Study of the structural complex between human MKK7 and human Gadd45B protein structures

Introduction: the biological problem

NF-kB/Rel factors control programmed cell death, and this control is crucial to oncogenesis, cancer chemoresistance, and antagonism of tumor necrosis factor (TNF) α-induced killing. With TNF, NF-kB-mediated protection involves suppression of the c-Jun-N-terminal kinase (JNK) cascade, and Gadd45β, a member of the Gadd45 family, was identified as a pivotal effector of this activity of NF-kB. Inhibition of TNFα-induced JNK signaling by Gadd45β depends on direct targeting of the JNK kinase, MKK7/JNKK2. The mechanism by which Gadd45β blunts MKK7, however, is unknown (Papa et al, JBC 2007, Tornatore et al, JMB 2008).

(Papa et al, JBC 2007) modelled the interaction between Gadd45β and MKK7 and validated the model through a number of complicated mutagenesis experiments.

The formation of the Gadd45β-MKK7 complex enables insertion of the Gadd45β acidic loop 1 into the MKK7 catalytic pocket, where this could engage in H-bonds and polar interactions with MKK7 Lys149 and other residues normally binding to the three-phosphate group of ATP. This engagement of the MKK7 ATP-binding site by Gadd45β loop 1 seems to prevent access of the kinase to ATP.
Therefore, the Gadd45β-MKK7 complex formation inhibits MKK7 by preventing it from binding ATP.
They also found that residues within putative Gadd45β acidic loops 1 and 2 – i.e. **Glu65**, **Glu66** and **Glu113** – directly contact basic residues within the catalytic pocket of MKK7 – i.e **Lys149**, **Lys157** and **Arg162**.

Before starting the actual study of the interaction, we need to do a number of things to get ready.

1) The crystal structure of MKK7 is available in the PDB. We need to identify the PDB code of the most appropriate conformation.

2) The crystal structure of Gadd45β is not available in the PDB. We need therefore to build a structural model using the homology modelling technique. We will do it with **HHPred** (http://toolkit.tuebingen.mpg.de/hhpred)

3) Structural models of complexes are more reliable if we can apply "restraints". These need to be extracted from our knowledge of the biology of the problem. In our case, we know that Gadd45β acidic loop 1 interacts with MKK7 Lys149, i.e. with the ATP binding site. Therefore, we have to identify such residue in the sequence of MKK7 using UniProt (**MKK UniProt AC: O14733**). We will use MKK7 Lys149 as constraint in building the model of the complex. This means that we also have to identify the ATP binding residue in the MKK7 PDB file. Notice that PDB residue numbering may differ from UniProt numbering.

4) Once we have collected these data, we are ready to use ClusPro (http://cluspro.bu.edu/home.php).

1. Identification of the PDB code of the MKK7 most appropriate conformation

This can be done a) by directly going to the PDB or b) passing through the UniProt.

a) By directly going to the PDB (http://www.rcsb.org/pdb/home/home.do)

Click on "Advanced search".
In the "Choose a Query Type" drop down menu choose "UniProtKB Accession Number(s)" and paste the MKK7 UniProt AC (**O14733**) in the text box.
Click on "Result Count". How many structures are available for MKK7?
Click on "Submit Query" and inspect the structures available. Which one seems the most appropriate for our purposes?

b) By passing through the UniProt
Go to UniProt (http://www.uniprot.org), type the MKK7 UniProt AC (O14733) in the text box at the top and click on Search.
On the result page, go to the structural information (you can scroll-down until you reach the Structure section or directly click on the "Structure" link on the left).
Select the RCSB PDB link destination. The structures available in the PDB for MKK7 will appear in the "Entry" column. You will have to inspect each of them (by clicking on their PDB codes) in order to decide which is the most appropriate for our purposes.

2. Homology model of Gadd45β

a) First, we have to identify the protein sequence of Gadd45β. Go to UniProt (http://www.uniprot.org), type the Gadd45β UniProt AC (**O75293**) in the text box at the top and click on Search. On the result page, go to the sequence information (you can scroll-down until you reach the Sequence section or directly click on the "Sequence" link on the left). Click on the FASTA link and copy the sequence in FASTA format.
b) Second, we have to identify a suitable template for Gadd45β. Go to HHPred and paste the Gadd45β in FASTA format into the Input text box. Inspect all the Search Options (but keep the default ones).
In the "Job Options" section, specify a Job-ID (e.g. gadd45B). Then Submit your job. The run may last up to a few minutes. However, this Gadd45β is a small protein (160 aa) and the template search should be quite fast.
c) In the HHpred result page, inspect the proposed templates. Which is the best one? Why? Go to the alignment between Gadd45β and its best template and take sometime to inspect it. Is it reliable? Do you think it might be manually improved? Has it a good coverage? What is the sequence identity? What is the e-value? Are the values of these two parameters good enough to proceed with the model building?
c) Once you are satisfied with the best template inspection, select it, then click on the "Create model" link immediately below the "Result, Histogram, etc." menu bar. In the

resulting page, check whether the selected best template is the one that you actually selected, then click on the "Create model from manually selected template(s)" button.
d) You will end up in the "Modeller" page. The target-template alignment will appear in the Input text box ("Paste multiple alignment"). Notice that, if you want to run a local version of Modeller, you have to copy this alignment and paste it to a local text file with .ali extension. This is NOT what we are going to do here. Here, we will use the Modeller installation provided by The Bioinformatics Toolkit. Therefore, check the options and insert a MODELLER-key (you can use "modeliranje") and a name for your Job in the Job-ID text box (I suggest "gadd45B_model" or something similar). Then, Submit your Job.
e) Congratulations! You have built a model for Gadd45β. You can Save it now.
f) After downloading it, open the file with your favourite Text Editor (gedit, vim, pico, nano, etc) and take a few minutes to inspect the file.
g) Before going back to the prediction of the Gadd45β-MKK7 complex, we will spend a few minutes to check the quality of our model.
Go to the QMEAN Server for Model Quality Estimation (http://swissmodel.expasy.org/qmean/cgi/index.cgi), provide a name for your project (I suggest gadd45B_model_quality) and upload the gadd45B_model.pdb file and submit your job. Notice that the quality assessment may take some time. We will discuss the result of the QMEAN run later.

3. Identification of Lys149 in the MKK7 protein sequence (O14733) and in the protein structure (PDB 2DYL).

a) Go to the "Function" section of the UniProt entry for MKK7. Check whether the ATP binding site residue has the same number as the one reported in the literature (K149).
b) Click on the residue Position(s) and identify the sequence context of K149 (e.g. the three residues before and after: IAV**K**QMR).
c) Go to the MKK7 PDB entry (http://www.rcsb.org/pdb/explore/explore.do?pdbId=2DYL) and click on the "Sequence" button in the menu bar. Identify the IAV**K**QMR sequence and the residue number corresponding to K149.
This can also be done by inspecting the PDB text file that you can download by selecting the "PDB file" option from the "Display Files" drop-down menu.
You should also verify to which PDB chain this residue belongs. 2DYL is a single chain structure

4. Modelling of the Gadd45β-MKK7 complex using ClusPro

Now we are ready to model the structure of the Gadd45β-MKK7 complex. The data we are going to use are:
1) MKK7 PDB code: 2DYL
2) The file containing the homology model for Gadd45β: gadd45B_model.pdb
3) The PDB number of the MKK7 residue interacting with the ATP (2DYL K165)

a) Go to the ClusPro webpage (http://cluspro.bu.edu/home.php). Here, you may want to create an account by clicking "Sign up for an account". You will receive an email with your account details that you can use to sign in.

b) We are not going to create an account. Therefore, click on "Use the server without the benefits of your own account".

c) In the ClusPro input page that will appear, you have to specify a Job Name (I suggest "mkk7-gadd45b"), insert the MKK7 PDB code (**2DYL**) in the Receptor text box and upload gadd45B_model.pdb for the Ligand. Do we need to also specify the chain? Why? Then, we will use the Advanced Options to set the restraint.

We will use the "Attraction and Repulsion" option to specify that the MKK7 K165 residue has attractive properties (toward Gadd45β). This can be done by specifying "a-165" in the "Attraction" text box for the Receptor.

d) Spend a few minutes inspecting the other options and try to understand what they do. In particular, what can you do with the "Restraints" option? How could we use this option in our case?

5. Analysis of ClusPro results

In the output page you can:

a) Review the details of your Job (what input structures? What options have you used?)

b) View the Model Scores. The most important aspect is to check whether there are docking solutions that form large clusters (in our case Cluster 0 has 127 members, which is a good number). In principle, you should choose – as your final solution - a conformation belonging to the largest cluster.

c) However, there are other parameters that count. An important one is whether the predicted complex fulfils the input restraints, i.e. (in our case), whether MKK7 Lys165 actually interacts with a loop of Gadd45β (and which loop). In order to check this, you have to inspect the models displayed in the output page (each model is a representative of a cluster) and download those that seem a potentially correct conformation.

d) Then you have to use molecular graphics software to display each model, highlight MKK7 Lys165 and verify its position relatively to Gadd45β. During this manual analysis, you may also use additional knowledge from the literature. For example (Papa et al, JBC 2007) predicted that Gadd45β acidic loop 1 interacts with Lys165.

e) In the output page, you will see four different choices for your docking results, "Balanced", "Electrostatic-favored", and so on. Which one should you choose?

CluPro provides many different options for docking because they believe good results go hand-in-hand with experimental knowledge of the complex. If you don't have any prior knowledge of what forces dominate in your complex, they recommend using the balanced coefficients.

For example, if you know from experiments that the receptor and the ligand have been observed to separate when subjected to ionic force *in vitro*, the interaction will likely be "Electrostatic-favored". Instead, if the two partners tend to separate when immersed in apolar solvents, they will likely be "Hydrophobic separated". If your complex is antibody-antigen, they recommend using the antibody mode.

f) In our case, we know that interactions between MKK7 and Gadd45β occur between basic (alkaline) MKK7 residues and acidic Gadd45 residues. Therefore, it makes sense to also have a look at "Electrostatic-favored" solutions.

6. Run a new docking experiment with further restraints from the literature

As mentioned in the biological introduction to the problem, (Papa et al, JBC 2007) found that residues within putative Gadd45β acidic loops 1 and 2 – i.e. **Glu65**, **Glu66** and **Glu113** – directly contact basic residues within the catalytic pocket of MKK7 – i.e **Lys149**, **Lys157** and **Arg162**.

Now we want to add some restraints and see what happens. In particular, we want to specify **Lys149**, **Lys157 as attractive residues for the receptor and Glu65, Glu66 as attractive residues for the ligand.** However, first we have to identify the residue numbers in the 3D structures, similarly to what we have done for MKK7 Lys149.

	sequence residue num	sequence context	structure residue num
MKK7	K149	IAV**K**QMR	K165 (**2DYL**)
	K157	SGN**K**EEN	K173 (**2DYL**)
Gadd45β	E65	IDEE**E**EDD	E65 (gadd45B_model.pdb)
	E66	IDEEE**E**DD	E66 (gadd45B_model.pdb)

Once we have the correct residue number in the protein structures, we can go to ClusPro and submit our query consisting in 2DYL PDB code for MKK7 and the gadd45B_model.pdb coordinate file for the model of Gadd45β. We can then use the "Attraction and Repulsion" option to specify that the MKK7 K165 and K173 residues have attractive properties toward Gadd45β residues E65 and E66, respectively. This can be done by specifying "a-165 a-173" in the "Attraction" text box for the Receptor and "a-65 a-66" in the "Attraction" text box for the Ligand.

You have to assign a name to your Job (I suggest mkk7-gadd45b-restr) and run it.

The procedure to analyse the docking result will be similar to the previous case. You can compare the results of the previous run (with only one restraint) and this one and see which are more accurate.

7. References

Docking
- Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. PROTEINS: Structure, Function, and Generics 47: 409-443
- Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A., Aflalo,C. and Vakser,I.A. (1992) Molecular surface recognition—determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl Acad. Sci. USA, 89, 2195–2199.
- Vakser,I.A. (1996) Low-resolution docking: prediction of complexes for underdetermined structures. Biopolymers, 39, 455–464.
- Ritchie,D.W. and Kemp,G.J.L. (2000) Protein docking using spherical polar Fourier correlations. Proteins, 39, 178–194.

Clustering

- Shortle,D., Simons,K.T. and Baker,D. (1998) Clustering of low-energy conformations near the native structures of small proteins. Proc. Natl Acad. Sci., USA, 95, 11158–11162
- Camacho,C.J. and Gatchell,D. (2003) Successful discrimination of protein interactions. Proteins, 52, 92–97

ClusPro

- Comeau, S. R., D. Gatchell, S. Vajda, and C. J. Camacho. 2004. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics. 20:45–50

Haddock

- Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *JACS* 125, 1731
- de Vries SJ, van Dijk M, Bonvin AMJJ (2010) The HADDOCK web server for data-driven biomolecular docking. Nature Protocols, 5, 883-897

Modeller

- Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali (2006) Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30
- M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 291-325
- Sali & T.L. Blundell (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815
- Fiser, R.K. Do, & A. Sali (2000) Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000

Swiss-Model

- Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22,195-201.
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*. 37, D387-D392.
- Guex, N., Peitsch, M.C., Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30(S1), S162-S173.

Other

- Chothia C, Lesk AM (1986): The relation between the divergence of sequence and structure in proteins. EMBO J. 5:823-836.
- Rost B (1999): Twilight zone of protein sequence alignments. Protein Eng.

12:85-94.

- Sander C, Schneider R (1991): Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 9:56-68.
- Venselaar H, Krieger E, Vriend G (2000) Homology Modeling
- Papa S, Monti SM, Vitale RM, Bubici C, Jayawardena S, Alvarez K, De Smaele E, Dathan N, Pedone C, Menotti Ruvo, Franzoso G (2007) Insights into the Structural Basis of the GADD45_-mediated Inactivation of the JNK Kinase, MKK7/JNKK2. JBC 282:19029-19041
- Tornatore L, Marasco D, Dathan N, Vitale RM, Benedetti E, Papa S, Franzoso G, Ruvo M, Monti SM (2008) J. Mol. Biol. 378: 97–111
- Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C (2010) Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure 18:1233-1243