

# Molecular modeling of DICER and identification of phosphorylation sites

Priyanka Dhar\*, Sayak Ganguli, Abhijit Datta

Defence Institute of High Altitude Research, Defence Research & Development Organization, C/o 56 APO, 901205, India; Priyanka Dhar  
E-mail: priyankodrd@gmail.com; Tel / Fax: 01982-252096; \*corresponding author.

Received October 29, 2009; accepted November 15, 2009; published March 31, 2010

## Abstract:

RNA interference (RNAi) is a post-transcriptional process initiated by double-stranded RNA molecules that degrade a complementary target RNA. In the first step, long double-stranded RNA molecules are chopped into shorter duplexes by an endonuclease dubbed DICER. This work focused on the identification of phosphorylation sites in DICER and proceeded towards comparative modeling. The domain of unknown function in DICER showed homology with the chromatin remodeling protein domains. This suggests that DICER may also perform chromatin remodeling. The Ramachandran analysis revealed that the models had 96% residues in the permissible regions and could be used for future ligand binding studies.

**Keywords:** RNAi, DICER, protein phosphorylation, chromatin remodeling, conserved domain.

## Background:

A conserved biological response to double-stranded RNA, known variously as RNA interference (RNAi) or post-transcriptional gene silencing, mediates resistance to both endogenous parasitic and exogenous pathogenic nucleic acids, and regulates the expression of protein-coding genes. RNAi has been cultivated as a means to manipulate gene expression experimentally and to probe gene function on a whole-genome scale. It is an ancient gene-silencing process that plays a fundamental role in a number of important functions, including viral defense, chromatin remodeling, genome rearrangement, development, morphogenesis and stem cell maintenance. RNA interference activities depend upon DICER, an enzyme's molecular structure is a critical step. Specificity of RNAi-dependent gene silencing is provided by small (21–26 nt) guide RNAs. Two main types of small RNAs that differ in origin and function are known to play integral roles in RNAi: siRNAs (small interfering RNAs) and miRNAs (micro- RNAs). siRNAs are the cleavage products of longer dsRNAs that are generated by RNA-dependent RNA polymerases, or from bidirectional transcription of genes or transposable elements. The siRNAs usually guide the RNAi machinery to carry out mRNA degradation and/or chromatin modification.

Based on genetic and biochemical studies, the canonical RNAi pathway has been divided into two stages: initiation and effector. The initiation stage involves the generation of siRNAs [1-4] and miRNAs [5, 6] from long dsRNA or hairpin RNA precursors respectively. Cleavage of precursors into si/miRNAs is mediated by the type III RNase Dicer [7]. The effector stage requires the transfer of the si/miRNAs into ribonucleoprotein complexes known as RISCs (RNA-induced silencing complexes) [8]. Functional RISCs contain only single-stranded siRNAs or miRNAs [9]. Gene silencing by RISC is accomplished via homology-dependent mRNA degradation [10-12], translational repression [13] or transcriptional gene silencing [14-16]. Its purported function in gene-silencing pathways is to amplify siRNAs and possibly miRNAs in order to increase the speed and/or the magnitude of the RNAi response after detection of the dsRNA trigger. This work focused on the identification of phosphorylation sites in the amino acid sequences of DICER, search for conserved domains, comparative modeling of DICER in seven important model organisms and proceeded towards phylogenetic analysis of DICER with the Chromatin Remodeling Complexes.

## Methodology:

### Dataset

The work was performed absolutely *in silico* and the materials are existing database sequences available at the public databases such as Swiss-Prot and GenBank of NCBI [17, 18]. The accession numbers of the DICER protein sequences used for the proteomic analysis are provided in Table 1 and Table 2.

## Computational proteomics

The phosphorylation sites in the DICER protein sequences were located using the NetPhos 2.0 server [19]. Then the conserved domains were identified at the NCBI conserved domain database [20]. After that the DICER sequences were modeled using the software Modeller 9v7 and tested using the Deepview package of the SWISS PDB viewer v4.0 [21, 22]. Finally the modeled structures were validated by the Ramachandran Plot analysis using the MolProbity analyzer [23].

## Phylogenetic analysis

The phylogenetic analysis of DICER with the Chromatin Remodeling Complexes was done using the PIR ClustalW and EBI ClustalW 2 tools [24, 25].

## Results:

### Detection of phosphorylation sites

Several phosphorylation sites were detected in the DICER protein sequences (Figure 1; Figure 2) indicative of the fact that DICER may also be regulated by the mechanism of protein phosphorylation. The three most abundant residues that are most frequently phosphorylated are serine, tyrosine and threonine.

### Recognition of conserved domain

The search for conserved domains reveals that DICER has five different conserved domains along its entire sequence. The conserved domains are: RIBOC (ribonuclease domain), the PAZ (Piwi, Argonaute and Zwillig) domain, the Helicase domain, the Domain of unknown function and the Dead box helicase domain.

### Molecular modeling

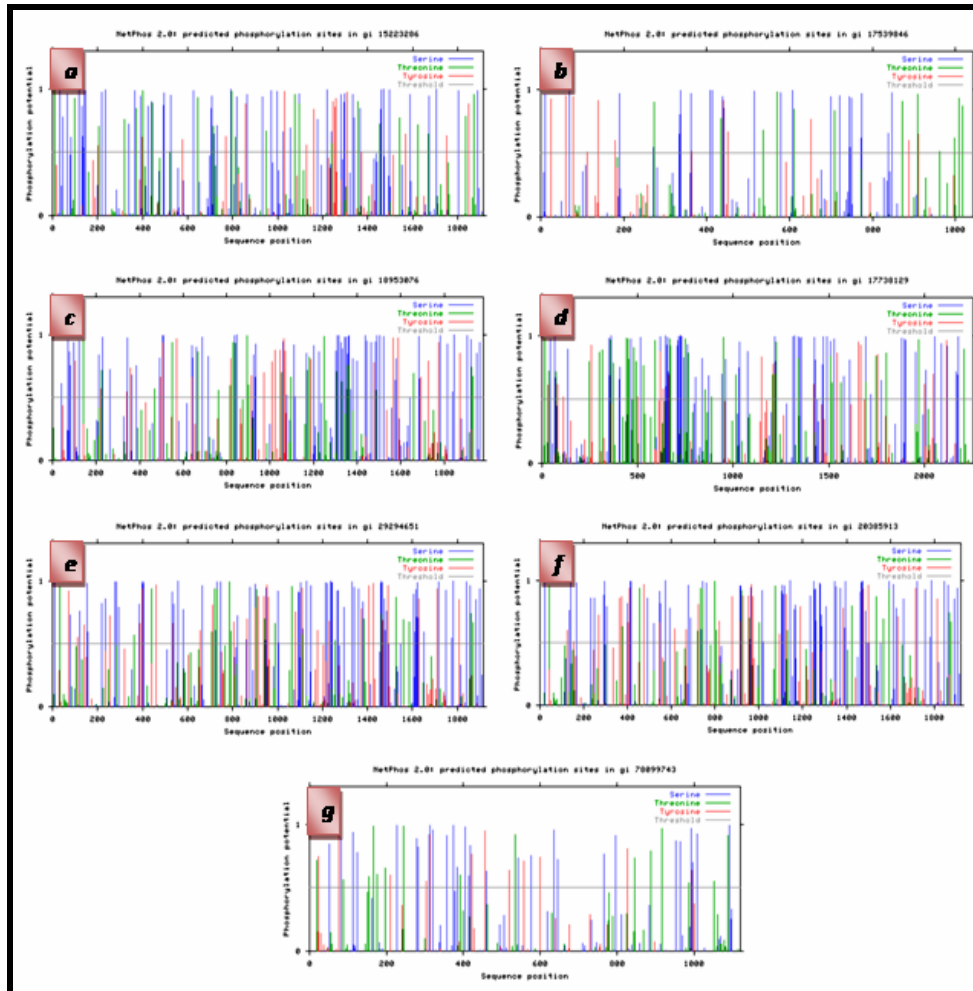
Molecular modeling of DICER was achieved using the “Modeller” software (Figure 3). The structure obtained reveals a basal helix-loop-helix domain and as evolution has occurred the complexity of the structure has gradually increased with the addition of pleated sheets.

### Model validation by Ramachandran analysis

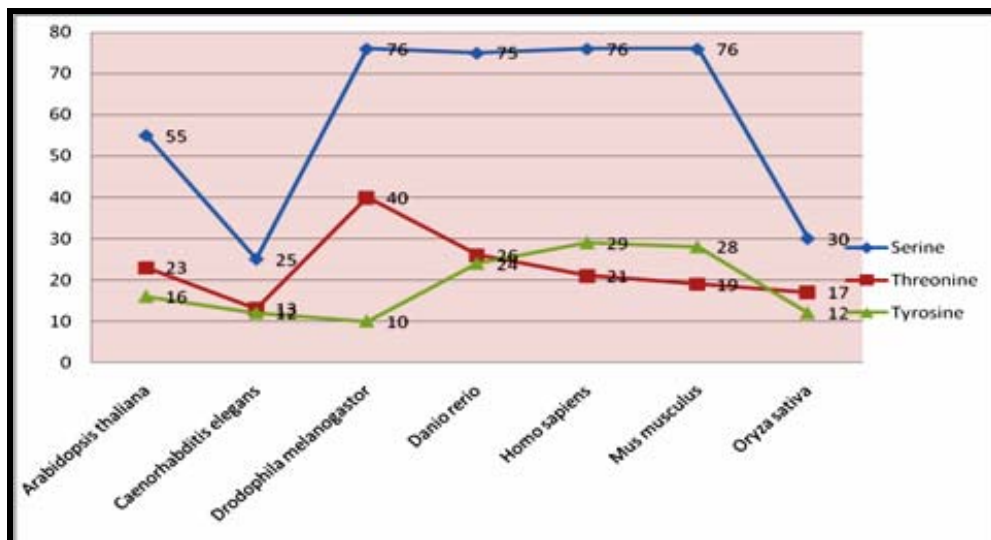
The Ramachandran analysis revealed that the models had 96% residues in the allowable regions (Figure 4).

### Phylogenetic analysis

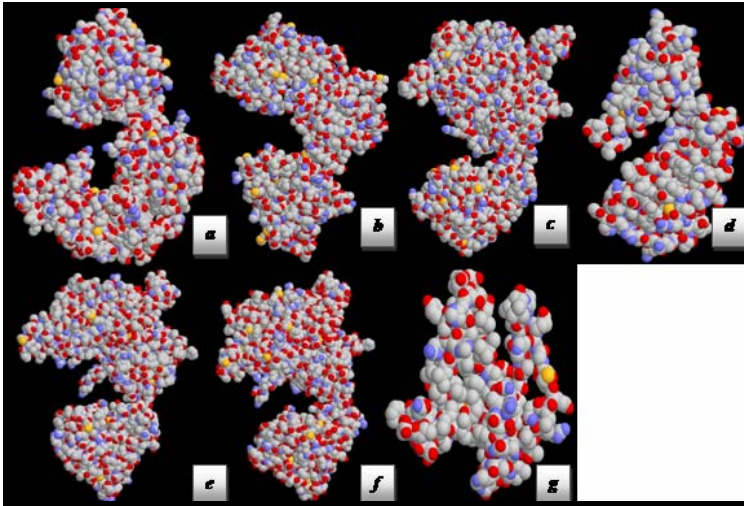
Multiple sequence alignment (PIR, ClustalW) indicates that subunit of the SWI/SNF chromatin remodeling complex that regulates transcription by remodeling chromosomes, required for transcription of many genes, including ADH1, ADH2, GAL1, HO, INO1 and SUC2; Swi3p of *Saccharomyces cerevisiae* and dicer1 of *Homo sapiens* originates from the same node (Figure 5). On the other hand multiple sequence alignment (EBI, ClustalW 2) shows that Dicer Related Helicase family member (drh-1) of *Caenorhabditis elegans* and ATP-dependent chromatin remodeling protein SNF2-related protein of *Danio rerio* originates from the same node (Figure 6).



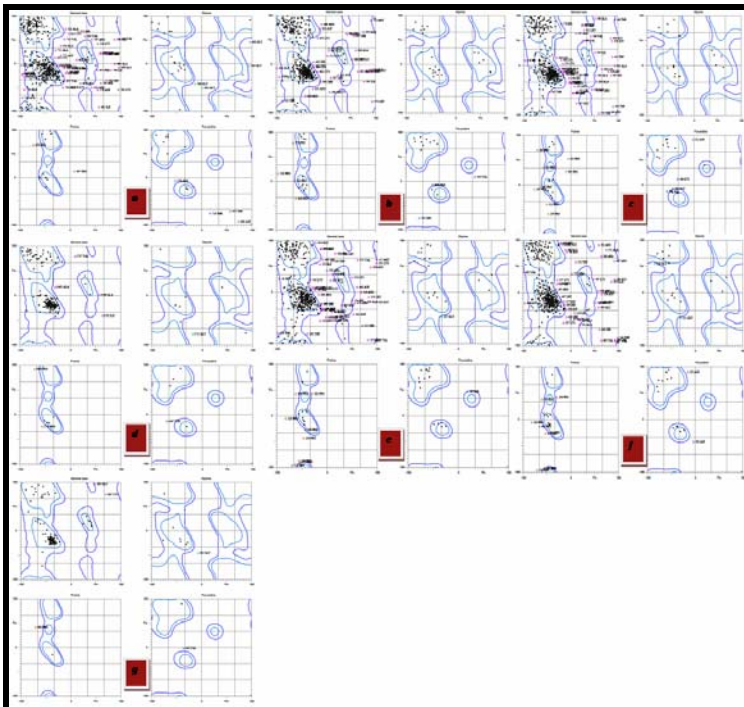
**Figure 1:** Identification of phosphorylation sites of DICER sequences of seven model organisms. a, *Arabidopsis thaliana* b, *Caenorhabditis elegans* c, *Danio rerio* d, *Drosophila melanogaster* e, *Homo sapiens* f, *Mus musculus* g, *Oryza sativa*



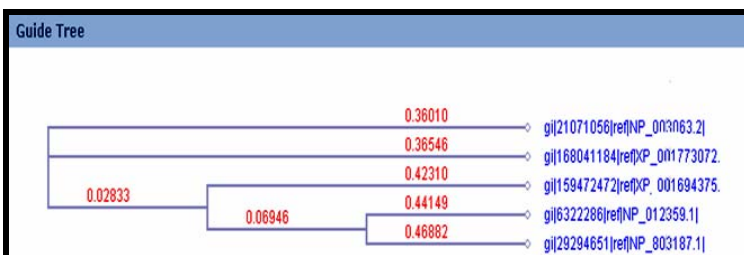
**Figure 2:** Predicted phosphorylation sites on DICER in seven model organisms



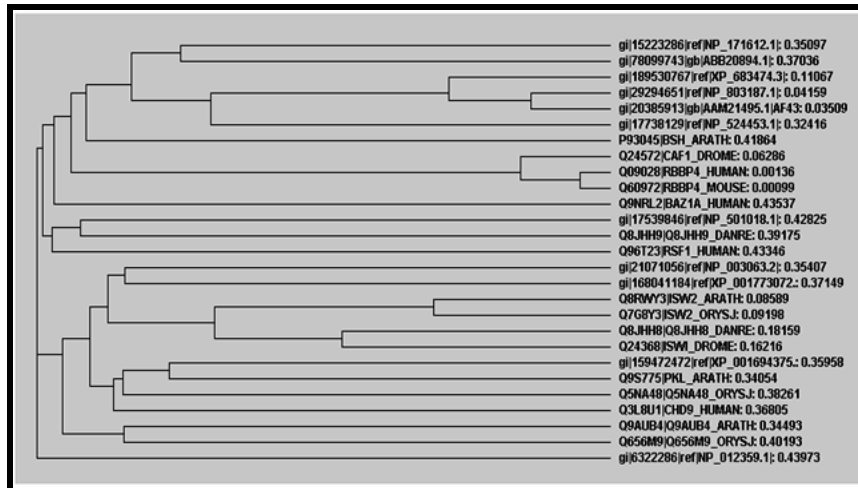
**Figure 3:** Molecular modeling of DICER sequences of seven model organisms. a, *Arabidopsis thaliana* b, *Caenorhabditis elegans* c, *Danio rerio* d, *Drosophila melanogaster* e, *Homo sapiens* f, *Mus musculus* g, *Oryza sativa*



**Figure 4:** Model validation: Ramachandran plot analysis of DICER sequences of seven model organisms. a, *Arabidopsis thaliana* b, *Caenorhabditis elegans* c, *Danio rerio* d, *Drosophila melanogaster* e, *Homo sapiens* f, *Mus musculus* g, *Oryza sativa*



**Figure 5:** Multiple sequence alignment (PIR) shows that >gi|6322286|ref|NP\_012359.1| Subunit of the SWI/SNF chromatin remodeling complex, which regulates transcription by remodeling chromosomes; required for transcription of many genes, including ADH1, ADH2, GAL1, HO, INO1 and SUC2; Swi3p [*Saccharomyces cerevisiae*] and >gi|29294651|ref|NP\_803187.1| dicer1 [*Homo sapiens*] were originated from the same node.



**Figure 6:** Multiple sequence alignment (EBI ClustalW 2) shows that >gi|17539846|ref|NP\_501018.1| Dicer Related Helicase family member (drh-1) [*Caenorhabditis elegans*] and >Q8JHH9|Q8JHH9\_DANRE ATP-dependent chromatin remodeling protein SNF2-related protein (fragment) from *Danio rerio* (Zebrafish) and *Brachydanio rerio* were originated from the same node.

**Discussion:**

Protein phosphorylation is known to play a key role in regulating a variety of processes in eukaryotes, from the cell division cycle to neuronal plasticity. The most commonly observed phosphorylations affect serine, threonine, and tyrosine residues although phosphorylation of histidines and aspartates has also been reported. Protein phosphorylation is catalyzed by enzymes called protein kinases, which are usually specific for either tyrosine or serine/threonine, with few of them being able to modify all three residues indistinguishably. The results obtained reveals that the most abundant phosphorylation site in case of DICER is Serine residues with over 50 sites in each protein under study. The protein from *Caenorhabditis elegans* has the least number of phosphorylation sites while that from *Homo sapiens* and *Drosophila melanogaster* has the largest number of phosphorylation sites (Figure 2). One of the most interesting observations from the molecular modeling is that the basic structure shares significant homology with the chromatin remodeling proteins. This suggests that DICER apart from being predominantly a ribonuclease may also perform chromatin remodeling at some stage. The Ramachandran analysis confirmed that the models were satisfactory and could be used for future ligand binding studies. Phylogenetic analysis by multiple sequence alignment showed that chromatin remodeling protein Swi3p of *Saccharomyces cerevisiae* and dicer1 of *Homo sapiens* originates from the same node. Likewise drh-1 of *Caenorhabditis elegans* and ATP-dependent chromatin remodeling protein SNF2-related protein of *Danio rerio* originates from the same node.

**Conclusion:**

RNAi technology is a recent advancement and by this we can treat many complex disorders by gene silencing. It has other immense uses also. DICER is an endonuclease that plays a significant role in the RNAi pathway. The study focused on the detection of the phosphorylation sites in the DICER protein as it plays the role of key regulator of protein activity. The other aspects of the work are the molecular modeling of DICER, conserved domain analysis and phylogenetic analysis of DICER with chromatin remodeling proteins. It is the first in the literature that reports the *in-silico* analysis of DICER, the regulator of RNA induced silencing complex (RISC). From multiple sequence alignment it can be concluded that DICER and chromatin remodeling proteins are predominantly xenologues but also have the orthologous properties.

The present study is thus revealed different aspects of DICER which is significant in the context of RNAi technologies and can usher new era of interference.

**References:**

[1] A Hamilton *et al.*, *EMBO J.* **21**: 4671 (2002) [PMID: 12198169]  
 [2] RC Lee, V Ambros, *Science* **294**: 862 (2001) [PMID: 11679672]  
 [3] NC Lau *et al.*, *Science* **294**: 858 (2001) [PMID: 11679671]  
 [4] M Lagos-Quintana *et al.*, *Science* **294**: 853 (2001) [PMID: 11679670]  
 [5] BJ Reinhart *et al.*, *Genes Dev.* **16**: 1616 (2002) [PMID: 12101121]  
 [6] W Park *et al.*, *Curr. Biol.* **12**: 1484 (2002) [PMID: 12225663]  
 [7] E Bernstein *et al.*, *Nature (London)* **409**: 363 (2001) [PMID: 11201747]  
 [8] SM Hammond *et al.*, *Nature (London)* **404**: 293 (2000) [PMID: 10749213]  
 [9] J Martinez *et al.*, *Cell* **110**: 563 (2002) [PMID: 12230974]  
 [10] T Tuschl *et al.*, *Genes Dev.* **13**: 3191 (1999) [PMID: 10617568]  
 [11] PD Zamore *et al.*, *Cell* **101**: 25 (2000), [PMID: 10778853]  
 [12] AJ Hamilton, DC Baulcombe, *Science* **286**: 950 (1999) [PMID: 10542148]  
 [13] A Grishok *et al.*, *Cell* **106**: 23 (2001) [PMID: 11461699]  
 [14] IM Hall *et al.*, *Science*, **297**, 2232 (2002) [PMID: 12215653]  
 [15] M Pal-Bhadra *et al.*, *Mol.Cell*, **9**: 315 (2002) [PMID: 11864605]  
 [16] TA Volpe *et al.*, *Science* **297**: 1833 (2002) [PMID: 12193640]  
 [17] <http://www.ncbi.nlm.nih.gov/>  
 [18] <http://www.expasy.ch/sprot/>  
 [19] <http://www.cbs.dtu.dk/services/NetPhos/>  
 [20] <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>  
 [21] <http://salilab.org/modeller/>  
 [22] <http://spdbv.vital-it.ch/>  
 [23] <http://molprobitry.biochem.duke.edu/>  
 [24] <http://pir.georgetown.edu/pirwww/search/multialn.shtml>  
 [25] <http://www.ebi.ac.uk/Tools/clustalw2/index.html>

Edited by P. Kanguane

Citation: Dhar, Bioinformatics 4(9): 412-416 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Accession numbers of the DICER protein sequences used for the proteomic analysis

GenBank accession	Name	Source
gi 15223286 ref NP_171612.1	DCL1 (DICER-LIKE1); ATP-dependent helicase/ ribonuclease III	<i>Arabidopsis thaliana</i>
gi 17539846 ref NP_501018.1	Dicer Related Helicase family member (drh-1)	<i>Caenorhabditis elegans</i>
gi 189530767 ref XP_683474.3	PREDICTED: Dicer1, Dcr-1 homolog	<i>Danio rerio</i>
gi 17738129 ref NP_524453.1	Dicer-1 CG4792-PA	<i>Drosophila melanogaster</i>
gi 29294651 ref NP_803187.1	dicer1	<i>Homo sapiens</i>
gi 20385913 gb AAM21495.1	AF430845_1 dicer-like protein	<i>Mus musculus</i>
gi 78099743 gb ABB20894.1	dicer-like protein	<i>Oryza sativa (indica cultivar-group)</i>

Table 2: Accession numbers of the chromatin remodeling protein sequences

GenBank/ Swiss-Prot accession	Name	Source
gi 21071056 ref NP_003063.2	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a4 isoform B	<i>Homo sapiens</i>
gi 159472472 ref XP_001694375.1	SWI/SNF chromatin remodeling complex component	<i>Chlamydomonas reinhardtii</i>
gi 6322286 ref NP_012359.1	Subunit of the SWI/SNF chromatin remodeling complex, which regulates transcription by remodeling chromosomes; required for transcription of many genes, including ADH1, ADH2, GAL1, HO, INO1 and SUC2; Swi3p	<i>Saccharomyces cerevisiae</i>
gi 168041184 ref XP_001773072.1	SWI/SNF class chromatin remodeling complex protein	<i>Physcomitrella patens subsp. patens</i>
P93045 BSH_ARATH	Chromatin structure-remodeling complex protein BSH	<i>Arabidopsis thaliana</i>
Q8RWY3 ISW2_ARATH	Putative chromatin-remodeling complex ATPase chain	<i>Arabidopsis thaliana</i>
Q9AUB4 Q9AUB4_ARATH	Putative chromatin remodeling protein SYD	<i>Arabidopsis thaliana</i>
Q9S775 PKL_ARATH	CHD3-type chromatin-remodeling factor PICKLE	<i>Arabidopsis thaliana</i>
Q8JHH9 Q8JHH9_DANRE	ATP-dependent chromatin remodeling protein SNF2-related protein (Fragment)	<i>Danio rerio</i>
Q8JHH8 Q8JHH8_DANRE	ATP-dependent chromatin remodeling protein SNF2-related protein (Fragment)	<i>Danio rerio</i>
Q24368 ISWI_DROME	Chromatin-remodeling complex ATPase chain Iswi	<i>Drosophila melanogaster</i>
Q24572 CAF1_DROME	Probable histone-binding protein Caf1	<i>Drosophila melanogaster</i>
Q3L8U1 CHD9_HUMAN	Chromodomain-helicase-DNA-binding protein 9	<i>Homo sapiens</i>
Q9NRL2 BAZ1A_HUMAN	Bromodomain adjacent to zinc finger domain protein 1A	<i>Homo sapiens</i>
Q96T23 RSF1_HUMAN	Remodeling and spacing factor 1	<i>Homo sapiens</i>
Q09028 RBBP4_HUMAN	Histone-binding protein RBBP4	<i>Homo sapiens</i>
Q60972 RBBP4_MOUSE	Histone-binding protein RBBP4	<i>Mus musculus</i>
Q5NA48 Q5NA48_ORYSJ	Putative chromatin remodeling factor CHD3	<i>Oryza sativa subsp. japonica</i>
Q7G8Y3 ISW2_ORYSJ	Probable chromatin-remodeling complex ATPase chain	<i>Oryza sativa subsp. japonica</i>