

Homology Modeling and Structure Prediction of Bloom syndrome Causing Protein Isolated from *Liparis tanakae*

Santosh Kumar^{1*}, Minisha Bagaria², Namrata Jha³, O.P. Verma⁴, Nem Kumar Jain⁵

Asst. Professor, Department of Life Sciences, ITM University, Gwalior, M.P., India¹

Student, Department of Life Sciences, ITM University, Gwalior, M.P., India^{2,3}

Associate Professor, Department of Molecular and Cellular Engineering, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India⁴

Asst. Professor, School of Pharmacy, ITM University, Gwalior, M.P., India⁵

ABSTRACT: Bloom syndrome is a rare human disorder which is an autosomal recessive disease which belongs to the group of chromosomal breakage syndrome. Firstly, Dr. David Bloom, a New York dermatologist described this disease in 1954 in three children as congenital telangiectatic erythema resembling lupus erythematosus in dwarfs. This can be examined by genetic instability, which includes high number of sister chromatid exchanges. The predominant and clinical feature of bloom syndrome is pre and postnatal growth retardation. Other clinical features of bloom syndrome are narrow facies, nasal prominence, facial sun sensitive, immunodeficiency, respiratory tract and gastrointestinal tract infection. There is 10 folds increase in the rate of sister chromatid exchanges as compared with normal cells. Bloom syndrome can be characterized by prenatal and postnatal growth. This is caused by the loss of the function of blm genes which codes for RecQ helicase, which causes bloom syndrome. This is a prototypical chromosomal instability syndrome and somatic mutations that occur as a result of that instability are responsible for increased cancer risk also.

Physiochemical characterization of protein isolated from *Liparis tanakae* responsible for Bloom syndrome disease with 429 amino acids was done to interpret properties like pI, EC, AI, GRAVY and instability index. Homology modeling was performed to generate good quality models. The assessment of generated three dimensional structure against structure verification tools PROCHECK showed that model generated by Geno3D was acceptable as compared to SWISS MODEL. The predicted model can be used in structure based drug designing and vaccine development

KEYWORDS: Bloom syndrome, RecQ helicase, homology modeling, PROCHECK, GENO3D

I. INTRODUCTION

Bloom syndrome results due to the formation of the mutant gene which is the BLM gene, which has been seen in the family of RecQ DNA helicase. Bloom Syndrome patients are diagnosed with the extra chromatid sisters with an extraordinary level from its normal level, which also shows a mark of genomic instability in the patient's body [1]. RecQ DNA helicase can be used along with the use of ATP molecules in the presence of the co-factor Mg²⁺; it helps to unwind the different type of DNA structures, it works in the 3' → 5' direction. Human cells contain five different types of helicases, encoded by the BLM, WRN, RECQL/RECQ1, RECQL4 and RECQL5 genes. When the person is diagnosed in the deficiency of BLM, WRN, and RECQL4 proteins that results in the autosomal recessive disorder in that particular human being [2].

In the recent studies it has been demonstrated that decrease in the level of BLM shows the survival response to that of chemotherapeutic agents working in the place of the deficient proteins, providing a particular amount of increase in the level of the developing BLM which acts as a biomarker [3]. Bloom syndrome helicase (BLM), a member of the conserved RecQ helicase family, has been implicated in anti-crossover activity in mitotic cells. Bi-allelic mutations in BLM give rise to Bloom syndrome, a recessive human genetic disorder characterized by short stature, sterility, and cancer pre-disposition [4]. In vitro studies have demonstrated that BLM can disassemble D-loops and, with its partners topoisomerase 3 α , RMI1,

and RMI2 (the BTR complex), catalyze dHJ dissolution [5]. In vivo studies in *Drosophila* showed that in the absence of Blm, mitotic SDSA is severely diminished and mitotic crossovers are increased [6]. Bloom syndrome is an inherited autosomal recessive genetic trait which is a genetic disorder, describing features are short stature, increased skin sensitivity to ultraviolet rays from the sun (photosensitivity) causing formation of some brown and grey spots on the patient's body, multiple small dilated blood vessels (telangiectasia) over the nose and cheeks resembling a butterfly in shape, weak immune system with increased susceptibility to infections, and they are very much vulnerable to cancer like lymphoma and gastro intestinal cancer. They are also diagnosed with a distinctive, narrow, small head and face, shorter and underweight, than a normal human body proportions. Bloom syndrome is caused due to genetic condition in which chromosomes are broken down and DNA is repaired abnormally, which causes high risk for cancer [7].

Characteristic feature of bloom syndrome have a characteristically high-pitched voice, dental abnormalities, prominent ears, cysts at the base of the spine, or extra fingers. Occasionally, other abnormalities of the eyes, ears, hands, or feet may also be present. In males, sperm production ceased which causes sterility; and in females, menstruation occurs abnormally at early age in bloom syndrome. Immune system becomes delicate which results in middle ear infection or sometimes pneumonia [8].

A recent study demonstrated that depletion of BLM undermines cellular survival in response to chemotherapeutic agents; providing a rationale for developing BLM as a biomarker for chemotherapeutic responsiveness [9]. The chromosome abnormalities observed in BS cells pointed to a defect in some fundamental process of DNA metabolism necessary for genomic stability. Retarded replication-fork elongation [10], abnormal replicational intermediates and the absence of defects in known DNA-repair systems [11].

II. MATERIALS AND METHODS

1. Retrieval of target sequence- The amino acid sequence of the protein causing Bloom syndrome isolated from *liparis tanakae* of 429 amino acids was obtained from the sequence database of NCBI (http://www.ncbi.nlm.nih.gov/protein/NP_112025.1). It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank [12]. Hence the present work for developing the 3D model of the protein from *Liparis tanakae* was undertaken.

2. Physico-chemical characterization- The values of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient [13], instability index [14], aliphatic index [15] and grand average hydropathicity (GRAVY) [16] were computed using the ExPASy's ProtParam server [17] (<http://us.expasy.org/tools/protparam.html>).

3. Secondary structure prediction- Secondary structure has been predicted using PHYRE2 software (<http://www.sbg.bio.ic.ac.uk/phyre2>) where the FASTA format of the sequence was given as input. SOPMA [12] was employed for calculating the secondary structural features of the protein sequence considered for this study.

4. Model building and quality assessment- The modeling of the three dimensional structure of the protein was done using two homology modeling programs, Geno3D [18] and SWISS MODEL [19]. The overall stereo chemical property of the protein was assessed by Ramchandran plot analysis [20]. The evaluation of structure models obtained from the two software tools was performed by using PROCHECK [21]. The models were further checked with WHAT IF [22] for standard bond length and bond angle determination.

III. RESULTS AND DISCUSSION

The physicochemical parameters computed using ExPASy's ProtParam tool for the protein causing Bloom syndrome had the molecular weight 48048.03D theoretical pI 9.05, number of negatively charged residues (Asp+Glu) equal to 50, extinction coefficient 40840, instability index 39.63, aliphatic index 63.03 and grand average of hydropathicity (GRAVY) -0.657 which characterizes the physico chemical properties of the protein represented in Table 1. A protein having instability index smaller than 40 is predicted as stable, on the other hand a value above 40 predicts that the protein may be unstable [23]. Instability index of 39.63 indicates the stable nature of the protein. The high extinction coefficient (40840) indicates presence of high concentration of Cystine, Tryptophan and Tyrosine. The aliphatic index is considered as a positive factor for the increase of thermal stability. High aliphatic index (63.03) of the protein suggests that the protein may be stable for a wide temperature range. The Grand Average hydropathicity (GRAVY) value is low (-0.657) which indicates the possibility of better interaction with water.

S.No.	Parameters	Values
1.	Number of amino acids	429
2.	Molecular weight	48048.03
3.	Theoretical Pi	9.05
4.	Total number of negatively charged residues(Asp+Glu)	50
5.	Total number of negatively charged residues(Arg+Lys)	60
6.	Extinction coefficient	40840
7.	Extinction coefficient*	40340
8.	Instability index	39.63
9.	Aliphatic index	63.03
10.	Grand average of hydropathicity	-0.657

Table 1-Physicochemical parameters computed using Expasy's ProtParam tool

The secondary structure of Bloom syndrome protein was predicted by two softwares namely SOPMA (Self Optimized Prediction Method with Alignment) and Phyre 2. SOPMA predicts 69.5% of amino acids correctly to describe secondary structure prediction. Table 2 represents the results of SOPMA. These results show higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns) (Fig. 6). default parameters (Window width: 17, similarity threshold: 8 and number of states: 4) were taken by SOPMA for secondary structure prediction. Secondary structure and disorder prediction was made using phyre 2 which is shown in figure 7.

S.No.	Parameters	Value
1.	Alpha helix	40.79%
2.	3 ₁₀ helix	0.00%
3.	Pi helix	0.00%
4.	Beta bridge	0.00%
5.	Extended strand	8.62%
6.	Beta turn	3.96%
7.	Bend region	0.00%
8.	Random coil	46.62%
9.	Ambiguous state	0.00%
10.	Other state	0.00%

Table 2- Calculated secondary structure elements by SOPMA

Three dimensional structures of proteins are predicted due to unavailability of such data. There is no experimental structure found for the protein considered. The three dimensional structure prediction of the protein was done using two programs Geno3D and Swiss Model. The results obtained from these programs were compared in table 3. This comparison suggests that the model generated by Swiss model (Fig. 1) is much better than the model generated by Geno3D due to presence of maximum residues (93.07%) in the most favoured region. Finally chosen model was visualized by Rasmol. The evaluation of predicted structure generated by SWISS MODEL for the stereochemical quality was done using Ramchandran plot (Fig. 2-Fig. 5), calculations done with the PROCHECK. The 88.3% of residues were found in the core right handed alpha helices(A), beta sheets (B)and left-handed alpha helix(L) region.10.7% of residues were found in the allowed right- handed alpha helix(a), beta sheets(b) and left -handed alpha helices regions. The 0.7% of the residues were found in the generously allowed alpha helices (~a), beta sheets (~b), left handed alpha helices (~l) and epsilon (~p) regions. The 0.3% of the residues was found to be localized at the disallowed regions. Above results indicate to a good quality of predicted model. The analysis of modeled structure of protein of Bloom syndrome by structure verification server WHATIF revealed such RMS Z-score (0.965) which was nearly meeting the required score for a high quality model.

Server	Parameters	Value (%)
Geno 3D	Residues in the most Favoured Region	73.39%
	Residues in additionally allowed region	20.54
	Residues in generously allowed region	4.19%
	Residues in disallowed region	1.99
Swiss Model	Residues in the most Favoured Region	93.07
	Residues in additionally allowed region	5.32
	Residues in generously allowed region	1.53
	Residues in disallowed region	0.08

Table 3 - Ramachandran plot calculation and Comparative analysis of the models from Geno3D and Swiss-model computed with the PROCHECK program

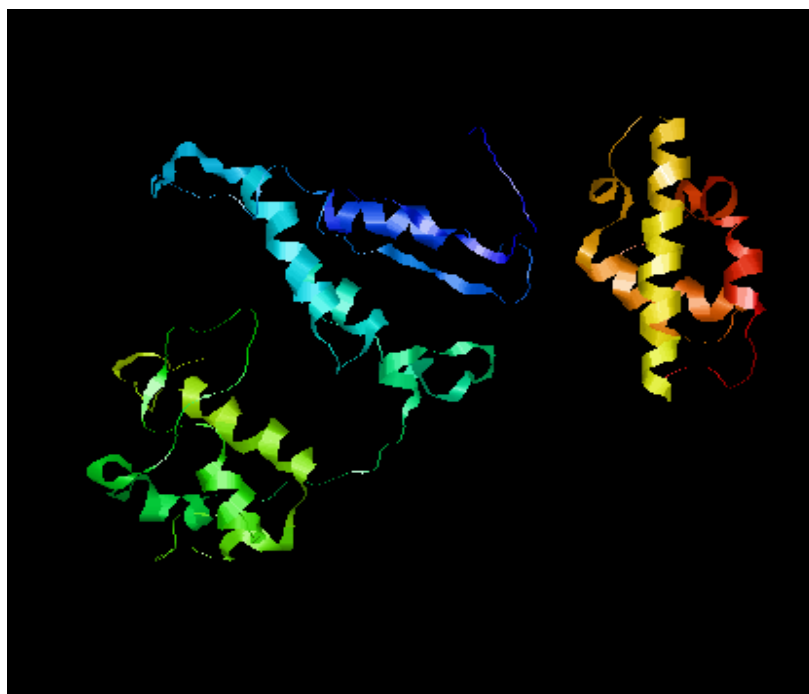


Fig 1: Structure generated by SWISS MODEL software

Ramachandran Plot

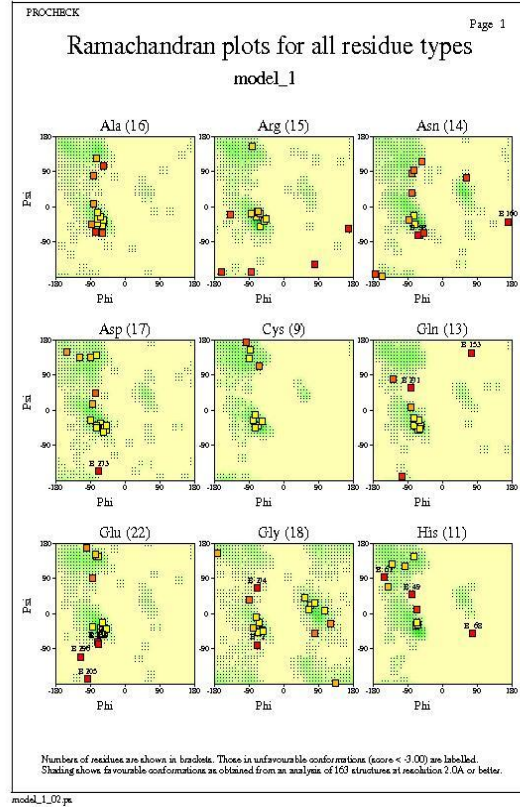
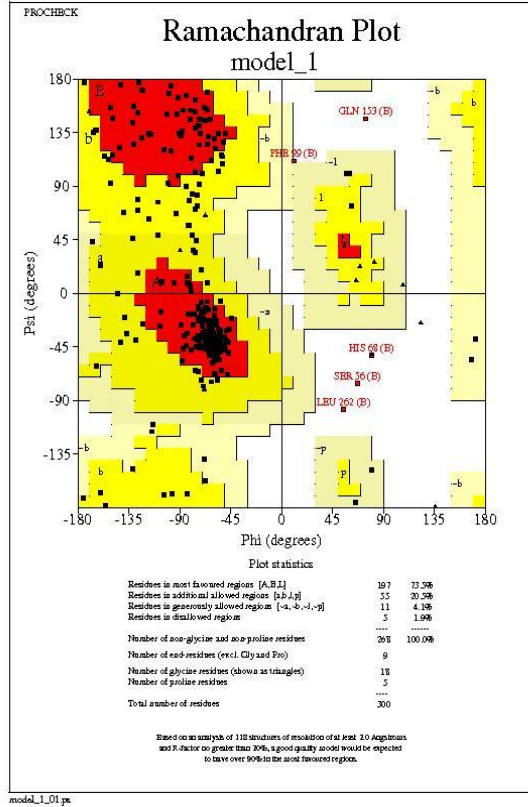


Fig 2: Ramachandran plot for all residues

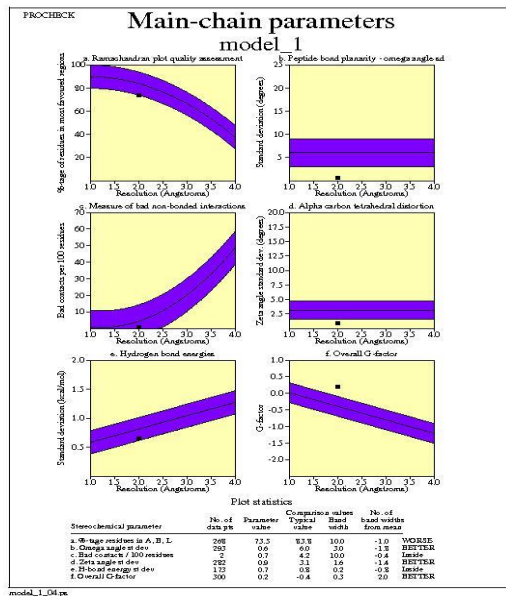


Fig 3: Main chain parameters

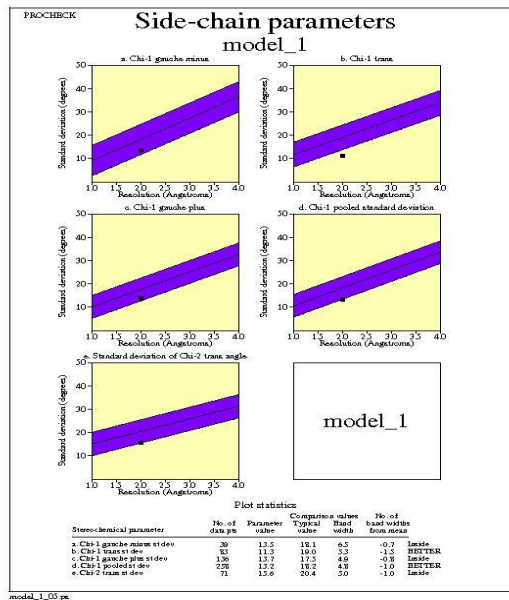


Fig 4: Graph for side chain residues

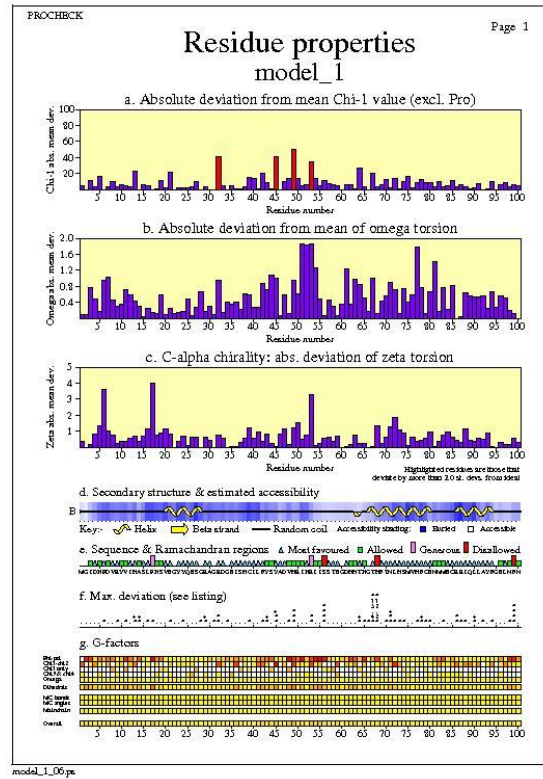
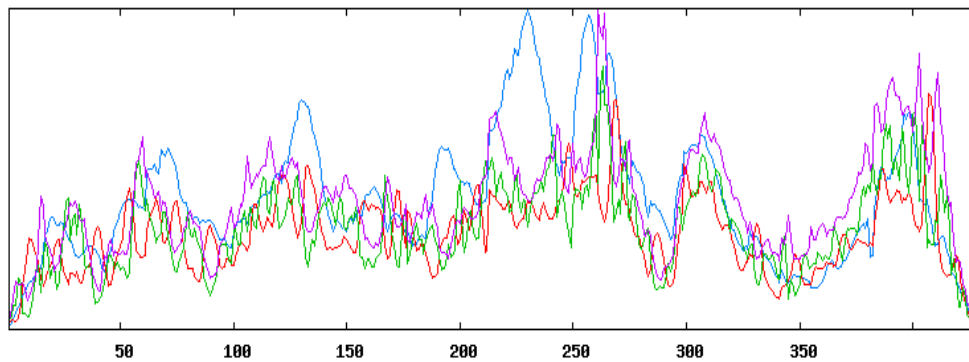


Fig 5: Graph for residual properties

SOPMA RESULTS

Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments.



10 20 30 40 50 60 70

```

| | | | |
MGIDKPDVRYVIHASLPKSVEGYQESGRAGRIDGEISHCILFYSYADVHRIKRIISSTEGDHHTKGTHTFT
etccttceeeeeccccchhhhhhhhtccccttcchhheeeccccchhhhhheehccccchhhhhhh
NLHSMVHFCENMMECRRIQLLAYFGELNFNKSFCKEHADVSCDNCAKPNQYKMRNVTEDVKKIVRFAQES
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhccccccccccccccccccccchhhhhhhhhhhhhhhhhhhhh
CEKVGSFRGRTAQQNRLTLNMLVDIFLGSKTAKVQTMFGMGAAYSRRHNADRLFKKLVLDNILEEDLYIT
cccccccccccccheehhhhhhhhtcccetceccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhheec
NNNQASVSARDEKVQECLKEVTDLCKQLGKAFGIHYFNIFSTATLKKIAEKLSPDPEVLLQIDGVTEDEKL
ctccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
EKYGAEVIQVLQKYSEWQLAEEQTDNAEDGWIDTAQGRSYGANEDDTESSYFRSQAAQGQKRKKAPFFK
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
FSKKKKPYSTTSANSKGRGYSGNKSWSASSRGGSTAAGRGRSSTGDAAASRRPGFMALPAPQTNQRPF
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
LKPSFSHVG
Cccccccc
    
```

Fig. 6: SOPMA results showing deviations, coils, helix and sheets

PHYRE 2 Result

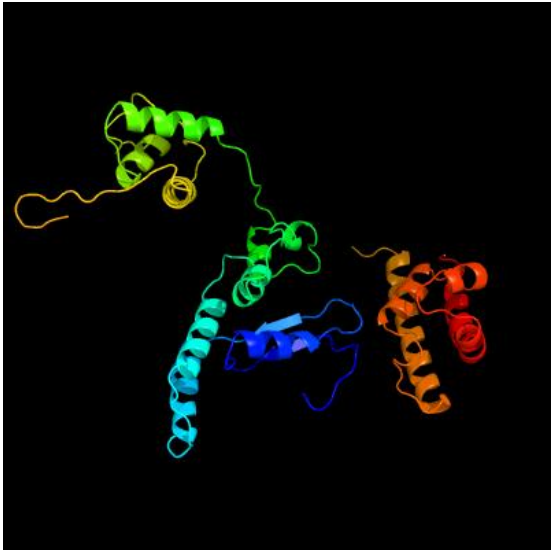


Fig. 7: Secondary structure generated by Phyre 2 software

IV. CONCLUSION

On the basis of various structural and physiochemical parameters assessment, it can be concluded that the predicted three dimensional structure of protein of Bloom syndrome is stable. The Ramchandran plot data showed that nearly 73% of the residues for each protein lies in the core region. The maximum percentage of the residues lying in the core region represents the best stereo chemical quality of the protein structure which is further supported by its nearly 98% residues lying in the most favoured and allowed regions. Only 0.08 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best

stereochemical quality and stability. The three dimensional structures generated using SWISS Model software and validated by PROCHECK software showed the best identity, lowest energy in Kcal/mol and root mean square deviation (RMSD) were considered as the best optimized and superimposed model. Since no effective therapeutic or vaccine is available for Bloom syndrome. Structural information of this model can be effectively used and can be further implemented in future drug designing.

V.ACKNOWLEDGEMENTS

We would like to acknowledge ITM University, Gwalior for its timely support and also the groups who supported us directly or indirectly in making this research a success. We would also thank the softwares without whose support this work would have not been completed.

REFERENCES

1. Karen H. Almeida: Kalmeida@ric.edu, Critical Interaction Domains between Bloom Syndrome Protein and RAD51, Protein J. 2011 January; 30(1): 1–8. doi:10.1007/s10930-010-9295-8.
2. Mao FJ, Sidorova JM, Lauper JM, Emond MJ, Monnat RJ. The human WRN and BLM RecQ helicases differentially regulate cell proliferation and survival after chemotherapeutic DNA damage. Cancer Res. 2010; 70:6548–6555. [PubMed: 20663905].
3. Mao FJ, Sidorova JM, Lauper JM, Emond MJ, Monnat RJ. The human WRN and BLM RecQ helicases differentially regulate cell proliferation and survival after chemotherapeutic DNA damage. Cancer Res. 2010; 70:6548–6555. [PubMed: 20663905].
4. Talia Hatkevich¹ and Jeff Sekelsky, Bloom Syndrome Helicase in meiosis: Pro-crossover functions of an anti-crossover protein, Bioessays. 2017 September; 39(9): .doi:10.1002/bies.201700073.
5. Bachrati CZ, Borts RH, Hickson ID. Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase. Nucleic Acids Res. 2006; 34:2269–79. [PubMed: 16670433].
6. Adams MD, McVey M, Sekelsky J. Drosophila BLM in double-strand break repair by synthesis independent strand annealing. Science. 2003; 299:265–7. [PubMed: 12522255].
7. NORD gratefully acknowledges Chris Cunniff, MD, Director of the Bloom's Syndrome Registry, and Amir R. Djavid, BA, Research Assistant at the Bloom's Syndrome Registry, for assistance in the preparation of this report.
8. NORD gratefully acknowledges Chris Cunniff, MD, Director of the Bloom's Syndrome Registry, and Amir R. Djavid, BA, Research Assistant at the Bloom's Syndrome Registry, for assistance in the preparation of this report.
9. Mao FJ, Sidorova JM, Lauper JM, Emond MJ, Monnat RJ. The human WRN and BLM RecQ helicases differentially regulate cell proliferation and survival after chemotherapeutic DNA damage. Cancer Res. 2010; 70:6548–6555. [PubMed: 20663905]
10. Giannelli, F., Benson, P.F., Pawsey, S.A. and Polani, P.E. (1977) Nature, 265, 466–469.
11. Lonn, U., Lonn, S., Nylen, U., Winblad, G. and German, J. (1990) Cancer Res., 50, 3141–3145.
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE, The Protein Data Bank. Nucleic Acids Res, 28 (1): 235-42, 2000.
13. Guruprasad K, Reddy BVP, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng 4: 155-164.
14. Gill SC, Von Hippel PH (1989) Extinction coefficient. Anal Biochem 182: 319-328
15. Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. J Biochem 88: 1895- 1898
16. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. J Mol Biol 157: 105- 132
17. Gasteiger E (2005) Protein Identification and Analysis Tools on the Expasy Server. In: John M. Walker, The Proteomics Protocols Handbook, Humana Press 571-607.
18. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-684.
19. Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics 18: 213-214
20. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201
21. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.
22. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8: 477-486
23. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56.