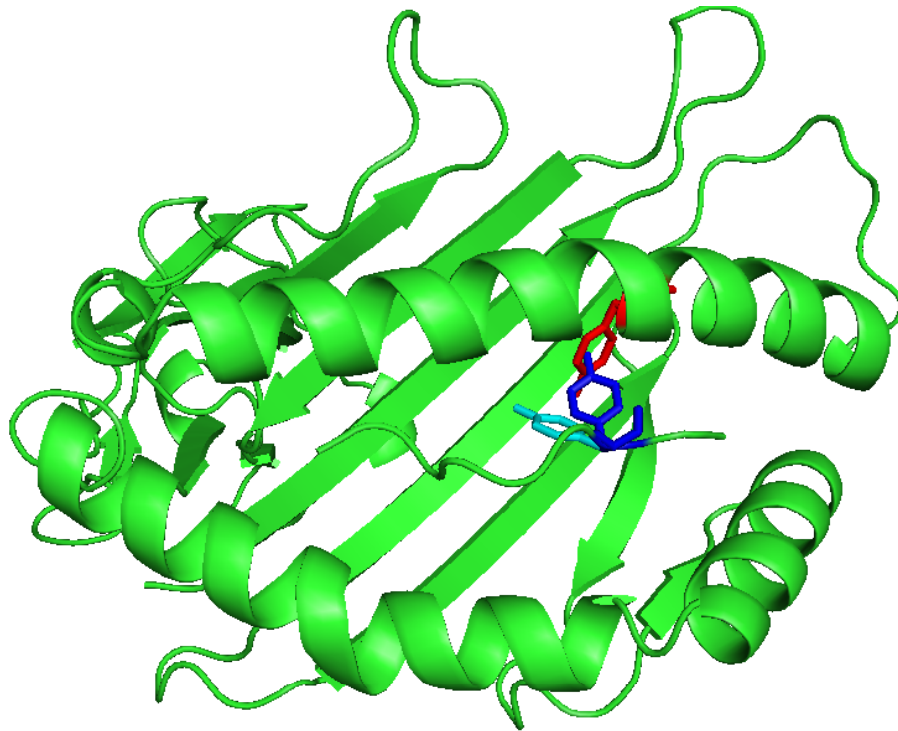
PPI / PCI Analysis

※ PPI-Validation is a 'PDB chains similarity'-based PPI prediction system. It searches for PDBs with a similarity of 40% or above among amino acid sequences and suggests some candidate proteins of your interest.

Gene ID 1	Gene Symbol 1	Gene ID 2	Gene Symbol 2	PDB Chain 1	PDB Chain 2	% Sim 1	% Sim 2	% Sim Avg.
Gid_7309912	Ccel_1099	Gid_7309582	Ccel_0728	2vn6^B_64	2vn6^A_151	96.830	100.000	98.415

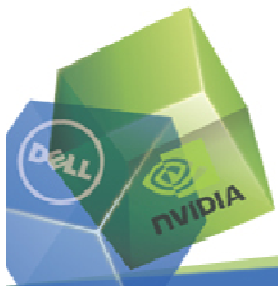


# 기술한계 극복1

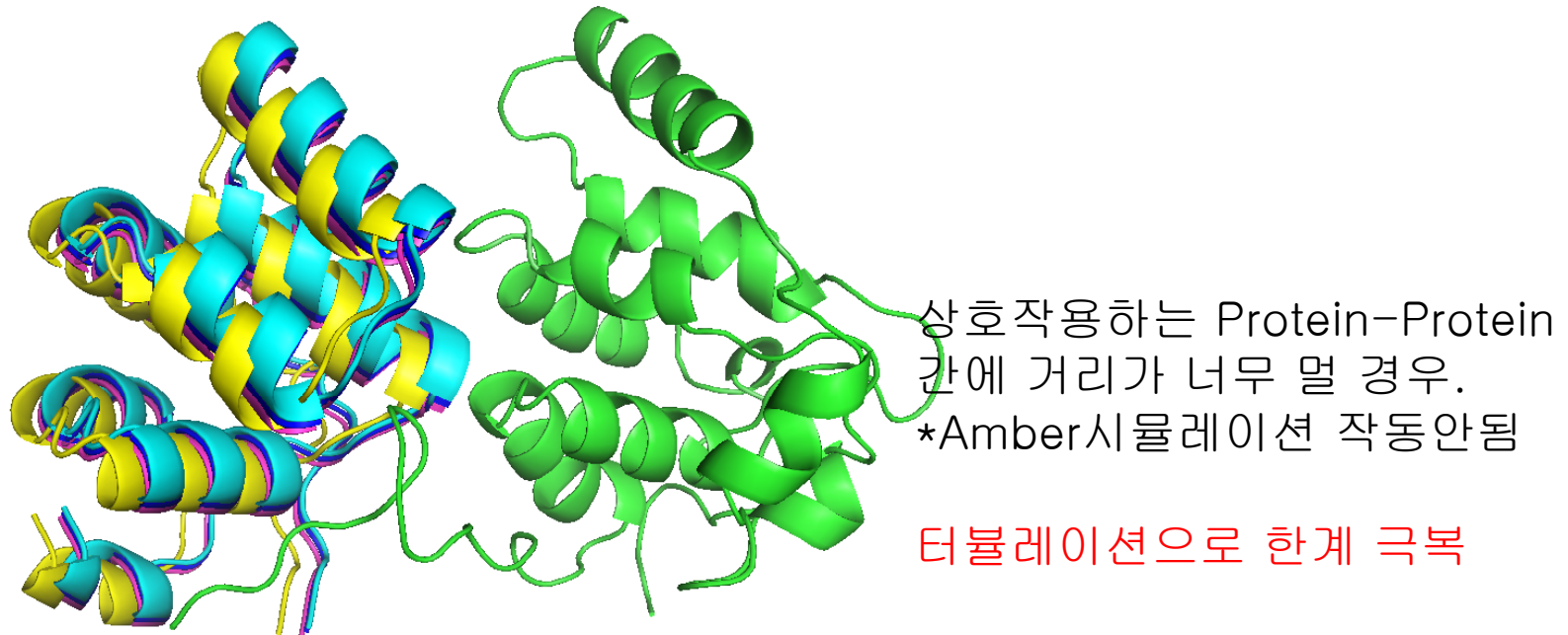


상호작용하는 Protein-ligand  
간에 crash가 있는 경우  
\*Amber시뮬레이션 작동안됨


3D조작(Editing)으로 한계 극복

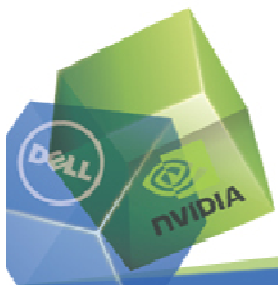


# 기술한계 극복2

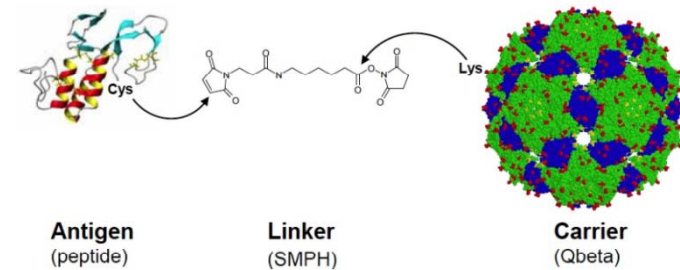
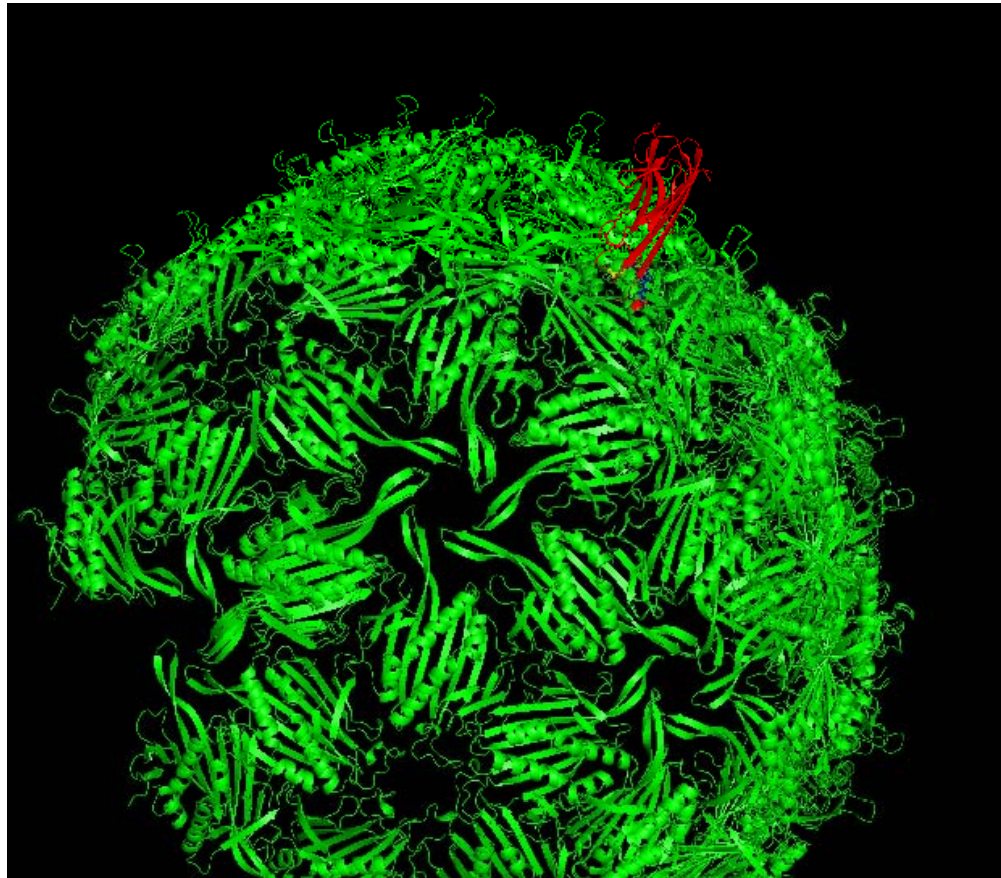


터볼레이션으로 한계 극복

- |   |                    |   |                                      |
|---|--------------------|---|--------------------------------------|
|  | Tube death domain  |  | Perturbed pelle death domain( upper) |
|  | Pelle death domain |  | Perturbed pelle death domain(lower)  |
|   |                    |  | Perturbed pelle death domain(right)  |



# 기술한계 극복3



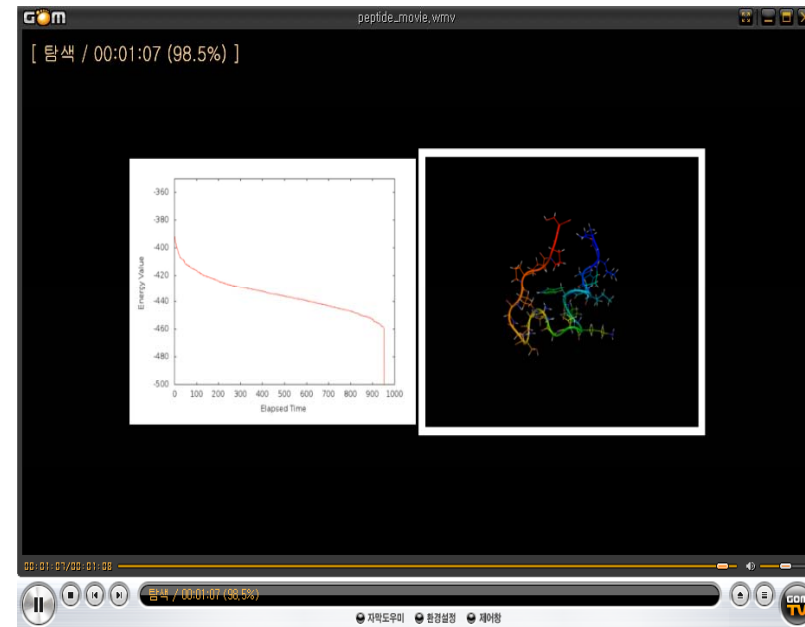
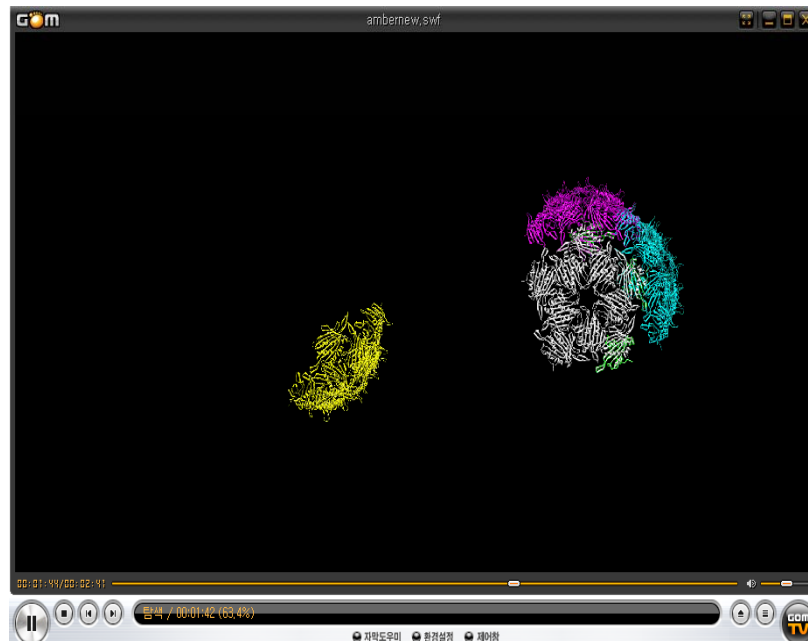
사이즈가 클 경우 어셈블리가 잘 안됨.

터볼레이션/시뮬레이션 반복으로 극복



## Virus-Like Particle

# Complex Modeling



동영상  
4개 시연





## VECSim (Amber-GPU)

**Gene Compound SNP CMAP BIORI**

**3D Modeling System**

**In Silico Epitope Screening**

**Epitope Screening Drug Screening Complex Modeling**

**Experimental Validation**

**VECSim System (10~20 Times faster with GPU)**  
 V: virtual  
 E: epitope, drug, RNA/DNA, any kind of ligands  
 C: complex modeling  
 sim: simulation

**Things that SyntekBio guarantees for 1 year**

- how to manipulate 3D homology modeling
- how to prepare pre-setting of simulation
- how to perturb 3D structure for complex modeling
- how to do simulation with AMBER-GPU
- how to find drugs and targets with local-browser
- fast system setting and delivery (less than one month)
- giving full scale instruction to bioinformatics specialist
- full technical support regarding the system

**Intel(R) Xeon(R) CPU X5690 @ 3.47GHz \* 2 = 12 core  
 Memory : DDR3 128GB  
 448 \* 4 (Tesla C2075)**

**SyntekBio Inc.** 서울 성북구 아암로 39-1번지 한국과학기술연구원 벤처타운 992호  
 Tel:02-960-0982 service@syntekbio.com www.Syntekbio.com

## BioMap System

**CPU: Xeon, 40(80)Cores, Memory: 1TB SAS RAID 10TB**

**BMAP (BioMap)**

**Web Programming**

**Public DB Search**

**Info Integration**

**CMAP (CellMap)**

**바이오구글링 & 자동워크플로우 기반 바이오맵 시스템**

- 검색대상 유전자의 homologues (%homology>40 or 사용자 정의)에 의해서 맵핑되는 모든 유전자들을 공용DB 약100 여 개에서 완벽하게 검색
- 검색대상 SNP의 LD block ( $r^2 = 0.7$  or 사용자 정의)에 의해서 맵핑되는 모든 SNP 들을 공용DB 약100 여개에서 완벽하게 검색
- 검색대상 화합물과 유사한 화합물 (tanimoto > 0.8 or 사용자 정의)에 의해서 맵핑되는 모든 화합물을 공용DB 약100 여개에서 완벽하게 검색
- 위의 Gene, Compound, SNP에 의해서 검색된 정보중 PPI (protein-protein interaction), PCI (protein-chemical interaction) 에 의해서 연결되는 모든 정보들에 대한 네트워크를 중 사용자가 원하는 유전자들 기준으로 상세-맵 작성가능
- 검색된 PPI및 PCI 에 대한 validation 결과 생성
- 위의 검증된 결과에 대한 공용DB상에 존재하는 모든 Evidence 생성 방법 제공

**SyntekBio Inc.** 서울 성북구 아암로 39-1번지 한국과학기술연구원 벤처타운 992호  
 Tel:02-960-0982 service@syntekbio.com www.Syntekbio.com





# 감사합니다.

Syntekabio Inc. 대표이사 정종선  
Channel & Alliance  
[jung@syntekabio.com](mailto:jung@syntekabio.com)

