Journal of Bioinformatics and Sequence Analysis Vol. 5(1), pp. 10-15, January, 2013
Available online at http://www.academicjournals.org/JBSA
DOI: 10.5897/JBSA11.022
ISSN 2141 2464 ©2013 Academic Journals

Full Length Research Paper

Generation of a 3D model for human cereblon using comparative modelling

Samina Bilal*, Shumaila Barkat Ali, Sahar Fazal and Asif Mir

Department of Computer Sciences and Bioinformatics, Mohammad Ali Jinnah University Islamabad Sihala Road, Pakistan.

Accepted 17 February, 2012

Three-dimensional (3D) protein structures provide helpful insights into the molecular association of a gene, its purpose also allow efficient drug designing experiments, such as the structure-based design of specific inhibitors. Recently, it has been shown that protein (cereblon) is involved in various tissues and brain and is revealed to be related with mental retardation. After this first report of cereblon (CRBN) involvement, it was necessary to further study this protein. Therefore a 3D structure of cereblon was developed using comparative modeling approach. By comparing the templates-target sequence, a model was created using MODELLER, a program for homology modeling. The accuracy of the predicted structure was checked using Ramachandran plot which showed that the residue falling in the favoured region was 88.4%. The predicted cereblon model can be used to understand the pathogenesis of mutations in cereblon that causes adenosine-5'-triphosphate (ATP)-dependent degradation of proteins in memory and learning.

Key words: CRBN, 3D structure, 1M4Y, MODELLER, comparative homology modeling.

INTRODUCTION

Mental retardation (MR) is the most common developmental disability and ranks first among the chronic conditions which are causing major activity limitations. A genetic or inherited metabolic etiology was implicated in two-thirds of mental retardation cases (Xin et al., 2008; Curry et al., 1997) and a recessive mode of transmission accounts for nearly one-fourth of these cases (Wright et al., 1959). Several of these cases affected basic cellular mechanisms such as cell signaling pathways (Matsuura et al., 1997; Albrecht et al., 1997; Costa et al., 2002; Xing et al., 1996) regulation of gene expression (Galdzicki et al., 2001; Shahbazian et al., 2002; Petrij et al., 1995; Wang et al., 1994; Darnell et al., 2001) and alterations in hippocampal dendrite morphology (Yang et al., 2004). There are three different types of non syndromic intellectual disability, one is X-linked (NS-XLMR), second is autosomal dominant (NS-ADID) and third one is autosomal recessive (NS-ARID) (Chelly et al., 2006). There are 6 genes accountable for NS-ARID these are TUSC3, TRAPPC9, PRSS12, CRBN, GRIK2 and CC2DIA. All of these six genes play a crucial role in the involvement of many biochemical pathways (Najmabadi et al., 2007). Higgins and colleagues found that the homozygous C > T nonsense mutation at nucleotide position 1,274 of a novel complementary deoxyribonucleic acid (cDNA) (1,274C > T) is involved in autosomal recessive nonsyndromic- mental retardation (ARNSMR) in a large kindred.

The gene was named CRBN (cereblon, NM_016302) based on its putative role in cerebral development and the presence of its large, highly conserved Lon domain. The nonsense mutation causing a premature stop codon in CRBN interrupts an N-myristoylation site and eliminates a casein kinase II phosphorylation site at the C terminus (Xin et al., 2008). CRBN are specifically involved in neural development and calcium signaling in the cerebral cortex and hippocampus. Although little is known about the function of human CRBN, its relationship to mild cognitive deficits suggests that it is involved in the basic processes of human memory and learning (Higgins et al., 2008). ARNSMR which is

*Corresponding author. E-mail: shumaila.ali86@gmail.com. Tel: +92519019351.

Table 1. Percentage similarity between target and template sequence.

Model number	Tool used	Template	Similarity (%)	No. of residues modeled
1	Modeller	1M4Y	56	442
2		2Q9U	29	442
3	3Djigsaw	------	------	105
4	CPH models	1M4Y	56	237
5	ESyPred3D	2ANE	8.4	107

associated with IQs ranging from 50 to 70 was widely studied (Xin et al., 2008). The predicted protein of CRBN, LON protease might belong to a family of ATP-dependent protease which is highly conserved across species (Lu et al., 2003).

Energy-dependent proteolysis plays a key role in prokaryotic and eukaryotic cells by regulating the availability of certain short-lived regulatory proteins, ensuring the proper stoichiometry of multiprotein complexes, and ridding the cell of abnormal and damaged proteins (Wang et al., 1993). Some ATP-dependent Lon proteases were reported to be associated with mental retardation. Comparative modeling is a useful technique in bioinformatics because this process constructs three dimensional models that are related to known protein structure (template) (Sali and Blundell, 1993; Marti-Renom et al., 2000). Thus this approach is relevant to structural based functional annotation. As a result, it enhances impact of structure and function on biology and medicine. By using bioinformatics tools, three dimensional structure of CRBN was constructed in the present study through comparative homology modeling approach.

MATERIALS AND METHODS

Comparative modeling consists of the following steps.

1. Search for related protein structures, selection of one or more templates.
2. Target-template alignment.
3. Model building and model evaluation. If model is not satisfactory, some or all of the steps can be repeated.

The amino acid sequence of cereblon was obtained from sequence database at National Center for Biotechnology Information (NCBI). It contains 442 amino acids. It was ensured that the three-dimensional structure of the protein was not available in Protein Data Bank therefore the present work of predicting the 3D modeling of the cereblon was performed. Template protein was searched by BLASTP, scanning the non redundant protein sequence database at NCBI with e-value cut off lesser than threshold, and retaining up to two templates with significant e-value. This searching provides us two templates. We used templates 1M4Y and 2Q9U high resolution X-ray diffraction in cereblon. Web based tools 3Djigsaw CPH models (Lund et al., 2002), ESyPred3D (Lambert et al., 2002) obtained templates automatically without any user intervention. 3Djigsaw looks for homologue templates in sequence database (PFAM+PDB+nr).

CPH models sought templates by iteratively aligning the target sequence to non redundant protein sequence database and searching the template pdb in protein structure database. ESyPred3D uses PSI-Blast at NCBI. All the templates are listed in Table 1. The target and template sequences were then aligned using the align2d command of MODELLER [http://www.salilab.org/modeller/8v1/] which uses global dynamic programming, with linear gap penalty function for aligning the two profiles. MODELLER takes target-template alignment file as input and without user intervention it generates 3D model. Initial step of model building is, identification of spatial restraints for example, distances and dihedral angles lying on the target sequence by aligning with template sequence. Interaction of many features of protein structure is analyzed statistically and used to derive spatial restraints on the target sequence. ESyPred3D use neural network method for increasing the alignment performance between the query and template sequence. CPH model uses profile-profile alignment between target and template.

Alignment between target and template that is, 1M4Y is shown in Figure 1 obtained through ClustalW web based tool. A three dimensional structure was developed from sequence alignment between CRBN and template protein using MODELLER8v1. It constructs model by satisfaction of spatial restraints. Distance and dihedral angle restraints on target sequence were derived from alignment with template structure. Stereochemical restraints such as bond angles and bond lengths were extracted from CHARM22 molecular mechanics force field. Statistical correlation of dihedral angles and non-bonded interatomic distance were extracted from database of family alignments that includes proteins with known three dimensional structures. CHARMM energy function and these spatial restraints were combined to obtain objective function. Final model was obtained by optimization of objective function using conjugate gradients and molecular dynamics with simulated annealing. 3Djigsaw, CPH models, ESyPred3D automatically build model by using their own set of modelling algorithms. CPH model uses segmod program from the GeneMine package.

It further refines the model using encad program from the GeneMine package. The constructed models were subjected to energy minimization by steepest descent, using GROMOS96 force field, implementation of Swiss-pdb Viewer Accuracy of the predicted model determines information that can be deduced from it; therefore all the models were subjected through a series of tests. Stereochemical properties were evaluated through procheck (Laskowski et al., 1993). Backbone conformation was evaluated by investigating PSi/Phi Ramachandran plot using Procheck and RAMPAGE (Sali, 1998). Packing quality and RMS of model was evaluated using Whatif packing quality control and protein analysis (Hooft et al., 1993).

RESULTS

Three dimensional structures are primarily important for providing valuable insight into molecular functions of

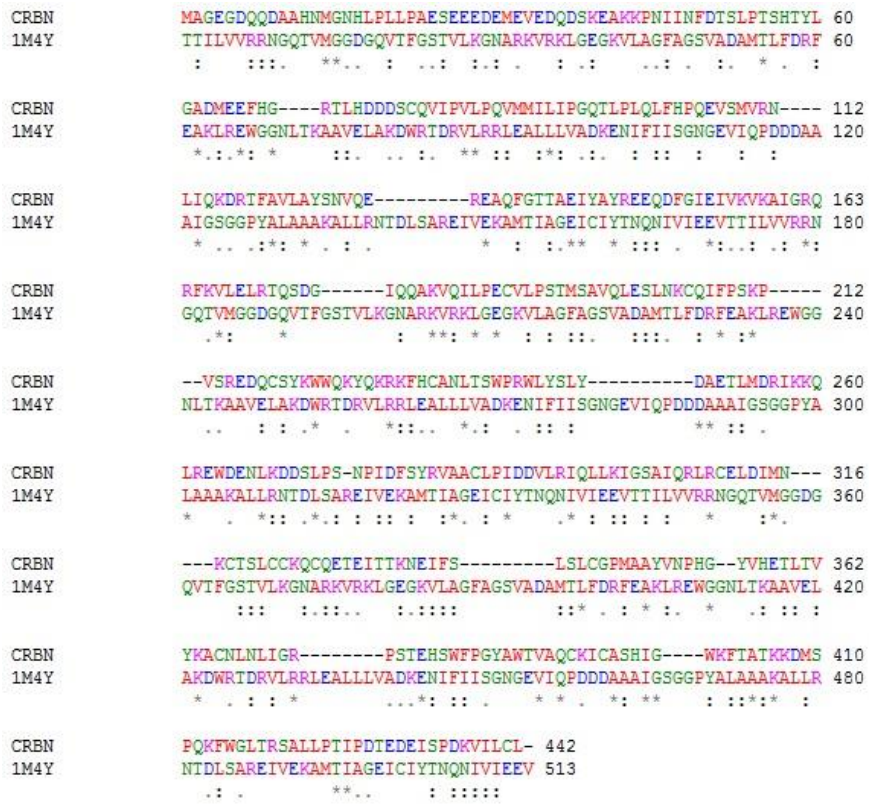

Figure 1. Alignment between CRBN and 1M4Y obtained through clustalW.

Table 2. Ramachandran plot values obtained through PROCHECK.

Model number	Ramachandran plot values			
	Core (%)	Allowed (%)	Generously (%)	Disallowed (%)
1	81.7	13.3	2.8	2.3
2	87.9	8.8	2.0	1.3
3	80.0	15.6	3.3	1.1
4	67.8	22.7	6.2	3.3
5	87.5	10.4	2.1	0.0

proteins. It plays a major role in site-directed mutagenesis, in studying disease related mutations and in drug designing process. Protein sequence of CRBN was obtained through NCBI, sequence database. Templates, high resolution X-ray diffraction 1M4Y and 2Q9U were obtained using blastp at NCBI. Secondary structure of predicted cereblon consists of 166 hydrogen bonds, 6 Helices, 11 strands and 58 Turns. Webs based tools obtained templates automatically and are shown in Table 1. Comparative modeling builds a three dimensional structure of the target protein based on sequence identity to known protein structures (Sali, 1998; Vitkup et al., 2001). Therefore, sequence identity is good determinant for the quality of the model. In general, sequence of at least one related structure must have more than 30% identity (Roberto and Andrej, 2000). Sequence similarity between target and template sequences has been shown in Table 1. Among the different alignments, the more related alignment is of models obtained through MODELLER and CPH models.

The template most used tools is 1M4Y. MODELLER and web-based tools were used for building the model and global energy minimization. After model building, the structures were validated through energy minimization. Refined models were checked through PROCHECK and RAMPAGE.

Values for the Ramachandran plot obtained through Procheck are shown in Table 2. The plot is subdivided in core, allowed, generously allowed and disallowed regions. The models obtained through MODELLER and

Table 3. Ramachandran plot values obtained through RAMPAGE.

Model number	Ramachandran plot values		
	No of residues in favoured region (%)	No of residues in allowed region (%)	No of residues in outlier region (%)
1	88.4	6.8	4.8
2	93.2	4.5	2.3
3	90.0	7.0	3.0
4	81.3	11.7	7.0
5	95.3	2.8	1.9

Table 4. Ramachandran plot values obtained through Whatif server.

Model number	Z-Score
1	-1.094
2	-0.549
3	-3.600
4	-5.612
5	-0.754

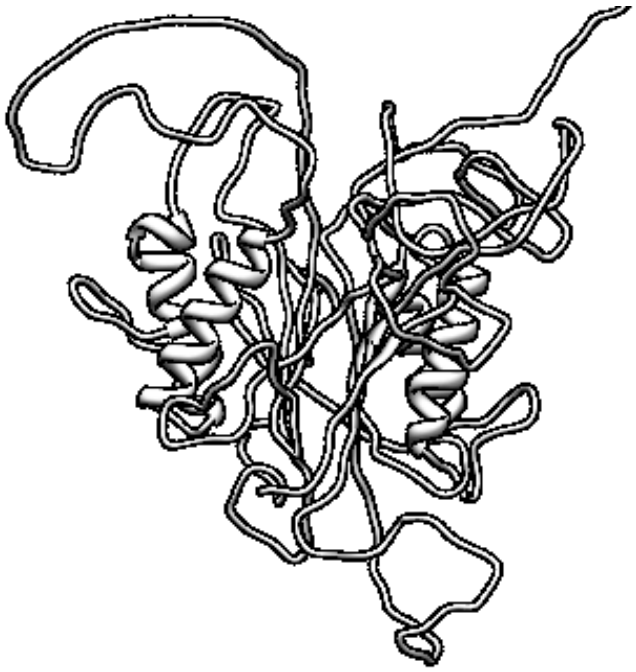

Figure 2. 3D Structure of human CRBN in raswin 2.7.5.

Esypred3D showed better Ramachandran plot values, as core region (>80%) accounts for better structure (Morris et al., 1992). Rampage assessment is shown in Table 3. Rampage derives Phi/Psi plots for Gly, Pro, Pre-Pro and other residues. The plot was divided into three regions that is, favored, allowed, and outlier regions. The result for models obtained through MODELLER and Esypred were significant, as denser number of residues in favored region (>90%) is the measure of good quality of a model (Morris et al., 1992), but Esypred3D created the model for 107 residues while MODELLER created the model for all 442 residues.

Values for Ramachandran plot obtained through Whatif Server are shown in Table 4. The score expressing how well the backbone conformations of all residues are corresponding to the known allowed areas in the Ramachandran plot is within expected ranges for well-refined structures. These results demonstrate that prediction of the best possible target would be a difficult task because the target performing well in one case was not found good in other cases. Esypred3D model tends to have better stereochemistry, whereas it does not hold good sequence similarity and is modeled for 107 residues only. For all the targets described herein, the structure obtained through MODELLER, using 1M4Y template, was found to be satisfactory based on the above results. This model is shown in Figure 2. Ramachandran plot analysis through procheck showed that 81.7% residues are within the core region (Figure 3). RMS and packing quality was evaluated through Whatif and found satisfactory for this model.

DISCUSSION

The identification of human monogenic disorders that solely affect cognition provides rare opportunities to study the mechanisms of human memory and learning. Studies have shown that CRBN, results in a phenotype characterized by mild mental retardation without the presence of dysmorphic features or physical anomalies (Higgins et al., 2004; 2000). The CRBN protein contains an N-terminal domain of the ATP-dependent Lon protease, a regulator of G-protein- signaling-like domain, a leucine zipper motif, and four putative protein kinase C phosphorylation sites (Jo et al., 2005). CRBN is highly expressed in the hippocampus and cerebral cortex. CRBN is also involved in regulating the surface expression and electrical properties of BKCa channels (Darnell et al., 2001). Mutations in CRBN cause

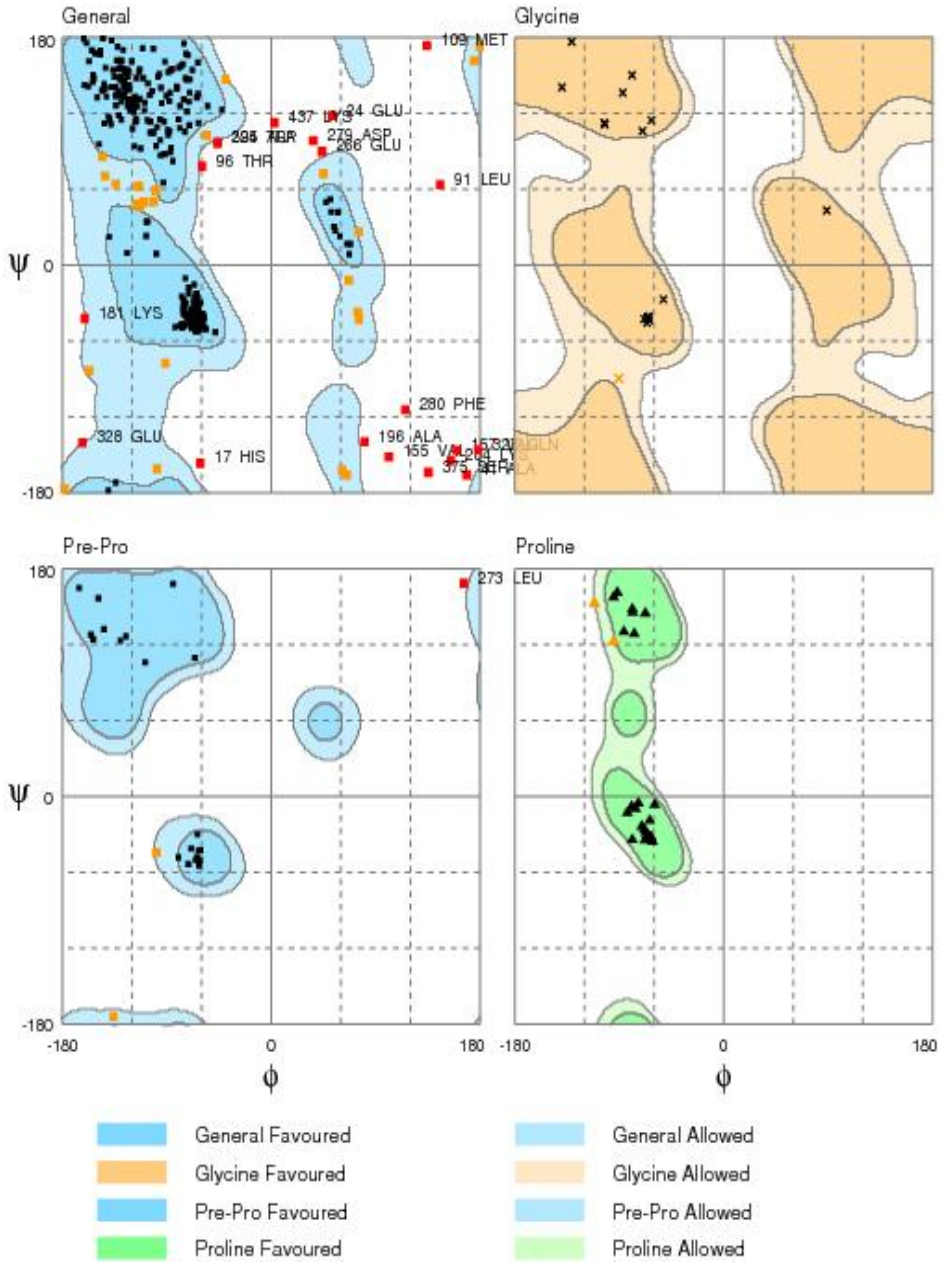

Figure 3. RAMPAGE values for best predicted model showing number of residues in favoured, allowed and outlier region.

mental retardation. This three dimensional structure of CRBN is useful for studying these disease-related mutations. Further analysis of this structure will help in finding binding clefts, for finding novel drug leads.

REFERENCES

Albrecht U, Sutcliffe JS, Cattanach BM (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat. Genet. 17:75-78 biologie/urbm/bioinfo/esypred/)

Costa RM, Federov NB, Kogan JH (2002). Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. Nature 415:526-530

Curry CJ, Stevenson RE, Aughton D (1997). Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. Am. J. Med. Genet. 72:468-477.

Darnell JC, Jensen KB, Jin P (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell 107:489-499

Galdzicki Z, Siarey R, Pearce R (2001). On the cause of mental retardation in Down syndrome: extrapolation from full and segmental trisomy 16 mouse models. Brain Res. Rev. 35:115-145.

Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP (2004). A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild

mental retardation. Neurology 63:1927-1931.
Higgins JJ, J Hao, BE Kososky (2008). Dysregulation of large conductance Ca^{2+} -activated K^{+} channel expression in nonsyndromal mental retardation due to a cerebelon p.R419X mutation. Neurogenetics. 9:219-223.
Higgins JJ, Rosen DR, Loveless JM, Clyman JC, Grau MJ (2000). A gene for nonsyndromic mental retardation maps to chromosome 3p25-pter. Neurology 55:335-340.
Hooft RWW, Vriend G, Sander C, Abola EE (1993). Errors in protein structures. Nature 381: 272-272. (Server: http://swift.cmbi.kun.nl/WIWWWI/)
Jo S, Lee KH, Song S, Jung YK, Park CS (2005). Identification and functional characterization of cereblon as a binding protein for large-conductance calcium-activated potassium channel in rat brain. J. Neurochem. 94:1212–1224
Lambert C, Leonard N, De Bolle X, Depiereux E (2002). ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18(9):1250-1256. (Server:http://www.fundp.ac.be/sciences)
Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst., 26: 283-291. (Server: http://www.csb.yale.edu/userguides/datamanip/prochec k/)
Lu B, Liu T, Crosby JA (2003). The ATP-dependent Lon protease of *Mus musculus* is a DNA-binding protein that is functionally conserved between yeast and mammals. Gene 306:45-55.
Lund O, Nielsen M, Lundegaard C, Worning P (2002). X3M a Computer Program to Extract 3D Models. CASP5 conference A102. (Server:http://www.cbs.dtu.dk/services/CPHmodels/output.php)
Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29:291-325.
Matsuura T, Sutcliffe JS, Fang P (1997). *De novo* truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat. Genet. 15:74-77.
Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992). Stereochemical quality of protein structure coordinates. Proteins 12:345-364.
Petrij F, Giles RH, Dauwerse HG (1995). Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature 376:348–351
Roberto S, Andrej S (2000). Comparative protein structure modeling: Introduction and practical examples with MODELLER. In: Protein Structure Prediction: Methods and Protocols. Ed: Webster, D. M., 97-129. Humana Press.
Sali A (1998). 100,000 protein structures for the biologist. Nat. Struct. Biol. 5:1029-1032.
Sali A, Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815.
Shahbazian M, Young J, Yuva-Paylor L (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron 35:243-254.
Vitkup D, Melamud E, Moult J, Sander C (2001). Completeness in structural genomics. Nat. Struct. Biol. 8(6):559-566.
Wang N, Gottesman S, Willingham MC (1993). A human mitochondrial ATPdependent protease that is highly homologous to bacterial Lon protease. Proc. Natl. Acad. Sci. USA 90:11247-11251.
Wang YH, Amirhaeri S, Kang S (1994). Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. Science 265:669-671.
Wright SW, Tarjan G, Eyer L (1959). Investigation of families with two or more mentally defective siblings; clinical observations. Am. J. Dis. Child 97:445-463
Xin W, Xiaohua N, Peilin C (2008). Primary function analysis of human mental retardation related gene CRBN. Mol. Biol. Rep. 35:251-256.
Xing J, Ginty DD, Greenberg ME (1996). Coupling of the RASMAPK pathway to gene activation by RSK2, a growth factor regulated CREB kinase. Science 273:959-963
Yang EJ, Yoon JH, Min do S (2004). LIM kinase 1 activates cAMP-responsive element-binding protein during the neuronal differentiation of immortalized hippocampal progenitor cells. J. Biol. Chem. 279:8903–8910.