
Prediction of mutations in ATM gene responsible for breast cancer in women – a cancer-informatics approach

R. Sindhuja

Department of Bioinformatics,
School of Chemical and Biotechnology,
Shanmugha Arts Science Technology & Research Academy
(SASTRA University),
Tanjore-613402, Tamilnadu, India
E-mail: rsindhu86@gmail.com

R. Seenivasagam

Department of Drug Discovery,
St. Joseph College,
P.B. 27094, 36, Lalbagh Road, Bangalore, Karnataka 560027, India
E-mail: seenivasagam.pharma@gmail.com

K. Hemavathi*

Department of Bioinformatics,
School of Chemical and Biotechnology,
Shanmugha Arts Science Technology & Research Academy
(SASTRA University),
Tanjore-613402, Tamilnadu, India
Fax: +04362 264120
E-mail: hema.bioinfo@gmail.com
*Corresponding author

Abstract: Cancer is an abnormal growth of cells caused by multiple changes in gene expression. Breast cancer is a cancer of the glandular breast tissue. The BRCA1 and BRCA2 genes were considered as greatly increasing occurrence of inherited breast cancer, however, it was found that the damaged ataxia telangiectasia mutated (ATM) genes were also equal culprits. The ATM gene (NC_000011.8) is very large and mutations occur throughout its amino acid coding portions. The present work discusses the use of bioinformatics tools in analysing the ATM gene's role in breast cancer.

Keywords: systems biology; in-silico analysis; mutations; breast cancer; immunogenetics.

Reference to this paper should be made as follows: Sindhuja, R., Seenivasagam, R. and Hemavathi, K. (2011) 'Prediction of mutations in ATM gene responsible for breast cancer in women – a cancer-informatics approach', *Int. J. Immunological Studies*, Vol. 1, No. 3, pp.229–248.

Biographical notes: R. Sindhuja graduated with MSc from the Department of Genetics at University of Leicester, UK. She completed her BTech from Department of Bioinformatics at SASTRA University.

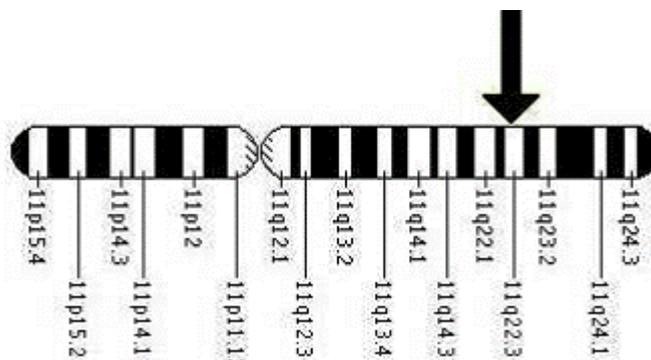
R. Seenivasagam is an MPharm graduate from the Department of Pharmacoinformatics, SASTRA University. He has been researching in drug design, drug development and drug discovery through pharmacoinformatic and pharmacogenomics approach.

K. Hemavathi is a faculty in the Department of Bioinformatics, SASTRA University. She has been researching in computational biology and biocomputing.

1 Introduction

Ataxia telangiectasia is a rare autosomal recessive disease (frequency of 1/40,000 or less, heterozygote carrier frequency of 1%) characterised predominantly by severe progressive cerebellar degeneration and increased rates of leukaemia and lymphoma (Broeks et al., 2000; Watts et al., 2002). The ataxia telangiectasia mutated (ATM) gene, cloned in 1995, is very large and mutations occur throughout its amino acid coding portions. The ATM gene occupies ~150 kb of 11q22.3-q23.1 (Figure 1).

Figure 1 The ATM gene at 11q22.3-q23.1



DNA damage is one of the most acute threats to cellular homeostasis and life. BRCA1 cannot begin the process of repairing cells when the damage is caused by some types of radiation therapies (Gatei et al., 2000). In this case, BRCA1 relies upon ATM protein to initiate repairs. Therefore, women with a flawed ATM gene are unable to repair damaged DNA. This causes cells with damaged DNA to multiply uncontrollably, thereby becoming cancer cells. ATM (AAB65827), the product of ATM gene, is a member of a family of large proteins found in various organisms that share a COOH-terminal PI3 kinase-like domain (Chenevix-Trench et al., 2002). ATM has serine/threonine protein kinase activity and mediates the activation of multiple signal transduction pathways. Although it has been well established that IR exposure activates the ATM kinase domain, the actual mechanism by which ATM responds to damaged DNA has remained enigmatic until recently. Initial evidences indicated that ATM

activation might involve autophosphorylation. It was found that ATM molecules are inactive in undamaged cells, being held as dimers or high-order multimers. In this configuration, the kinase domain of each molecule is blocked by the FAT domain of the other (Piret et al., 1999). Following DNA damage, each ATM molecule phosphorylates the other on a serine residue at position 1981 within the FAT domain, a phosphorylation that releases two molecules from each other's grip, turning them into fully active monomers. The existence of evolutionary constraint in the FAT domain, especially in the higher primates, underscores the importance of this region for the function of human ATM. At least two missense variants (V2424G and Y2470D) in this region have been reported to have a dominant negative effect on protein function or protein stability, suggesting that the region may be important for protein-protein interactions. However, one variant in the FAT region (C2464R) had normal protein activity, suggesting that not all such changes are deleterious (Stredrick et al., 2006).

The ATM gene (NC_000011.8) is very large and mutations occur throughout its amino acid coding portions. One particular mutation, V2424G, is stable but intrinsically defective as a kinase. It appears that, in cells with these mutant ATM alleles, p53 phosphorylation and stabilisation are reduced, which presumably decreases the effectiveness of the G1/S checkpoint (Khanna et al., 1998). These cells are also defective in DNA damage induced phosphorylation of BRCA1. The mutant ATM allele may exert a dominant negative effect on the wild-type allele for a variety of reasons, including direct competition with wild-type protein for binding to key substrates and regulators, such as the p53 and BRCA1 proteins (Edvardsen et al., 2007). These are biologically significant roles of the ATM gene compel to understand the characteristic features of the gene. Recent reports carried out in non-cancerous mouse BRCA1 or BRCA2 deficient embryonic stem (ES) cells, and hamster BRCA2-deficient cells have demonstrated that the targeted inhibition of Poly (ADP-Ribose) polymerase (PARP-1) kills BRCA mutant cells with high specificity (Buerkle, 2005). PARP, a protein involved in a number of cellular processes mainly involving DNA repair and programmed cell death. PARP activation rescues tumour cells from therapeutic DNA damage induced by chemotherapy agents (Bryant and Helleday, 2006). PARP repairs alkylated bases by recruiting scaffolding proteins (XRCC1), DNA ligase III, and DNA polymerases, and mediates base excision repair (BER). PARP activation in this setting is a basis for tumour resistance to chemotherapy. PARP inhibition, conversely, attenuates tumour resistance to alkylating agents and restores susceptibility of tumours to chemotherapy (Noël et al., 2003).

2 Methodology adopted

Efficient bioinformatics tools and databases were used for the analysis of the target gene (P53). All the programmes were run with default value.

2.1 Sequence analysis

The term 'sequence analysis' in biology implies subjecting a DNA or peptide sequence to sequence alignment, sequence databases, repeated sequence searches, or other bioinformatics methods on a computer. Sequence analysis in molecular biology and bioinformatics is an automated, computer-based examination of characteristically fragments, e.g., of DNA-strand. It basically includes five biologically relevant topics:

- 1 the comparison of sequences in order to find similar sequences (sequence alignment)
- 2 identification of gene-structures, reading frames, distributions of introns and exons and regulatory elements
- 3 prediction of protein structures
- 4 genome mapping
- 5 comparison of homologous sequences for construct a molecular phylogeny.

2.2 *Kyoto Encyclopedia of Genes and Genomes*

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a knowledge base for systematic analysis of gene function in terms of network of gene and molecules. The major component of KEGG is the pathway database that consists of graphical diagrams of biochemical pathways including most of the known metabolic pathways (<http://www.genome.jp/kegg/>).

2.3 *National Center for Biotechnology Information*

National Center for Biotechnology Information (NCBI) provides access to the whole genomes of over 1,500 organisms. Various data retrieval and submission tools may be found on the NCBI website, including text tern searching, sequence similarity searching, taxonomy, and sequence submission. This database was used in retrieving the sequence of the ATM gene (NC_000011.8) (<http://www.ncbi.nlm.nih.gov/>).

2.4 *European Molecular Biology Open Software Suite*

European Molecular Biology Open Software Suite (EMBOSS) is a suite of free software tools for sequence analysis package specially developed for the needs of the molecular biology. The software automatically copes with data in a variety of formats and even allows transparent retrieval of sequence data from the web. EMBOSS provides hundreds of programmes (applications) for,

- a sequence alignment
- b protein motif identification, including domain analysis
- c nucleotide sequence pattern analysis, for example to identify CpG islands or repeats
- d rapid identification of sequence patterns in large scale sequences
- e mutation tools (<http://www.ebi.ac.uk/Tools/emboss/align/>).

2.5 *Mutate sequence beyond all recognition*

Mutate sequence beyond all recognition (MSBAR) programme changes a sequence a lot or little, attempting to emulate various forms of mutation. It can act on the following sizes of sequence:

- a point (single base or residue change)
- b codon (not applicable for proteins)
- c block of sequence (of a specified minimum and maximum random size)
- d for each of the above size of sequence, it can produce the effects of any of the following types of mutation at a randomly chosen position
 - 1 insertion of a randomly generated sequence
 - 2 deletion
 - 3 change (deletion then insertion of a random sequence of the same size)
 - 4 duplication at an adjacent position
 - 5 move region from one position to another (without deletion of the original)
 - 6 any of the above, chosen at random
 - 7 none of the above (<http://www.msbar.org/>).

2.6 *ClustalW2*

ClustalW2 is a general purpose multiple sequence alignment programme for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing cladograms or phylograms (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

2.7 *MODELLER*

MODELLER is used for homology or comparative modelling of protein three-dimensional structures. The user provides an alignment of a sequence to be modelled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modelling by satisfaction of spatial restraints, and can perform many additional tasks, including de novo modelling of loops in protein structures, optimisation of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. (<http://www.salilab.org/modeller/>).

2.8 *Swiss-PdbViewer*

Swiss-PdbViewer (SPDPV) is an application that provides a user friendly interface that allows the user to analyse several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.

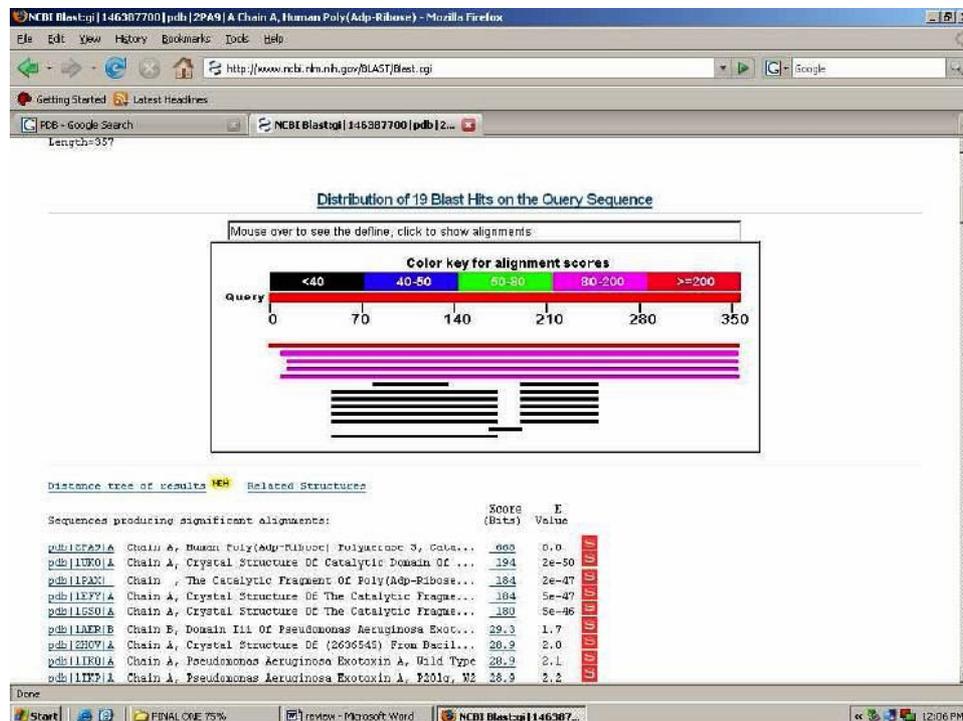
Moreover, SPDPV is tightly linked to Swiss-Model, an automated homology modelling server accessible from ExPASy (<http://spdbv.vital-it.ch/>).

3 Results and discussion

Cells have evolved various sophisticated pathways to sense and overcome DNA damage as a mechanism to preserve the integrity of the genome. Environmental attacks like radiations or toxins, or spontaneous DNA lesions, trigger checkpoint activation and consequent cell cycle arrest leading to DNA repair or apoptosis. Two key proteins that coordinate recognition of DNA damage and signal transduction to p53 are ATM and PARP-1 (Khanna et al., 1998). ATM and PARP-1 participate in distinct forms of DNA repair that partially compensate for each other. PARP-1 and ATM participate in BER and homologous recombination (HR), respectively. It is normally assumed that ATM signals for double strand breaks while PARP-1 participates in signalling from single DNA strand lesions. Here we report that these two proteins form a molecular complex that co-localises in DNA damage foci (D'Amours et al., 1999).

ATM and PARP-1 are two of the most important players in the cell's response to DNA damage. PARP-1 and ATM recognise and bound to both single and double strand DNA breaks in response to different triggers. ATM and PARP-1 form a molecular complex in vivo in undamaged cells. ATM is also modified by PARP-1 during DNA damage. While PARP-1 deficient cells display a defective ATM-kinase activity, PARP inhibition on itself is able to activate ATM-kinase. Inhibition of PARP also induces DNA double strand breaks which were dependent on the presence of ATM. As consequence ATM deficient cells display an increased sensitivity to PARP inhibition (Lavin et al., 2004).

Figure 2 Selection of template using the BLAST programme (see online version for colours)

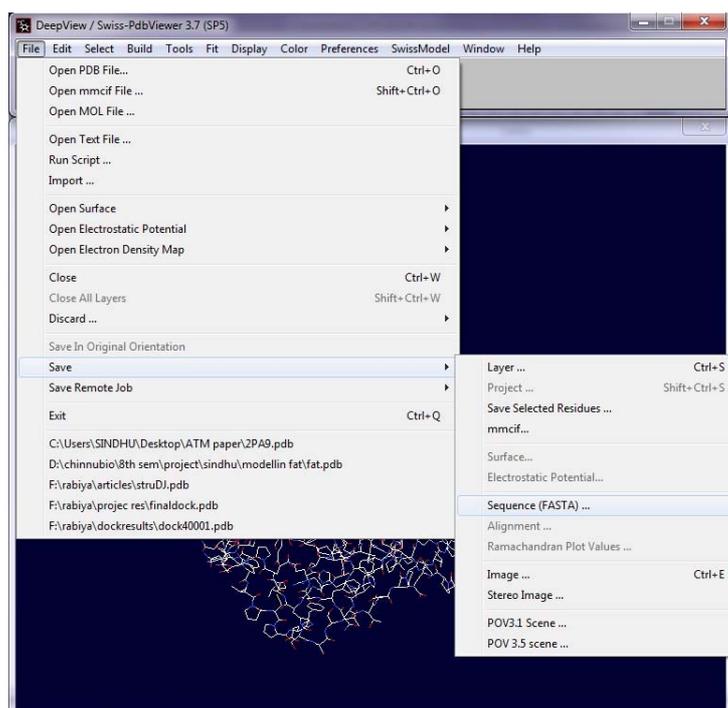


3.1 Homology modelling of PARP-1

3.1.1 Selecting template

Breast cancer was found to have many targets out of which the gene PARP1 was found to play a major role in preventing the proliferation of cells. Hence, PARP1 with the accession number NM_005485 was chosen as the target. It was seen to contain 357 amino acid residues. The target was then blasted using the protein-protein blast programme (<http://www.ncbi.nlm.nih.gov/BLAST/>). A total of 19 hits were obtained and the chosen template (1EFY, Chain A) was found to be less similar to the 30 target protein with a sequence identity of 35% and maximum gap penalty frequency was taken as template (Figure 2). The pdb file of the template was downloaded from the protein data bank which can be accessed at <http://www.rcsb.org/pdb>. The atom file (.atm) was created and the sequence of the template was obtained (Figure 3) by using the export function from the SPDBV tool (<http://expasy.org/spdbv/>). Thus, the sequence for target and template is retrieved.

Figure 3 Retrieving the FASTA file of the template using SPDBV (see online version for colours)



3.1.2 Aligning target with template

The alignment of the target and template sequences was done using the MultAlin (<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>) pair wise alignment tool for proteins with user defined gap penalties. BLOSUM62-12-2 matrix was used to find

the sequence similarity between target and template (Figure 4). The alignment was obtained.

Figure 4 Aligning target and template sequences using MultAlin (see online version for colours)

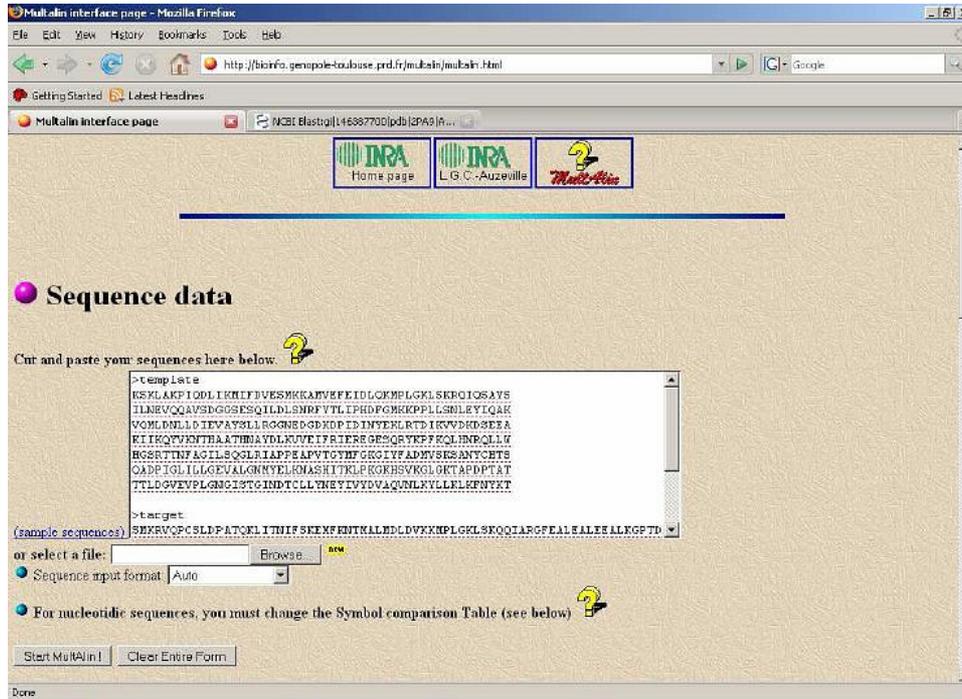


Figure 5 Input file for MULTALIN software (see online version for colours)

```
>template 350 Weight: 1.00
-----KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPLGKLSKRQ
IQSAYSILNEVQQAASDGGSESQILD-LSNRFYTLIPHDFGMKKPPLLSN
LEYIQAKVQMLDNLLDIEVAYSLLRGGNEDGDKD----PIDINYEKLRTD
IKVVDKDSSEAKIHKQYVKNTHAAATHNAYDLKVVVEIFRIEREGESQRYKP
FKQLHNRQLLWHGSRITTFAGILSQGLRIAPPEAPVTGYMFGKGIYFADM
VSKSANYCHTSQ--ADPIGLILLGEVALGNMYELKNASHITKLPK-GKHS
VKGLGKTAPDPTATTTLD----GVEVPLGNGISTG-INDTCLLYNEYIVY
DVAQVNLKYLKLFNYKT

>target 357 Weight: 1.00
SMKRVQPCSLDPATQKLITNIFSKEMFKNTMALMDLDVKKMPLGKLSKQQ
IARGFEALEALEEALGPTDGGQSLEELSSHFYTVIPHNFGHSQPPINS
PELLQAKKDMLLVLADIELAQAQAVSEQEKTVEEVPHPLDRDYQLKCKQ
LQLLDGAPEYKVIQTYLEQT-GSNHRCPTLQ--HIWKVNQEGEEDRFQA
HSKLGNRKLWHGTNMAVVAAILTSGLRIMPH----SDDRVRGKGIYFASE
NSKSAGYVIGMKCGAHHVGYMFLGEVALGREHHINTDNPSLKSPPPGFDS
VIARGHTEPDPTQDTELELDGQQVVVPQGPVPCPEFSSSTFSQSEYLIY
QESQRLRYLLEVH-----
```

3.1.3 Model building

The fourth step in comparative modelling is the model building step for the given target sequence. The programme MODELLER (version 9) was used. MODELLER is used for homology or comparative modelling of protein three-dimensional structures. The user provides an alignment of a sequence to be modelled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modelling by satisfaction of spatial restraints, and can perform many additional tasks, including de novo modelling of loops in protein structures, optimisation of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. Before using the MODELLER programme, the alignment file and the model-default file have to be created. The alignment file was created as follows:

Figure 6 Alignment file for MODELLER programme (see online version for colours)

```
>P1;template
structureX:template:662 :A :1011 :A :ferredoxin:Azotobacter
vinelandii: 1.90: 0.19
-----KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPLGKLSKRQ
IQSAYSILNEVQQAVSDGGSESQILD-LSNRFYTLIPHDFGMKKPPLSN
LEYIQAKVQMLDNLLDIEVAYSLLRGGNEDGDKD---PIDINYEKLRTD
IKVVDKDSEEAKIHKQYVKNTHAATHNAYDLKVVEIFRIEREGESQRYKP
FKQLHNRQLLWHGSRITNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADM
VSKSANYCHTSQ--ADPIGLILLGEVALGNMYELKNASHITKLPK-GKHS
VKGLGKTAPDPTATTTLD---GVEVPLNGISTG-INDTCLLYNEYIVY
DVAQVNLKYLLKLFNYKT*

>P1;target
sequence:target:1 : :357 : :ferredoxin:Peptococcus aerogenes:
2.00:-1.00
SMKRVQPCSLDPATQKLITNIFSKEMFKNTMALMDLDVKKMPLGKLSKQQ
IARGFEALEALEEALKGPTDGGQSLEELSSHFYTVIPHNFGHSQPPPINS
PELLQAKKDMLLVLADELAQALQAVSEQEKTVEEVPHPLDRDYQLLKCCQ
LQLLDGAPEYKVIQTYLEQT-GSNHRCPTLQ--HIWKVNQEGEEDRFQA
HSKLGNRKLLWHGTNMAVVAAILTSGLRIMPH---SGGRVVGKGIYFASE
NSKSAGYVIGMKCGAHHVGYMFLGEVALGREHHINTDNPSLKSPPPGFDS
VIARGHTEPDPTQDTELELDGQQVVVPQGPVPCPEFSSSTFSQSEYLIY
QESQCRLRYLLEVH-----*
```

Care should be taken while creating the alignment file. The template's alignment should precede the alignment of the target. All the file names should match with the respective names in the file. The chains and the length of the template and target have to be mentioned clearly. Once all the above conditions are fulfilled, the 'model-default' file has to be created. This was created as follows:

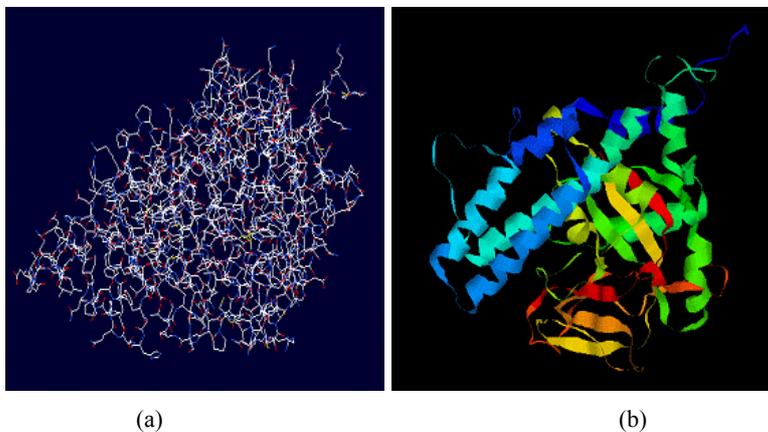
Figure 7 The 'model-default' file that was created (see online version for colours)

```

# Homology modeling by the automodel class
from modeller import * # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this
model in
# directories for input atom files
env.io.atom_files_directory = './../atom_files'
a = automodel(env,
alnfile = 'alignment.ali', # alignment filename
knowns = 'template', # codes of the
templates
sequence = 'target') # code of the target
a.starting_model= 1 # index of the first model
a.ending_model = 1 # index of the last model
# (determines how many models to
calculate)
a.make() # do the actual homology modeling

```

MODELLER was run using the command 'mod9v1 model-default.py'. The structure obtained was as follows. Figure 8(a) is the structure seen using SPDBV and Figure 8(b) is the structure seen in RASMOL.

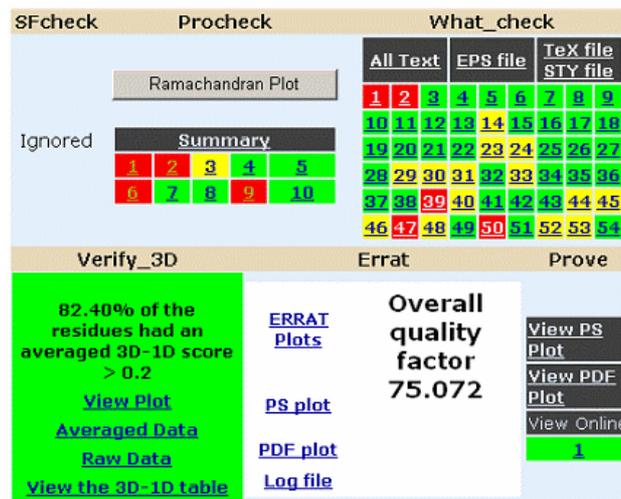
Figure 8 (a) Structure seen using SPDBV (b) structure seen in RASMOL (see online version for colours)

3.1.4 Evaluation and refinement of the model

A major problem in the determination of the 3D structure of the protein concerns the structural models obtained from the interpretation of experimental data. Once the model is built, it is vital to check for two errors: internal error and external error. In internal evaluation, we check if a model satisfies the restraints used in the calculation. It includes the stereo chemical check on the modelled structure, like bond lengths, dihedral angles,

etc. External evaluation relies on the information that was not used in the calculation of the model. Here the quality of the model is assessed using PROCHECK and WHATCHECK. The Structure Analysis Verification Server (SAVS) gives a total quality factor value based on the above programmes. An interaction between ATM and PARP-1 gene was obtained (Figure 9). PARP-1 and ATM were both found to be responsible for DNA repair mechanisms. The signalling cascades that are activated in response to such damage remain to be fully elucidated, but both proteins play critical roles. In response to ionising radiation, the kinase activity of ATM is enhanced, leading to phosphorylation of p53 on serine 15. The activated ATM kinase also phosphorylates and activates the checkpoint kinase Chk2, which in turn, phosphorylates p53 on serine 20. These interactions result in the stabilisation of p53 and its activation as a transcription factor of genes such as MDM2, GADD45 and BAX which is associated with cell cycle arrest, DNA repair, or apoptosis. ATM can also regulate the cell cycle by p53-independent pathways. Both of these multifunctional proteins are also involved in various other signalling pathways. Thus, abnormalities in either ATM or p53 expression could have dramatic consequences for both the control of normal physiological processes and the cellular response to DNA damage.

Figure 9 SAVS server results (see online version for colours)



ATM and PARP-1 were found to form a molecular complex in vivo in undamaged cells. While PARP-1 deficient cells display a defective ATM-kinase activity, PARP inhibition on itself is able to activate ATM-kinase. Inhibition of PARP also induces DNA double strand breaks which were dependent on the presence of ATM. As consequence ATM deficient cells display an increased sensitivity to PARP inhibition.

3.2 ATM gene

For many women who have a strong family history of breast cancer, knowing whether they have inherited gene mutations that predispose them to the disease is the key to their ability to manage their fears of developing cancer. But for more than half the women in this category, no genetic test is available because their family disease is caused by

something other than known BRCA1 or BRCA2 mutations, the primary mutations so far unearthed in severely affected families (Thorstenson et al., 2001). The study concluded that the presence of the mutations dramatically increased the odds of developing breast cancer to 16-fold, as high as that for some BRCA1 or BRCA2 mutations. Only women with a family history appeared to carry these mutations. New study reveals that BRCA1 may rely on the ATM protein to repair some types of cell damage. When DNA damage occurs in cells, BRCA1 helps the body repairs those cells – a process that may prevent tumour development (Szabo et al., 2004). However, BRCA1 cannot begin the process of repairing cells when the damage is caused by some types of radiation therapies. In this case, BRCA1 must rely on the ATM protein to initiate the repairs. ATM is important in recognising when chromosomes are broken. When functioning normally, ATM helps control cell division. Mutant alleles pack a powerful punch; not only are they faulty themselves, but they also corrupt the function of the normal ATM protein. Normal ATM protein activates members of the DNA repair crew – the BRCA1, BRCA2, and p53 gene products – by adding phosphate groups to them (Khanna et al., 1998). But in the presence of the mutant ATM, no phosphate is added. However, inherited mutations of the ATM gene from both the father and the mother have been associated with a loss of brain cells that causes a lack of muscle control and increased cancer rates. Inheriting one mutation of the ATM gene has also been shown to increase the risk for breast cancer in previous research, although the risk was thought to be very small and therefore not associated with a high frequency of breast cancer in some families (Angèle et al., 2000).

3.2.1 Sequence retrieval

The sequence of the ATM gene having an accession number AAB65827 was retrieved from the NCBI database (Figure 10). The length was found to be 3056 amino acids.

Figure 10 Retrieval of ATM sequence from NCBI (see online version for colours)

The screenshot shows the NCBI Protein database search interface. The search bar contains 'Protein' and the results page displays the following information for the ATM [Homo sapiens] entry (Accession: AAB65827):

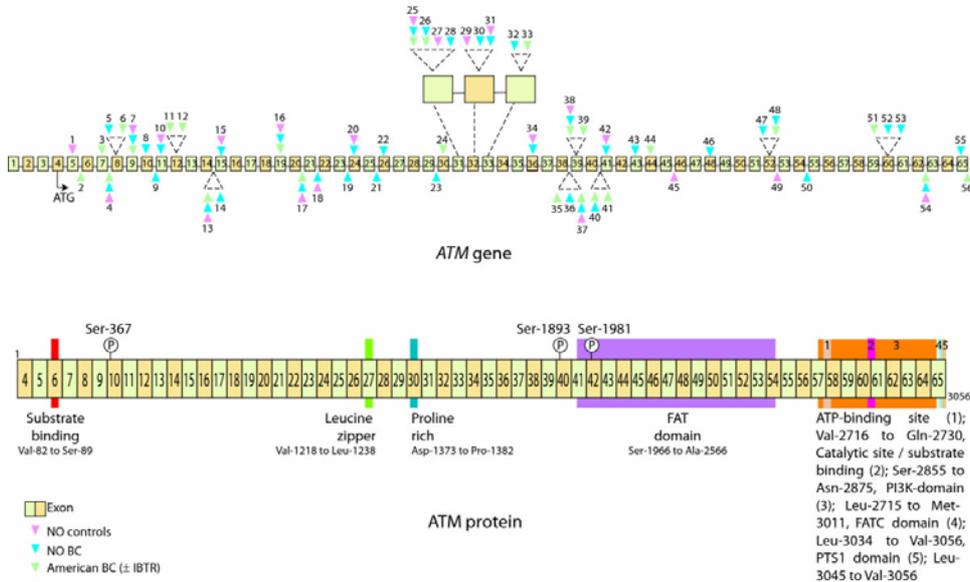
- LOCUS:** AAB65827 3056 aa linear PRI 06-AUG-1997
- DEFINITION:** ATM [Homo sapiens].
- ACCESSION:** AAB65827
- VERSION:** AAB65827.1 GI:2304971
- DBSOURCE:** locus HS082828 accession [U22828.1](#)
- KEYWORDS:** .
- SOURCE:** Homo sapiens (human)
- ORGANISM:** [Homo_sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
- REFERENCE 1 (residues 1 to 3056):**
AUTHORS: Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y. and Rosenthal, A.
TITLE: Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene
JOURNAL: Genome Res. 7 (6), 592-605 (1997)
PubMed: [919932](#)
- REFERENCE 2 (residues 1 to 3056):**
AUTHORS: Platzer, M., Rotman, G., Bauer, D., Savitsky, K., Shiloh, Y. and Rosenthal, A.
TITLE: Direct Submission
JOURNAL: Submitted (20-DEC-1996) Genomeanal, Inst. Molec. Biotechnol.,

3.2.2 ATM domain

The ATM gene codes for a protein with 3056 amino acids and a molecular weight of ~350 kDa which have been found to exist both in mono-meric (active) and di-meric (inactive) state. The protein contains several important domains such as

- 1 the C-terminal protein kinase domain (PI3K-domain)
- 2 the substrate binding domain in the N-terminal of the protein necessary for activation of p53 in response to DNA damage
- 3 the FAT domain – common for the PI3K-like family members FRAP, ATM and TRAPP
- 4 a proline rich region shown to bind c-Abl
- 5 an incomplete leucine zipper (Figure 11).

Figure 11 Schematic illustration of the ATM gene (see online version for colours)

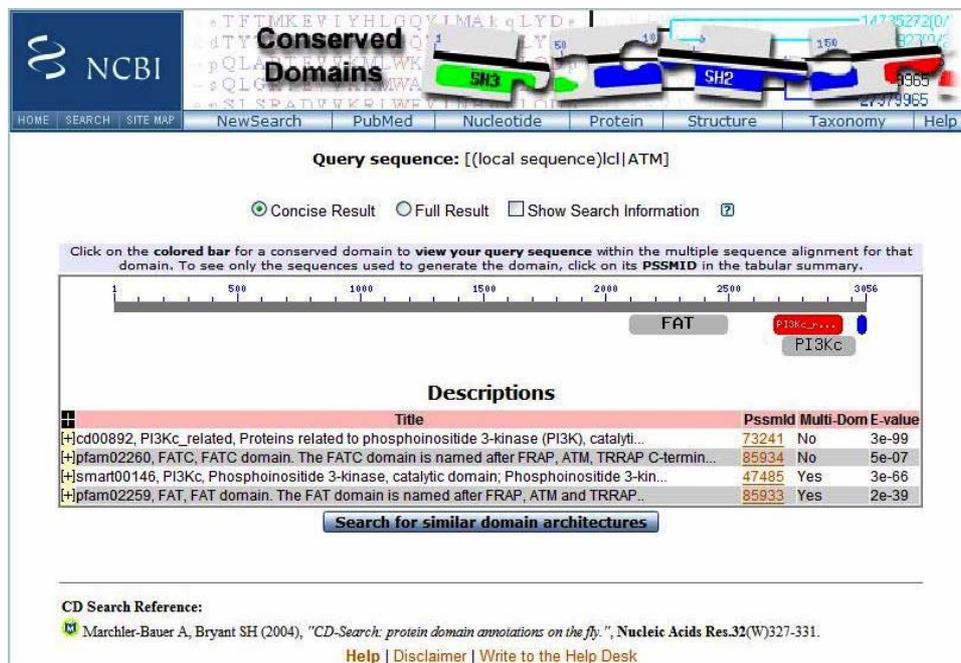


Notes: The distribution of the variations detected in the studied materials along the gene is shown in the upper panel with exonic variants indicated on top of the gene and intronic below the gene, illustrated by coloured triangles. Above is given an illustration of the protein with important areas such as substrate binding domains, Leucine zipper, ATP-binding domains, FAT domain and PI3K domain together with exonic information. (The size of the exons and the distance between them are not indicative of the sizes/distances in the gene/protein).

The putative domains were detected using the conserved domain search in NCBI (Figure 12). The result obtained showed the presence of four domains – a FAT domain, a FATC domain, a PI3K domain and a PI3Kc_related domain. It was seen that the FAT domain was 395 residues long and that most of the mutations occur with the FAT domain. Initial evidences indicated that ATM activation might involve autophosphorylation. It was found that ATM molecules are inactive in undamaged cells,

being held as di-mers or high-order multi-mers. In this configuration, the kinase domain of each molecule is blocked by the FAT domain of the other. Following DNA damage, each ATM molecule phosphorylates the other on a serine residue at position 1981 within the FAT domain, a phosphorylation that releases two molecules from each other's grip, turning them into fully active monomers. The existence of evolutionary constraint in the FAT domain, especially in the higher primates, underscores the importance of this region for the function of human ATM. At least two missense variants (V2424G and Y2470D) in this region have been reported to have a dominant negative effect on protein function or protein stability, suggesting that the region may be important for protein-protein interactions.

Figure 12 Conserved domain search for ATM using NCBI (see online version for colours)



NCBI Conserved Domains

Query sequence: [(local sequence)cl|ATM]

Concise Result Full Result Show Search Information

Click on the colored bar for a conserved domain to view your query sequence within the multiple sequence alignment for that domain. To see only the sequences used to generate the domain, click on its PSSMID in the tabular summary.

1 500 1000 1500 2000 2500 3056

FAT PI3Kc

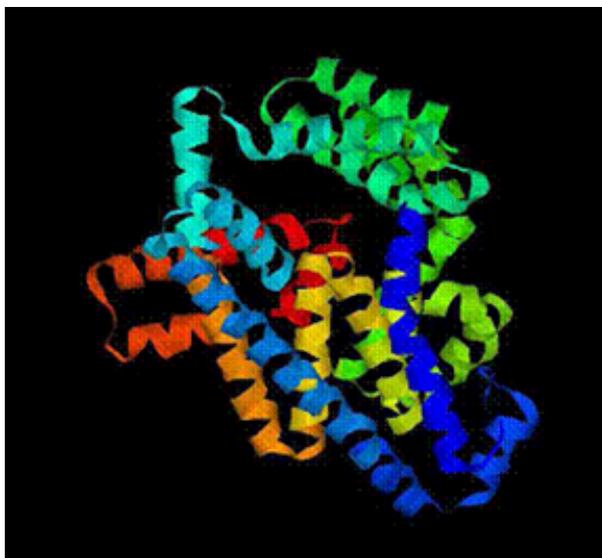
Descriptions

Title	Pssmid	Multi-Dom	E-value
[+]cd00892, PI3Kc-related, Proteins related to phosphoinositide 3-kinase (PI3K), catalyti...	73241	No	3e-99
[+]pfam02260, FATC, FATC domain. The FATC domain is named after FRAP, ATM, TRRAP C-termin...	85934	No	5e-07
[+]smart00146, PI3Kc, Phosphoinositide 3-kinase, catalytic domain; Phosphoinositide 3-kin...	47485	Yes	3e-66
[+]pfam02259, FAT, FAT domain. The FAT domain is named after FRAP, ATM and TRRAP...	85933	Yes	2e-39

[Search for similar domain architectures](#)

CD Search Reference:
 Marchler-Bauer A, Bryant SH (2004), "CD-Search: protein domain annotations on the fly.", *Nucleic Acids Res.*32(W)327-331.
[Help](#) | [Disclaimer](#) | [Write to the Help Desk](#)

- Conserved domains (4)
 - smart00146
 - Location: 2715–3011
 - Blast score: 639
 - PI3Kc; phosphoinositide 3-kinase, catalytic domain; phosphoinositide 3-kinase isoforms participate in a variety of processes, including cell motility, the Ras pathway, vesicle trafficking and secretion, and apoptosis.

Figure 14 Structure of the FAT domain of ATM (see online version for colours)

3.3.2 Mutations in the FAT domain

The T7271G (Val2424Gly) mutation was the only ATM mutation found in the AT family and is, therefore, likely to be pathogenic. This mutation leads to incorrect splicing of exon 11 and exon skipping, resulting in a frame shift starting at codon 355 and subsequent truncation of the protein at amino acid 419. In vivo analyses of cell lines established from individuals heterozygous for the T7271G and IVS10–6T→G mutations suggest that both of these ATM mutations act in a dominant negative manner so that the wild-type enzyme is unable to function normally in the presence of the mutant protein (Szabo et al., 2004). Missense mutations in ATM have been associated with an elevated risk of breast cancer. Complete mutation analysis of the ATM gene revealed a missense mutation at position 2424 where valine is substituted by glycine. The V2424G missense mutation caused a helix to coil conformational change, thus revealing an increase in the number of chromosomal aberrations. Mutation analysis was carried out using the MSBAR programme (Figure 16) available in EMBOSS. The V2424G missense mutational changes shown on Figure 17 and Figure 18, and its sequence mentioned as follows:

Figure 15 Results of the mutation analysis done using MSBAR program in EMBOSS (see online version for colours)

```

IENYMKSSSEFENKQALLKRAKEEVGLLREHKIQTNRYTVKVVQRELELDELALRALKEDRK 2460
IENYMKSSSEFENKQALLKRAKEEGGLLREHKIQTNRYTVKVVQRELELDELALRALKEDRK 2460

```

Figure 16 Mutation Analysis using MSBAR (see online version for colours)

EMBOSS

pepwindowall

PROTEIN MOTIFS

- antigenic
- digest
- epesfind
- fuzzpro
- fuzztran
- helixturnhelix
- odcomp
- patmatdb
- patmatnotifs
- pepcoll
- preg
- pscan
- sigcleave

PROTEIN MUTATION

- msbar
- shuffleseq

PROTEIN PROFILES

- profit
- prophecy
- prophet
- aindexextract
- domainer
- hetparse
- pdparse
- pdbtosp
- scopnr

OUTPUT FILE: outseq [RIGHT CLICK TO SAVE]

```
>AAB65027.1 ATM [Homo sapiens]
MSLVLNDLLICCRQLEHDRATERKKEVEKFKRLIRDPEIKHLDRHSDSKQKYLNDVAV
FRFLQKYIQKETECLRIAKPNVSASTQASRQKMQEISSLVKYFIKCANRRAPRLKQCEL
LNYIMDTVKDSSNGAIYVADCSNILLKDIILSVRFWCEISQQQWLELFVSYFRLYLKPSQ
DVHREVLVARIHAVITKGCSSQTDGLNSKFLDFFSKAIQCARQEKSSSGLNHLAALTIPL
KTLAVNFRIRVCELGDEILPTLLYIWTQHRLENDLSKEVIELFLQIYIHPKGAKTQEK
GAYESTKWRSLLYNLYDLLVNEISHIGSRGKYSSGFENLAVKENLIELMADIHQVFNED
TRSLIISQSYTTTQRESSDYSVPCKRKKIELGWEVVKDHLQKSNDFDLVPWQLIATQLI
SKYPASLPNCELSPLLMILSOLLPQQRHGERTPYVLRCLTEVALCQDKRNLSSQKSDL
LKLWNKIWCITFRGISSEQIQANFGLLGAIIQGSLEVEDREFWKLFTGSACRPPCPAVC
CLTLALTTISVPTVKMGIEQNMCEVNRSFSLKESIMRWLLFYQLEGDLENSTEVPIIH
SNFPHLVLEKILVSLTMKNCKAAMNFFQSVPECEHHQKDKELSPSEVEELFLQTTDFRM
DFLTIIVRECGIERHQSSIGFSVHQNLKESLDRCLLGLSEQLLNNYSSEITNSSETLVRCR
LLVQVLOCYCYMOVIAEEEAAYKSELFQAKSLMQCAGESITLFFKNKTNEEPRIGSLRNM
QLCTRCLSNCTKSPNKIASOFFLRLTSLKLMNDIADICKSLASFIKKPFDRQEVESMED
DTNGNLMEVEDQSSMNLFDYFDSVSDANEPGESQSTIGAINPLAEVLSKQDLLFLDM
LKFCLCVTTAQTNTVSFRAADIRRKLMLIDSSTLEPTKSLHLHMYLMLKELPGEYYP
LPMEDVLELLKPLSNVCSLYRRDQDVCKTILNHVHLHVKNLQGSNMDSENTDAQQGFLT
VIGAFWHLTKERKVI FSVRMALVNCLKTLLEADPYSKWAILNVMGKDFPVNEVFTQFLAD
NHHQVRMLAAESINRLFQDTKGDSSRLKALPLKQQTAFENAYLKAQEGMREMSHAEN
PETLDEIYNRKSIVLLTLIAVVLSCSPICEKQALFALCKSVKENGLEPHLVKVKLEKVS
ETFGYRREDFMASHLDYLVLEWNLQDTEYNLSSFPFILLNYTINIEDFYRSCYKVIIPHLV
IRSHFDEVKSIANQIQEDWKSLLTDCFPKILVNIILPYFAYEGTRDSGMAQQRETATKVYD
MLKSENLLGRQIDHLFISNLPETVVELLMTLHEPANSSASQSTDLCDFFSGDLLPAPNPH
FPSHVIKATFAYISNCHTKLKLKSLILEILSGSPDSYQKILLAICEQAAETNNVYKHHRIK
IVHIEVSLIEDIKSGLGGWAEVLRVIVYTIHVINDRPSCTMDVSLRSEFLCCDLISQ
```

Figure 17 Structure of the residue before mutation showing valine 2424 (see online version for colours)

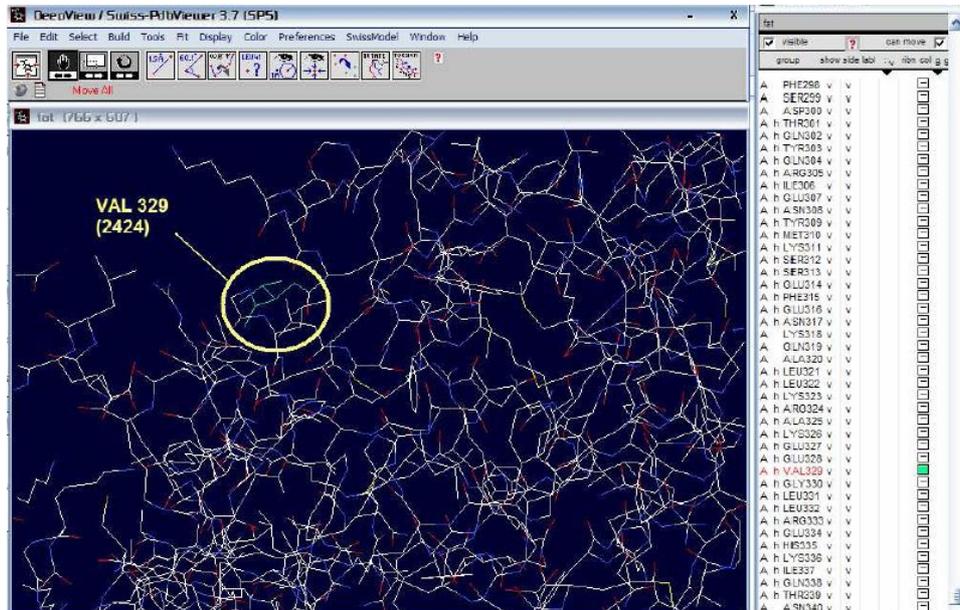


Figure 18 Structure of the residue after mutation showing glycine 2424 (see online version for colours)

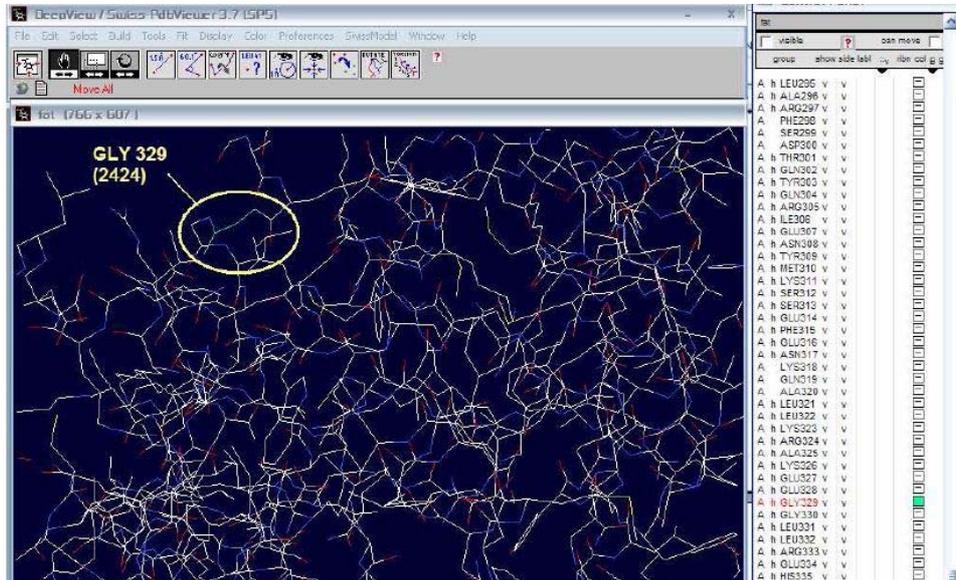
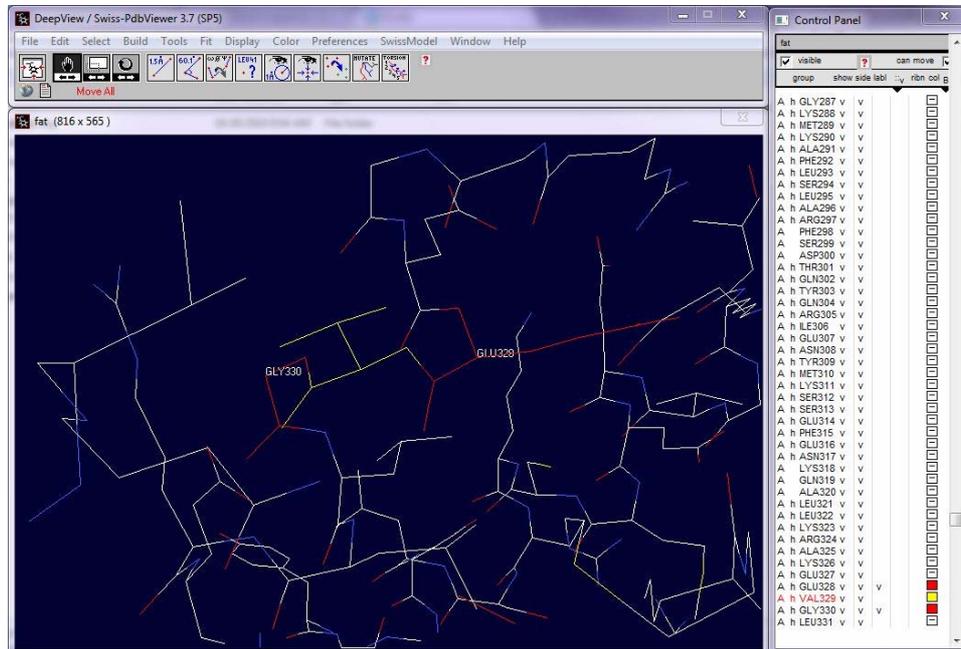


Figure 19 Representation of the hydrogen bond interactions between valine (in yellow) and its neighbouring residues Glu 328 and Gly 330 (see online version for colours)



The missense mutation resulted in an increase in the number of chromosomal aberrations. The hydrogen bond interactions were computed to find out any changes in the interactions of the residue at position 2424 and its neighbouring residues. The distance considered was 2 Å. Figure 19 portrays the various hydrogen bond interactions between the residue under consideration and its neighbouring residues. No interactions are seen with the CH₃-CH-CH₃ atoms in the valine residue. Since there is no interaction between these atoms, its mutation to glycine would also not depict any interactions. This is the case with residues present within 4 Å. Although no change is detected structure wise or interaction wise, the V2424G missense mutation was found to have an important functional aspect of increasing the chromosomal aberrations.

4 Conclusions

Certain ATM gene mutations, missense and truncating, are associated with a sufficiently high risk of breast cancer. Loss of ATM decreases the genetic stability of malignant cells but, at the same time, affects the sensitivity of malignant cells to radiation and thereby could provide a therapeutic gain. New drugs could be screened for their ability to exploit the differences between normal and malignant cells in cell cycle checkpoints and DNA repair. The above project will allow direct evaluation of breast cancer related to mutations in the ATM gene. This will facilitate the design of appropriate preventive and therapeutic strategies.

Acknowledgements

The authors gratefully acknowledge the help of Dr. M. Anand, iPower Solutions India Ltd., Chennai-88, India. This work was fully supported by a research grant from Department of Bioinformatics, School of Chemical and Biotechnology, Shanmugha Arts Science Technology & Research Academy (SASTRA University), Tanjore-613402, India.

References

- Angèle, S., Treilleux, I., Taniere, P., Martel-Planche, G., Vuillaume, M., Bailly, C., Brémond, A., Montesano, R. and Hall, J. (2000) 'Abnormal expression of the ATM and TP53 genes in sporadic breast carcinomas', *Clinical Cancer Research*, Vol. 6, pp.3536–3544.
- Brooks, A., Urbanus, J.H., Floore, A.N., Dahler, E.C., Klijn, J.G., Rutgers, E.J., Devilee, P., Russell, N.S., van Leeuwen, F.E. and van 't Veer, L.J. (2000) 'ATM-heterozygous germline mutations contribute to breast cancer-susceptibility', *Am. J. Hum. Genet*, Vol. 66, pp.494–500.
- Bryant, H.E. and Helleday, T. (2006) 'Inhibition of poly (ADP-ribose) polymerase activates ATM which is required for subsequent homologous recombination repair', *Nucleic Acids Research*, Vol. 34, pp.1685–1691.
- Buerkle, A. (2005) 'Poly(ADP-ribose). The most elaborate metabolite of NAD⁺', *FEBS Journal*, Vol. 272, pp.4576–4589.

- Chenevix-Trench, G., Spurdle, A.B., Gatei, M., Kelly, H., Marsh, A., Chen, X., Donn, K., Cummings, M., Nyholt, D., Jenkins, M.A., Scott, C., Pupo, G.M., Dork, T., Bendix, R., Kirk, J., Tucker, K., McCredie, M.R., Hopper, J.L., Sambrook, J., Mann, G.J. and Khanna, K.K. (2002) 'Dominant negative ATM mutations in breast cancer families', *Journal of the National Cancer Institute*, Vol. 94, pp.205–215.
- D'Amours, D., Desnoyers, S., D'Silva, I. and Poirier, G.G. (1999) 'Poly(ADP-ribose) action reactions in the regulation of nuclear functions', *Biochem Journal*, Vol. 342, pp.249–268.
- Edvardsen, H., Tefre, T., Jansen, L., Vu, P., Haffty, B.G., Fossa, S.D., Kristensen, V.N. and Borresen-Dale, A-L. (2007) 'Linkage disequilibrium pattern of the ATM gene in breast cancer patients and controls; association of SNPs and haplotypes to radio-sensitivity and post-lumpectomy local recurrence', *Radiation Oncology*, Vol. 2, p.25.
- Gatei, M., Scott, S.P., Filippovitch, I., Soronika, N., Lavin, M.F., Weber, B. and Khanna, K.K. (2000) 'Role for ATM in DNA damage-induced phosphorylation of BRCA1', *Cancer Res.*, Vol. 60, pp.3299–3304.
- Khanna, K.K., Keating, K.E., Kozlov, S., Scott, S., Gatei, M., Hobson, K. and Taya, Y., Gabrielli, B., Chan, D., Lees-Miller, S.P. and Lavin, M.F. (1998) 'ATM associates with and phosphorylates p53: mapping the region of interaction', *Nat Genet* Vol. 20, pp.398–400.
- Lavin, M.F., Scott, S., Gueven, N., Kozlov, S., Peng, C. and Chen, P. (2004) 'Functional consequences of sequence alterations in the ATM gene', *DNA Repair (Amst)*, Vol. 3, pp.1197–1205.
- Noël, G., Giocanti, N., Fernet, M., Megnin-Chanet, F. and Favaudon, V. (2003) 'Poly(ADP-ribose) polymerase (PARP-1) is not involved in DNA double-strand break recovery', *BMC Cell Biology*, Vol. 4, pp.1471–2121.
- Piret, B., Schoonbroodt, S. and Piette, J. (1999) 'The ATM protein is required for sustained activation of NF- κ B following DNA damage', *Oncogene*, Vol. 18, pp.2261–2271.
- Stredrick, D.L., Garcia-Closas, M., Pineda, M.A., Bhatti, P., Alexander, B.H., Doody, M.M., Lissowska, J., Peplonska, B., Brinton, L.A., Chanock, S.J., Struwing, J.P. and Sigurdson, A.J. (2006) 'The ATM missense mutation p.Ser49Cys (c.146C4G) and the risk of breast cancer', *Human Mutation*, Vol. 27, pp.538–544.
- Szabo, C.I., Schutte, M., Broeks, A. and Houwing-Duistermaat, J.J. (2004) 'Are ATM mutations 7271T3G and IVS10–6T3G really high-risk breast cancer-susceptibility alleles?', *Cancer Research*, Vol. 64, pp.840–843.
- Thorstenson, Y.R., Shen, P., Tusher, V.G., Wayne, T.L., Davis, R.W., Chu, G. and Oefner, P.J. (2001) 'Global analysis of ATM polymorphism reveals significant functional constraint', *Am. J. Hum. Genet*, Vol. 69, pp.396–412.
- Watts, J.A., Morley, M., Burdick, J.T., Fiori, J.L., Ewens, W.J., Spielman, R.S. and Cheung, V.G. (2002) 'Gene expression phenotype in heterozygous carriers of ataxia telangiectasia', *Am. J. Hum. Genet.*, Vol. 71, pp.791–800.