Methods for studying macromolecular systems
- Multivariate analysis of angular data

Carmay Lim
Institute of Biomedical Sciences
Academia Sinica

Workshop on analysis and its applications in biology and physiology
19 December 2014

Thanks to present and past group members.

Thanks also to my long-time collaborators

Dr. Hanna Yuan

Institute of Molecular Biology

Dr. Tse Wen Chang

Genomics Research Center

Supported by Academia Sinica & National Science Council, Taiwan

Plan of talk

1. Overview of our research interests
2. Multivariate analysis of angular data

Research Interests

1. Unravel the principles governing biological processes, and apply them to guide drug design & identify new drug targets.

2. Develop new methods for studying macromolecular interactions.

Principles governing metal selectivity in proteins

Revealed how metal binding in proteins depends on

1. the intrinsic properties of the metal in mono/polynuclear sites,
2. the 1st shell (CN, geometry, aa type, denticity),
3. the 2nd shell (& its interaction with the 1st shell),
4. the protein matrix and its aqueous environment, and
5. competing anions; e.g., Cl^-, from the cellular fluids.

Todor Dudev

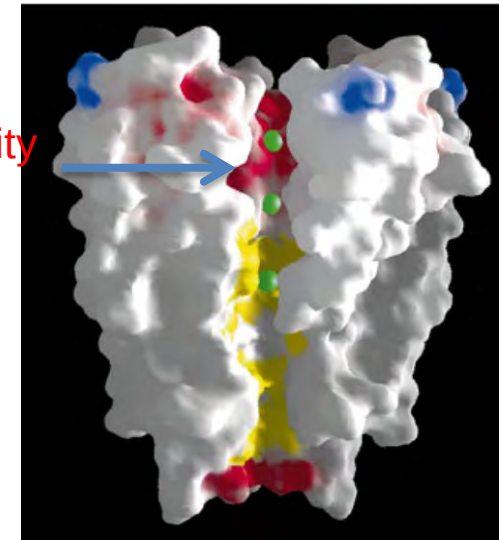
Competition Among Metal Ions for Protein Binding Sites. *Chem. Reviews* **2014**

Principles governing ion selectivity in ion channels

- Ion channels control signal transduction, cardiac, skeletal & smooth muscle contraction, hormone secretion, and taste & pain sensation.

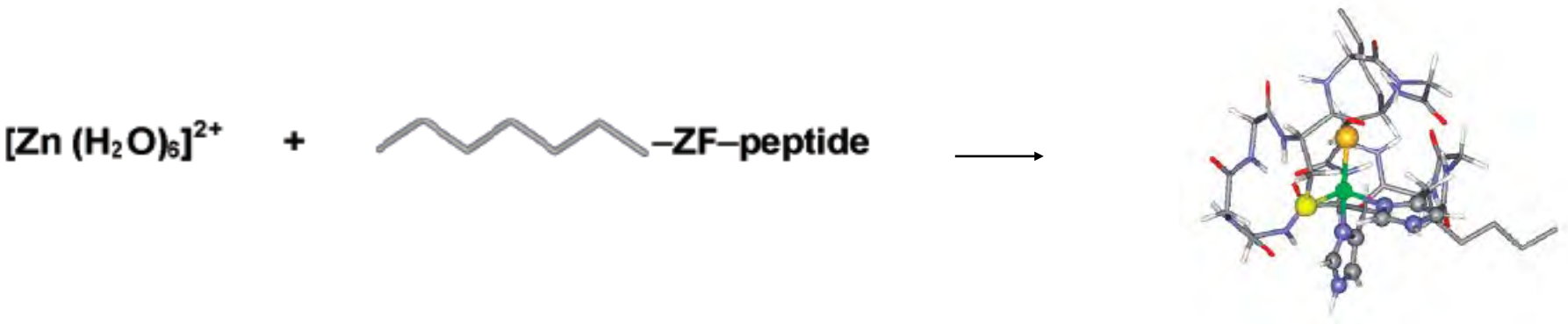
- Their malfunction cause various life-threatening disorders such as cardiac arrhythmias, hypertension, angina, heart attack, stroke, migraine, epilepsy, pain, cancer and autoimmune disorders.

- Sodium Channels: *Acc. Chem. Res.*, **2014**
- Potassium Channels: *J. Am. Chem. Soc.*, **2009**
- Magnesium Channels: *J. Am. Chem. Soc.*, **2013**
- Calcium Channels: *J. Phys. Chem. B*, **2012**
- Ion Channel Evolution: *J. Am. Chem. Soc.*, **2014**



Novel force fields for metalloprotein simulations

- More than 30% of cellular proteins bind metal ions, which play essential tasks such as protein structure stabilization, enzyme catalysis, signal transduction, muscle contraction, hormone secretion, taste and pain sensation, respiration, and photosynthesis.



Computer simulations of metalloproteins require an accurate forcefield

$$V_{Zn-i}(t) = \sum_{i=1} \frac{q_{Zn} q_i}{4\pi\varepsilon_0 r_{Zn-i}(t)} + 4\varepsilon_{Zn-i}\left[\left(\frac{\sigma_{Zn-i}}{r_{Zn-i}(t)}\right)^{12} - \left(\frac{\sigma_{Zn-i}}{r_{Zn-i}(t)}\right)^{6}\right]$$

A force field is described by

- the potential energy function and
- its associated parameters; e.g., the parameters modeling the interaction of the metal ion with water and biological ligands.

Conventional force fields do not account for q-transfer

$$V_{Zn-i}(t) = \sum_{i=1} \frac{q_{Zn} q_i}{4\pi\varepsilon_0 r_{Zn-i}(t)} + 4\varepsilon_{Zn-i}\left[\left(\frac{\sigma_{Zn-i}}{r_{Zn-i}(t)}\right)^{12} - \left(\frac{\sigma_{Zn-i}}{r_{Zn-i}(t)}\right)^{6}\right]$$

In conventional potential energy function, the Zn charge is constant, irrespective of the number and the type of protein ligand bound to Zn during a simulation!

Why the need for a new force field for metalloproteins

- The metal coordination # could decrease upon protein binding; e.g., in solution Zn is octahedrally bound to 6 water molecules.In most Zn proteins, Zn is tetrahedrally coordinated, but in some catalytic sites, it is found 5- or 6-coordinated.

- Current force fields cannot account for the change in the Zn coordination number, which occurs in the folding of Zn-finger proteins or in certain enzymatic reactions.

The new force field accounts for charge-transfer and local polarization effects

$$V_{Zn-j}(t) = V^{pol}(t) + \sum_j \frac{q_{Zn}(t)q_j(t)}{4\pi\varepsilon_0 r_{Zn-j}(t)} + 4\varepsilon_{Zn-j}\left[\left(\frac{\sigma_{Zn-j}}{r_{Zn-j}(t)}\right)^{12} - \left(\frac{\sigma_{Zn-j}}{r_{Zn-j}(t)}\right)^{6}\right]$$

The new force field differs from the conventional one in 3 ways:

1. It includes the polarization energy for the metal (Zn^{2+}) & its ligands.

2. It includes the charge transferred by the a.a. ligands to the metal. Thus, the charges on Zn and its ligand (L) atoms change depending on the Zn-L distance during the simulation.

3. The parameters are derived to reproduce the hydration free energies and structural properties of all cations of the same net charge.

Novel force fields for metalloprotein simulations

Satheesan Babu

Dmitri Sakharov

- Allows accurate simulations of ~40% proteins containing metal sites.

1. Sakharov & Lim (**2005**) Zn Protein Simulations Including Charge Transfer and Local Polarization Effects *J. Am. Chem. Soc*. 127: 4921.
2. Babu & Lim (**2006**) Empirical Force Fields for Biologically Active Divalent Metal Cations in Water. *J. Phys. Chem. A* 110: 691.
3. Sakharov & Lim (**2009**) Force Fields Including Charge Transfer and Local Polarization Effects: Application to Proteins Containing Multi/Heavy Metal Ions *J. Comp. Chem.* 2009

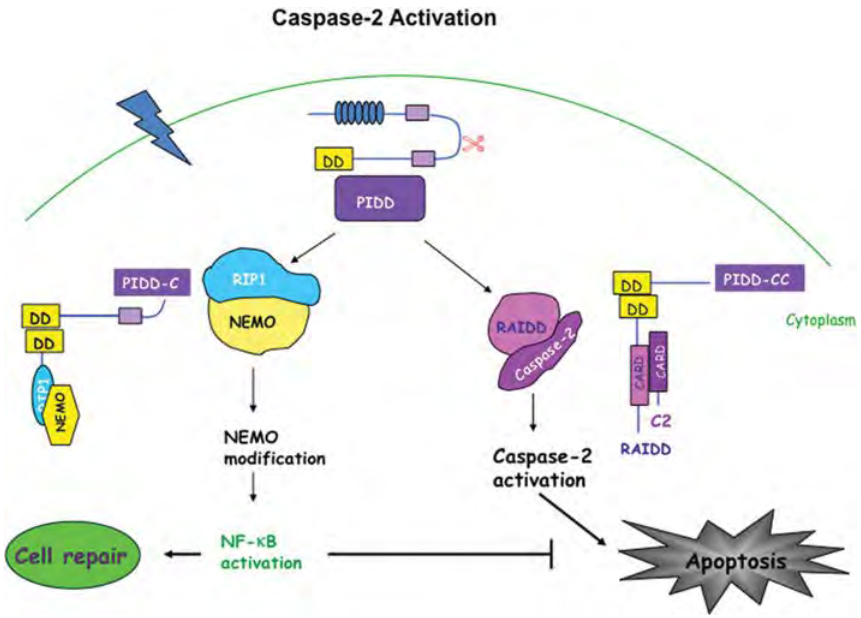
Structure-based prediction of ligand-binding sites

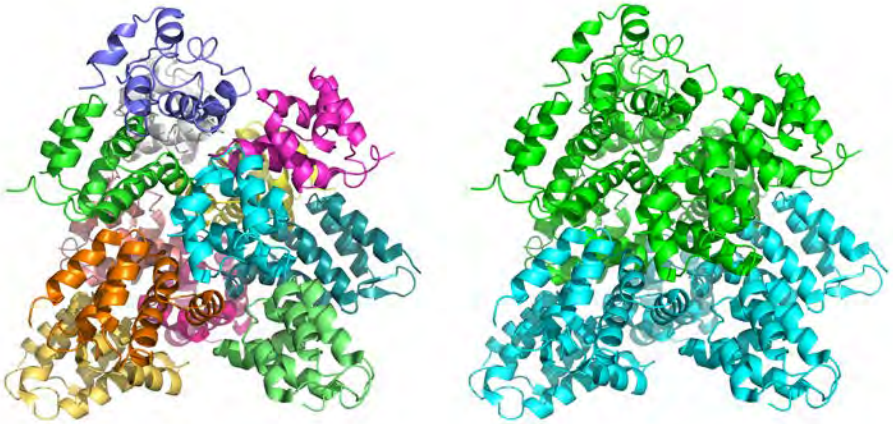
- Revealed a common physical basis underlying DNA/RNA/protein-binding sites & their distinguishing features.
- Used physical basis to predict functional residues. http://drbind.limlab.ibms.sinica.edu.tw/
- Help in functional annotation of the new structures & in reducing conformational search in docking simulations.

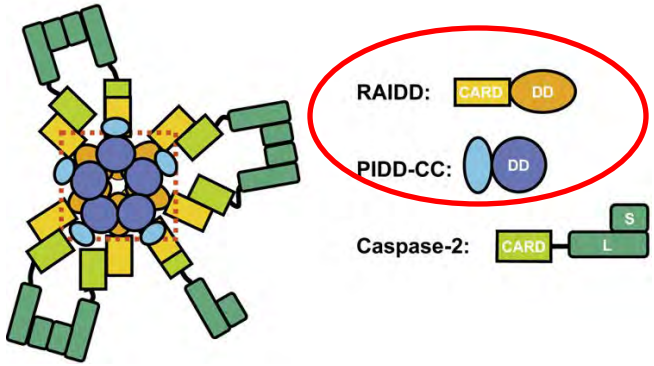
Backy Chen

Nucl. Acids Res. **2014**, **2012**, **2008**
Proteins: Struct. Func. & Bioinformatics **2007**

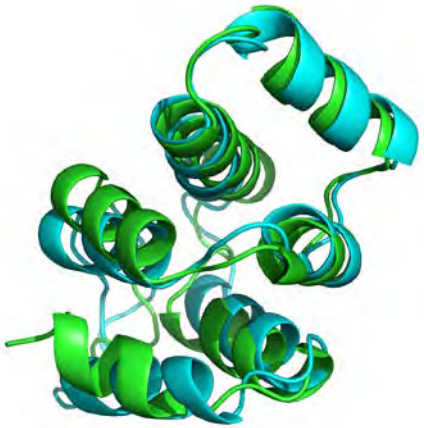
Applications of DR_bind2 to huge oligomeric complexes involved in cell apoptosis and inflammation mechanisms







Cell. 2007 Feb 9;128(3):533-46



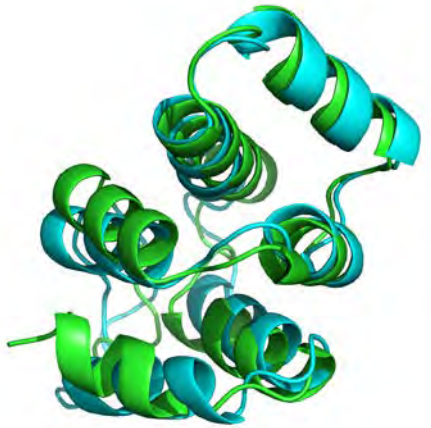
The Death-Domain Fold

The Death-Domain Fold

- The death fold is a tertiary structure motif commonly found in proteins involved in apoptosis or inflammation-related processes.

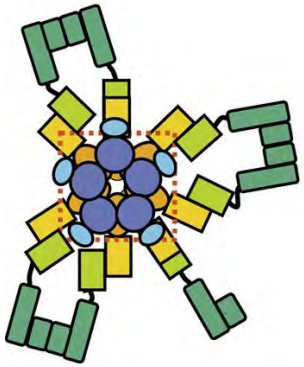
- This motif is commonly found in domains that participate in protein-protein interactions leading to the formation of large functional complexes, such as PIDDosome, Myddosome, and inflammasomes.

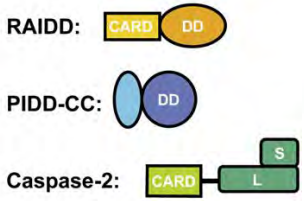
- Examples of death fold domains include the
  death domain (DD)
  death effector domain (DED)
  Caspase Recruitment Domain (CARD)
  pyrin domain (PYD)

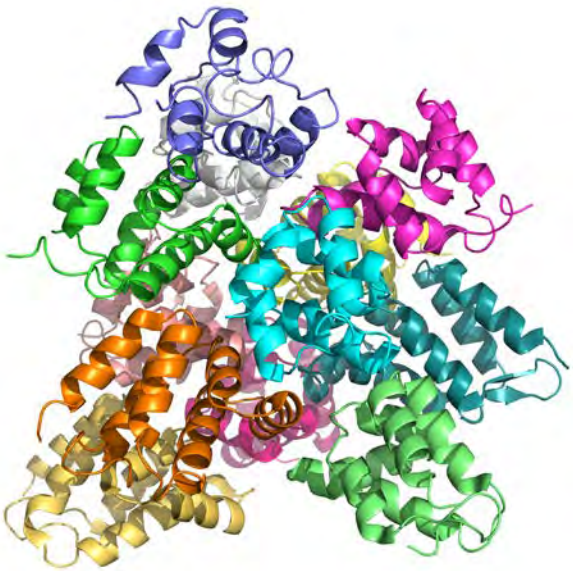


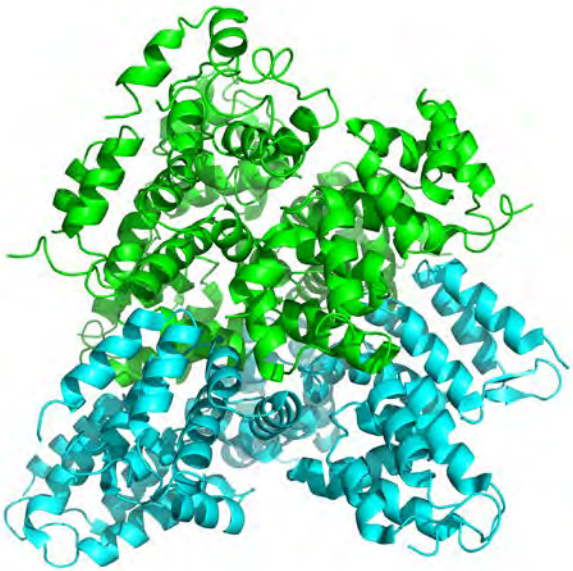
http://en.wikipedia.org/wiki/Death_fold

PIDDosome









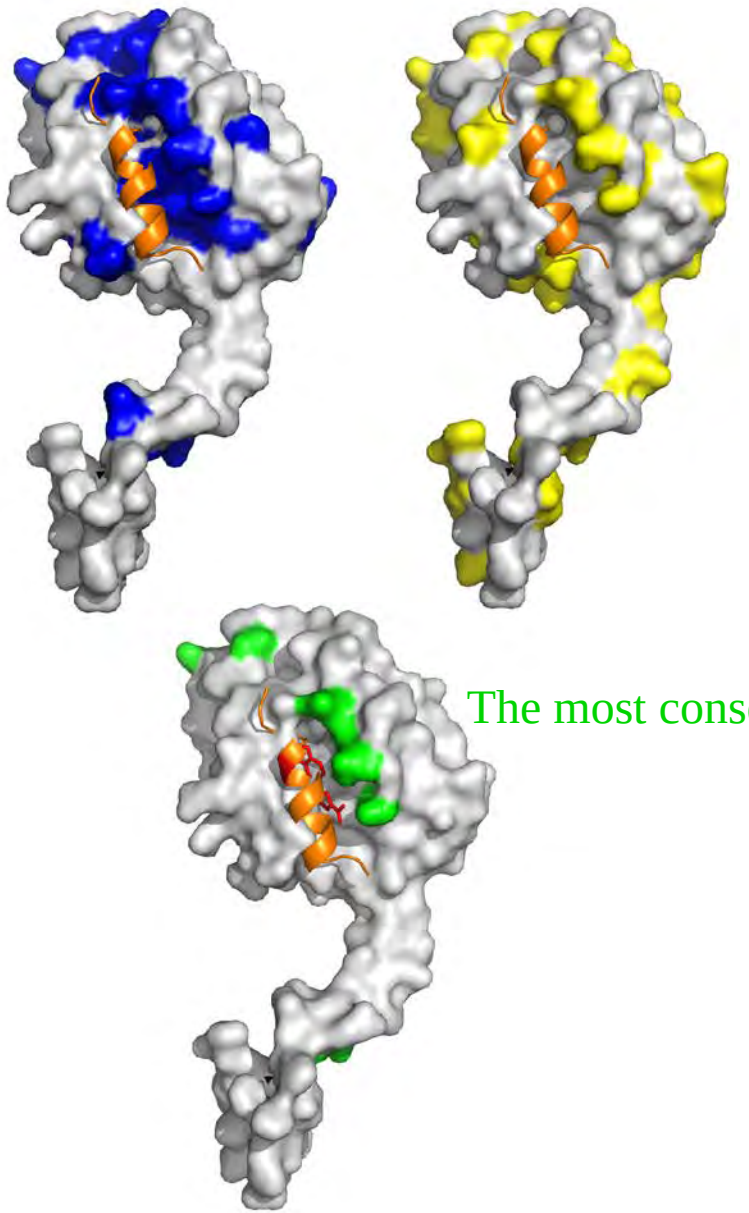
PDB entry 2OF5 **(**Cell **2007** 128:533-46**)**

Death domain-containing protein
CRADD: A, B, C, D, E, F, G

p53-induced protein with a death domain
PIDD: H, I, J, K, L

How DR_bind2 identifies functional residues

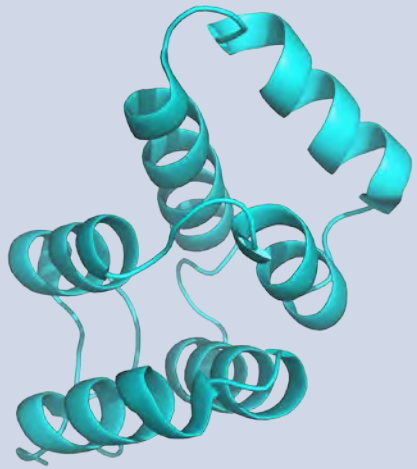
The most conserved regions



The most unstable regions

The most conserved and unstable regions

Functional residues predicted from PIDD structures



DR_bind2

D864

PDB entry 2OF5
Death domain-containing protein
CRADD: A, B, C, D, E, F, G
p53-induced protein with a death domain
PIDD: H, I, J, K, L

Experimental verification of "functional" residues

Interaction	a	b
Type I, R:P	N121D**, Q125A**	Y814A**
Type I, R:R	N121D**	Q142E*, L136E*
Type I, P:P	E830K, F837D**	Y814A**
Type II, R:P	Q169E**, Q169A*, R170A*	R862A**, Q863A*
Type II, R:R	Q169E**, Q169A*, R170A*	E188K, V189W, V189D
Type III, R:P	Y146A**, R147A**, R147E**	L801A, L828E**
Type III, R:R	Y146A**, R147A**, R147E**	V156D**
Type III, P:P	R815A*, R815E**, H822A*	L801A, L828E**
Shared residues†	H154A*, K149E**	R825A*, R825E*, D826K*

†These residues bury large surface area upon complex formation, but are not in Van der Waals contacts with any neighboring molecules.

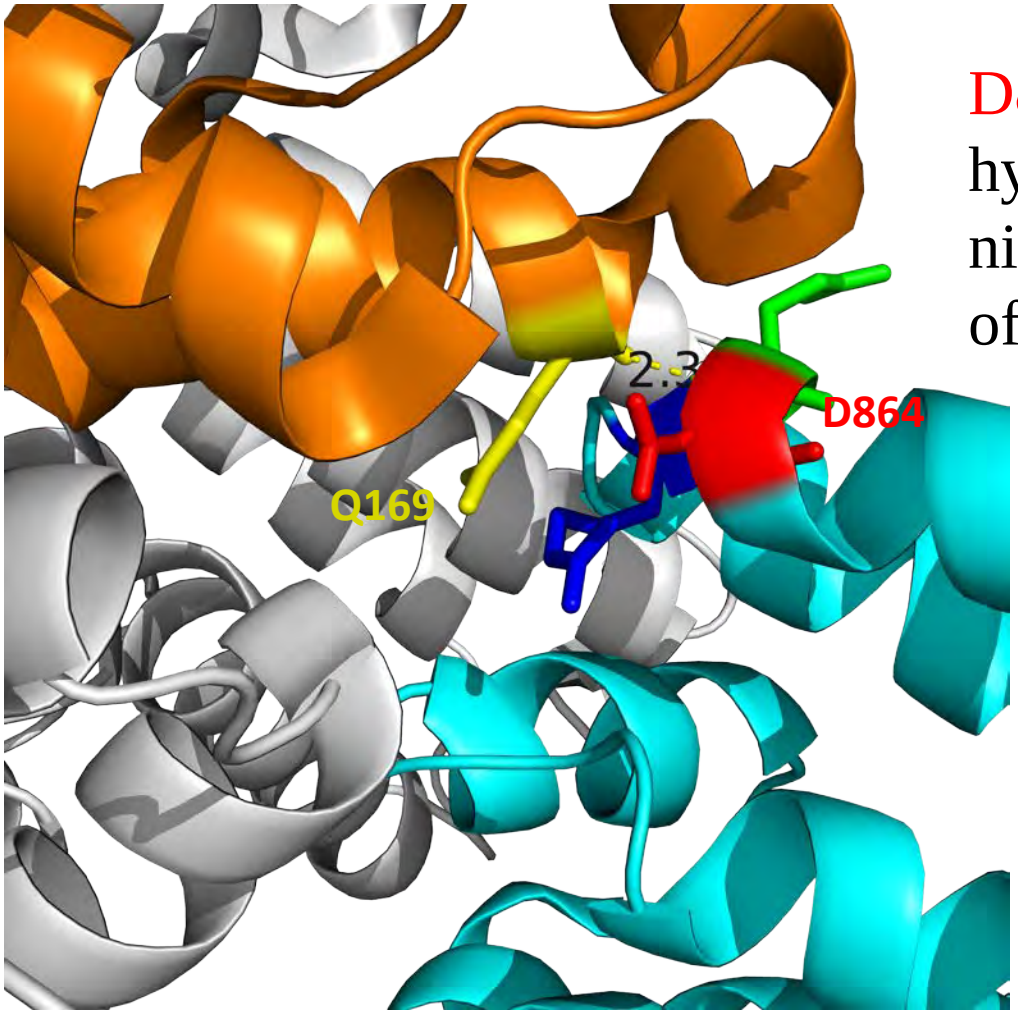
PIDD

D864

** means mutations completely disrupted PIDD:CRADD complex formation

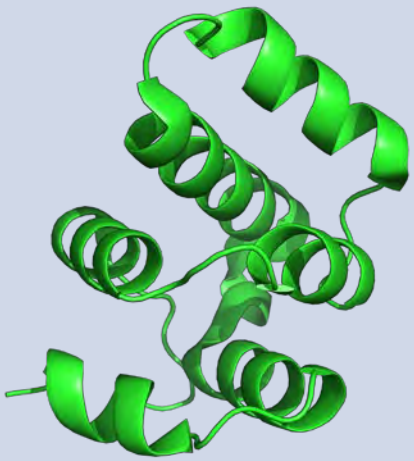
* means mutations partially disrupted complex formation

D864 participates directly in PIDD–CRADD interactions



D864 forms a backbone–backbone hydrogen bond between its amide nitrogen and the carbonyl oxygen of Q169.

Functional residues predicted from CRADD structures



DR_bind2

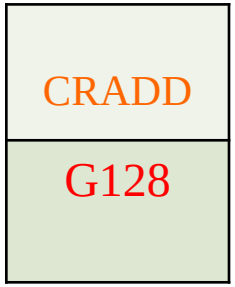
G128

PDB entry 2OF5
Death domain-containing protein
CRADD: A, B, C, D, E, F, G
p53-induced protein with a death domain
PIDD: H, I, J, K, L

G128 was not among the residues mutated!

Interaction	a	b
Type I, R:P	N121D**, Q125A**	Y814A**
Type I, R:R	N121D**	Q142E*, L136E*
Type I, P:P	E830K, F837D**	Y814A**
Type II, R:P	Q169E**, Q169A*, R170A*	R862A**, Q863A*
Type II, R:R	Q169E**, Q169A*, R170A*	E188K, V189W, V189D
Type III, R:P	Y146A**, R147A**, R147E**	L801A, L828E**
Type III, R:R	Y146A**, R147A**, R147E**	V156D**
Type III, P:P	R815A*, R815E**, H822A*	L801A, L828E**
Shared residues†	H154A*, K149E**	R825A*, R825E*, D826K*

†These residues bury large surface area upon complex formation, but are not in Van der Waals contacts with any neighboring molecules.



** means mutations completely disrupted PIDD:CRADD complex formation.

* Means mutations partially disrupted complex formation

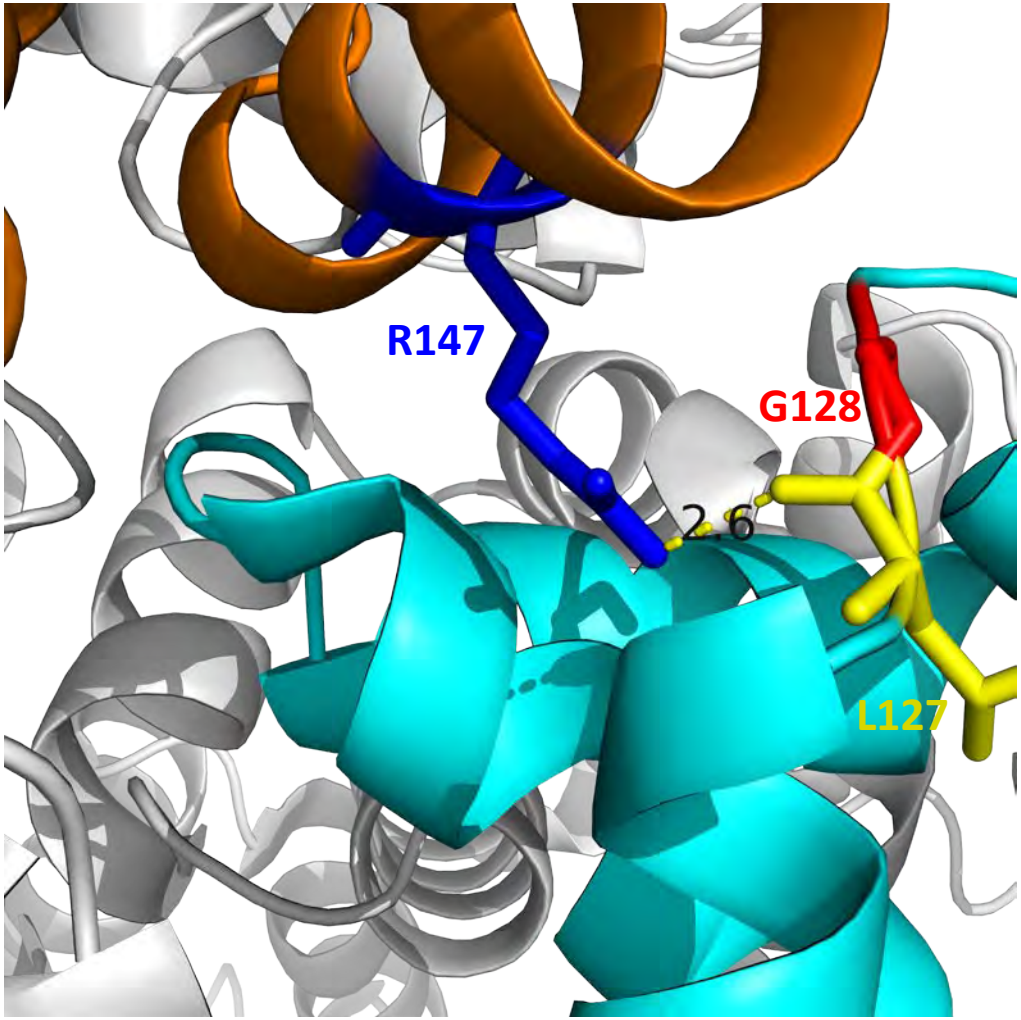
G128R in CRADD is related to Mental retardation

Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases

Received September 16, 2011; Accepted November 17, 2011; Published January 17, 2012

- Gly128Arg CRADD formed large aggregates when co-overexpressed with wild-type PIDD or the PIDD death domain.
- These data suggest that the Gly128Arg mutation alters one of the interaction surfaces of the CRADD death domain to decrease affinity for the PIDD death domain.

G128 participates indirectly in CRADD–CRADD interactions

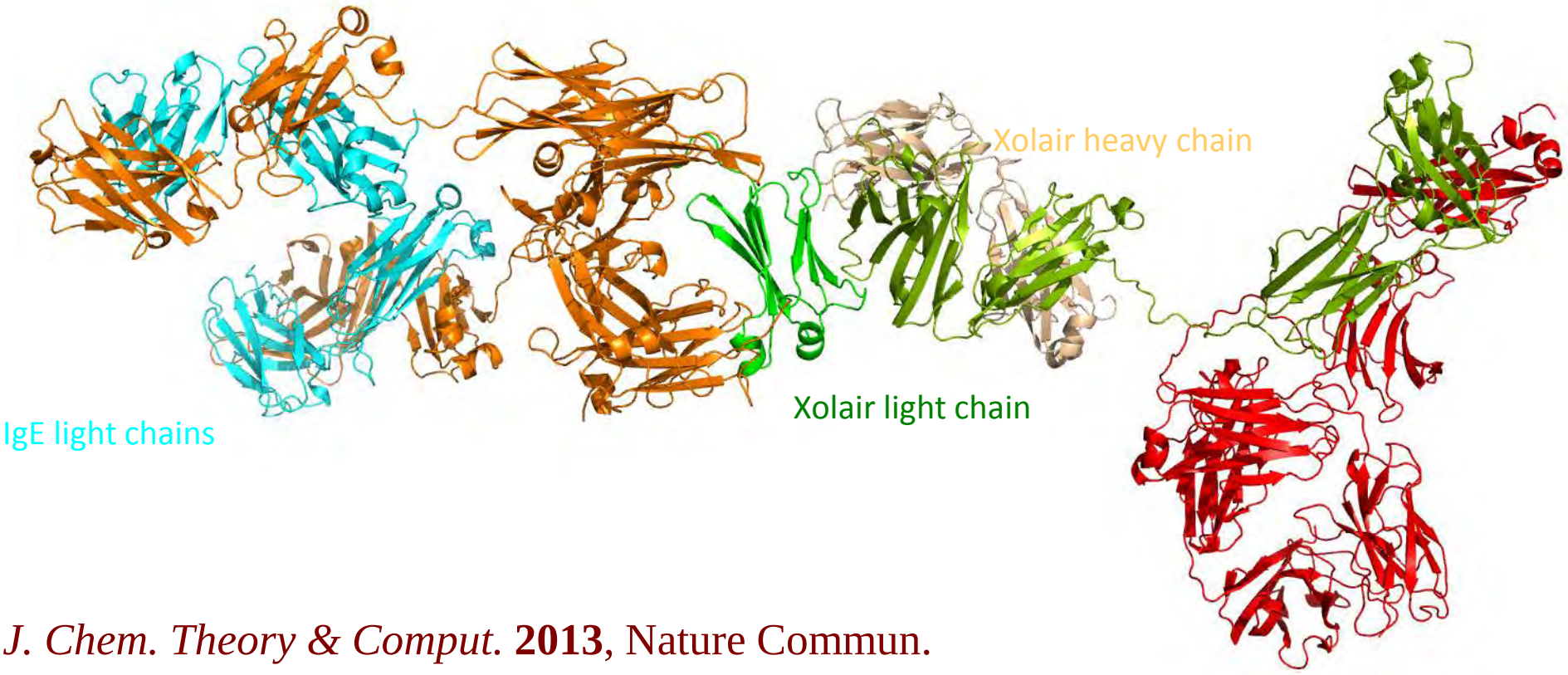


G128 participates indirectly in CRADD–CRADD interactions via a backbone–side chain hydrogen bond between its neighbor, L127 and R147. Its mutation to a positively charged Arg would likely repel the nearby positively charged R147 in another chain, thus disrupting the CRADD–CRADD interface, which in turn would decrease CRADD's affinity for PIDD.

Protein-protein docking using CHARMM and SVM

Jon Wright

Karen Sargsyan



J. Chem. Theory & Comput. **2013**, Nature Commun.

Strategies for finding sequence & structural motifs

Steven Wu

Minko Dudev

Karen Sargsyan

- Developed a method for finding 3D motifs across protein families. (*BMC Bioinformatics*, **2007;** *Nucl. Acids Res.* **2010**) and in large RNA structures (*Nucl. Acids Res.* **2010**).

- Developed a strategy for discovering 1D motifs in proteins (*J. Phys. Chem.* B, **2012**).

- Aid functional annotation of the new structures, detecting protein binding sites & new drug targets (*Scientific Reports*, **2014**).

Tools for conformational analyses using angles

- Angular data pose difficulties in multivariate analyses; e.g., the mean of 10^{o} and 350^{o} is 180^{o} instead of 0^{o}; thus, statistical analysis of linear data and Euclidean geometry cannot be applied directly to circular data.

- Developed the **Clustang** package for analysis of dihedral angles based on geodesics and spherical distance. http://pca.limlab.ibms.sinica.edu.tw/

Karen Sargsyan

- **Clustang** provides a useful way of visualizing, analyzing, and predicting conformations of complex macromolecules with many degrees of freedom.

GeoPCA - *Nucleic Acids Res.* **2012**

Polarization Energy Term

$$V^{pol}(t) = -\frac{1}{2}\sum_i \mu_i(t) E_i^0(t)$$

where

$$\mu_i(t) = \alpha_i E_i(t)$$

$$E_i(t) = E_i^0(t) + \sum_{j\neq i} T_{ij}(t)\mu_j(t) = \sum_{j\neq i} \frac{q_j \vec{r}_{ij}(t)}{r_{ij}^3(t)} + \sum_{j\neq i} \frac{\vec{\mu}_j(t)}{r_{ij}^3(t)}\left(\frac{3\vec{r}_{ij}(t)\vec{r}_{ij}(t)}{r_{ij}^2(t)} - 1\right)$$

$$\mu_i(t) = \alpha_i E_i^0(t) + \alpha_i \sum_{j\neq i} T_{ij}(t)\mu_j(t)$$

Electric fields caused by charges Electric field caused by induced dipoles.

is obtained by solving a set of coupled equations, which are solved iteratively.

How the CHARMM Zn^{2+} parameters were derived

The CHARMM parameters (ε_{Zn}= –0.25 kcal/mol, σ_{Zn} = 1.95 Å) were obtained initially by fitting to an ab initio-derived Zn^{2+}–H_2O potential energy surface, and then adjusted to reproduce the experimental

(i) first-shell Zn^{2+}–O(water) distance,

(ii) hydration number of 6, and

(iii) the absolute Zn^{2+} hydration free energy.