Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env.Pharmacol. Life Sci., Vol 4 [7] June 2015: 87-94
©2014 Academy for Environment and Life Sciences, India
Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.533
Universal Impact Factor 0.9804

ORIGINAL ARTICLE

In silico study of Cystatin C protein to Develop an ELISA kit using Computational tools and servers

Zahra Abdollah[1], Ali Karami[2]*, Jafar Amani[3], Samaneh Khodi[1], Elaheh Gheybi[4]

1. Applied Biotechnology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
2. Molecular Biology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
3. Molecular Microbiology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
4. Department of Biology, Faculty of Science , University of Guilan, Rasht, Iran

E. mail: alikarami1@yahoo.com

ABSTRACT

In this paper cystatin C (Cys-C) is analyzed and characterized using In silico tools to investigate whether the sequence contains suitable epitopes to be applied for designing an ELISA kit. Secondary structure consensus prediction was performed using SPOMA method, PSIPRED, ProtParam, Pepstats, Scratch Protein Predicton and Recombinant Protein Solubility Prediction. Tertiary structure prediction was performed by I-TASSER server. Bcepred, Discotope were applied to predict B cell epitopes from three dimensional protein structures. The codon usage bias in E. coli was upgraded the CAI from 0.71 to 0.87, and GC content and unfavorable peaks were optimized to prolong the half-life of the mRNA. Secondary structure study showed Cys-C as a hydrophobe, and insoluble protein with improbability of expression in inclusion bodies of 0.858. The aliphatic index computed by ExPasy's ProtParam inferred that Cys-C may be stable for a wide range of temperature. Structure analysis also showed Cys-C has predominant–α-helices most in the N-terminal which was in consistent with the results of tertiary structure analyses. HHpred detected the highest homology with cysteine protease inhibitor. Conclusively, the produced protein contains several predicted putative antigens which could be applied for antibody production and can be applied for designing an ELISA kit.

Keyword: Cystatin C, ELISA kit, Bioinformatics, Epitopes

Received 12.02.2015 Revised 09.04.2015 Accepted 20.04. 2015

INTRODUCTION

The estimated glomerular filtration rate (GFR) is the clinical parameter for evaluation of kidney function [8]. *Creatinine* as an endogenous marker and by product of muscle metabolism is affected by age, sex, diet, muscle mass, race and tubular Creatinine secretion especially in reduced GFR. For this reason the interest in cystatin C as a precise and accurate estimation of GFR has raised along the last few years [6]. Serum cystatin C concentration could be a good indicator (predictor) for acute kidney injury development [1]. In addition to role of cystatin C in kidney disease, it has been shown that to play few roles in cardiovascular disease, brain disorders for example Alzheimer's disease, patient with amyotrophic lateral sclerosis, thyroid dysfunction, glucocorticoid therapy, cancer and HIV function [4, 11, 16]. Cystatin C is a potent inhibitor proteinases and one of the most important extracellular inhibitors of cysteine proteases such as lysosomal cathepsins and proteases of parasites and microorganisms. It belongs to the type 2 cystatin gene family, a low molecular weight protein (13kD) consisting 120 amino acids that are removed from vessels by glomerular filtration rate decreases the blood levels of cystatin C increases [2, 14]. This single non-glycosylated polypeptide contains four disulfide-paired cystatin C residues that are present in all human body fluids highest concentrations in serum, and then milk, tears and saliva [15]. It isn't secreted by renal tubules but increasing in urinary cystatin C excretion is the result of (is occurred by) renal tubular injury [17, 19]. Therefore serum cystatin C measurement is diagnostic [10]. In this study we applied bioinformatics tools to better understanding and characterizing the Cystatin C. Codon Optimization, Investigation of RNA stability and comparison of C-terminal domain by itself and in complex with N-terminal, epitope mapping and prediction of tertiary structure are another aim of this study.

METHODS

Sequences, databases and construct design

The nucleotide sequence encoding Cys C (NM_000099.3) was obtained from the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov). The full length sequence was converted to FASTA format using the ReadSeQ sequence conversion server [7]. To optimize the chimeric gene, the in silico analysis was performed using an online optimization tool (http://www.genscript.com/index.html) and Kazusa codon usage database (http:// www.kazusa.or.jp/codon). The chimeric gene was designed for cloning and expression in *E. coli*. The chimeric gene was synthesized by BioMatick Molecular Biotech, Inc. (Canada). The program mfold (http://www.bioinfo.rpi.edu/applications/mfold) and CentroidFold (http://www.ncrna.org/centroidfold) were used to analyze the secondary structure of the gene mRNA [12]

Secondary structure prediction

Secondary structure consensus prediction was performed using *SPOMA* method (Self-Optimized Prediction Method with Alignment), and PSIPRED [5]. The amino acid composition and the Hydrophobicity and Hydrophilicity analysis of amino acid sequence were fulfilled using the pepstats analysis tool and Protscale program, respectively [9, 13]. To predict the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) ProtParam was used (http://web.expasy.org/protparam/). To predict the probability of inclusion body was studied by Pepstats. The solubility was predicted by Scratch Protein Predictor (http://scratch.proteomics.ics.uci.edu/) and Recombinant Protein Solubility Prediction (http://www.biotech.ou.edu/).

Tertiary structure prediction

Structure prediction was performed by I-TASSER server and was uploaded to the Swiss-PdbViewer server to depict the tertiary structural illustrations [18]. HHpred at Max-plank Institute for Developmental Biology (http://toolkit.tuebingen.mpg.de/hhpred) were run to detect homolog domains. I-TASSER was run for prediction of 3D structure. The resulted PDB file was used for conformational Epitope mapping. The generated model was subjected to several repeated cycles of energy minimization using SPDBV software Model stability was evaluated using Ramachandran plot analysis.

Prediction of B-cell epitopes

The amino acid sequence was analyzed using three web based B-cell epitope prediction algorithms; Bcepred (http:/ /www.imtech.res.in/raghava/bcepred/). Continuous B cell epitopes prediction methods based on physico-chemical properties on a non-redundant dataset, and the Discotope (http://www.cbs.dtu.dk/services/DiscoTope/ Server) for predicting discontinuous B cell epitopes from three dimensional protein structures. Briefly, chimeric proteins were analyzed first for continuous B-cell epitopes using Bcepred and then using the Discotope server to predict discontinuous B cell epitopes. Finally, we used the VaxiJen server to predict the immunogenicity of the whole antigen and its subunit vaccine [3].

RESULTS AND DISCUSSION

Bioinformatics analysis of the wild type and optimized CTD sequences Rare codons in mRNA leads to a higher-order secondary structures which takes time to drive ribosome through the critical region, therefore, to have a high-level protein production biased codon usage can be applied and rare codons tRNAs that are rarely used in E coli host cell can be replaced, including AGG/AGA (arginine), CGG (arginine), AUA (isoleucine) CUA (leucine) CCC (proline), and GGA (glycine). The repeated sequences were also avoided.

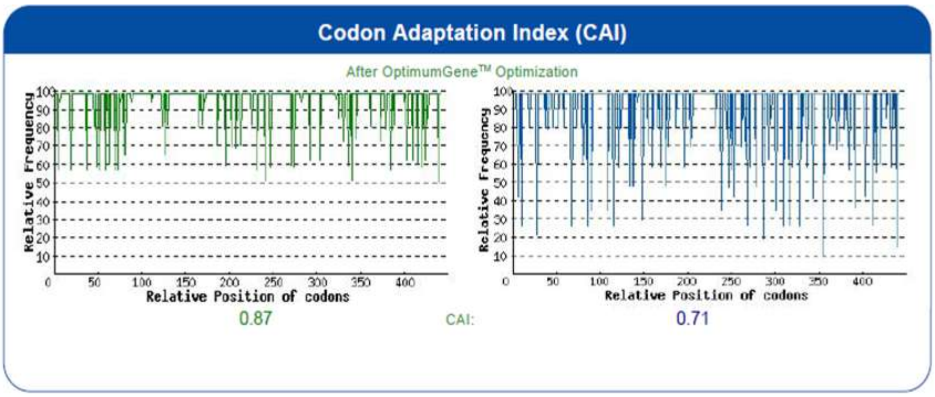

Fig. 1a: the distribution of codon usage frequency along the length of the gene sequence

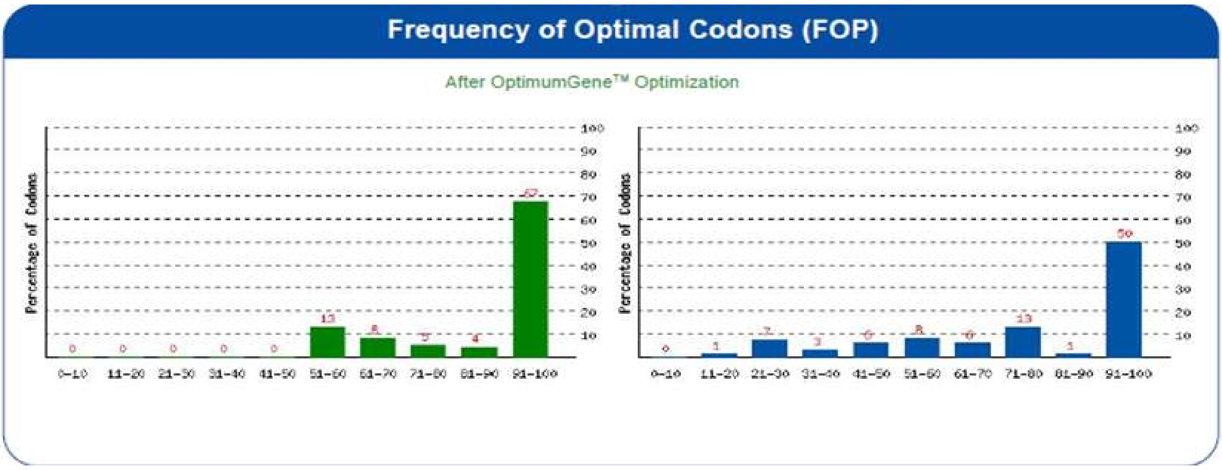

Fig. 1b: The percentage distribution of codons in computed codon quality groups

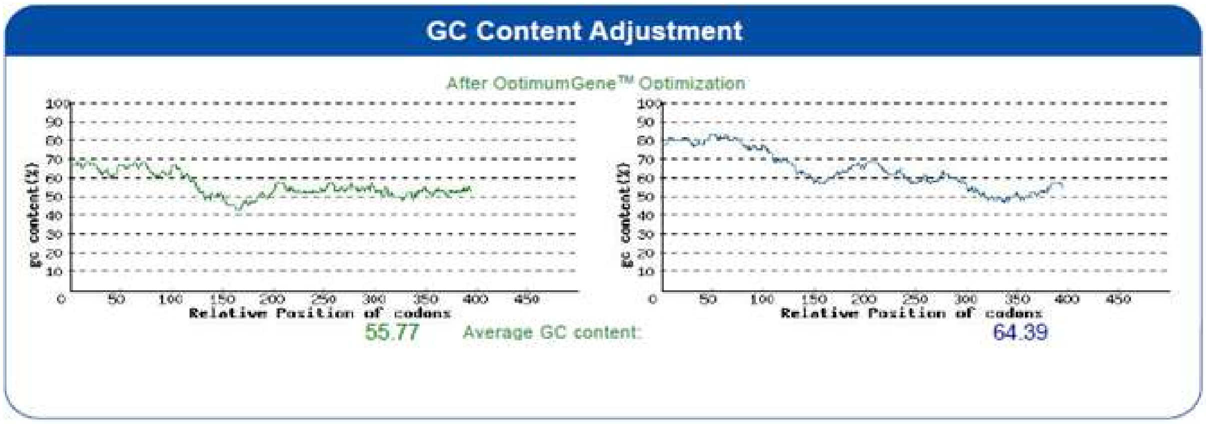

Fig. 2: The ideal percentage range of GC content is between 30-70%

A wide variety of factors regulate and influence gene expression levels, and OptimumGeneTM algorithm takes into consideration as many of them as possible, producing the single gene that can reach the highest possible level of expression. In this case, the native gene employs tandem rare codons that can reduce the efficiency of translation or even disengage the translational machinery. The codon usage bias in *E. coli* is changed by upgrading the CAI from 0.71 to 0.87 (fig. 1a). A CAI od 1.0 is considered to be perfect in the desired expression organism and a CAI of>0.8 is regarded as good, in terms of high gene expression level. The percentage distribution of codons showed 67% of codon quality groups in the value of 91-100 which is set for codon with the highest usage frequency for a given amino acid in the E. coli (fig. 1b).

GC content and unfavorable peaks were optimized to prolong the half-life of the mRNA, and peaks of %GC content in a 60 bp window were removed (fig. 2). The Stem-Loop structures, which impact ribosomal binding and stability of mRNA, were broken. In addition, the optimization process has screened and successfully modified those negative cis-acting sites as listed in the introduction. Furthermore, *Bam*HI restriction enzyme site was added in 5' and *Hin*dIII restriction enzyme site was added in 3' site. No major difference between the synthetic gene and the original one was observed and their structures were compatible with each other.

mRNA structure predictions

Mfold analysis showed a minimum free energy of -133.14 kcal/mol for secondary structure. The first nucleotides at the 5′ end did not have a long stable hairpin or pseudoknot. Predicted energy by Centroidfold server was -93.43 Kcal/mol (Fig. 3). The data showed the mRNA was stable enough for efficient translation in the new host

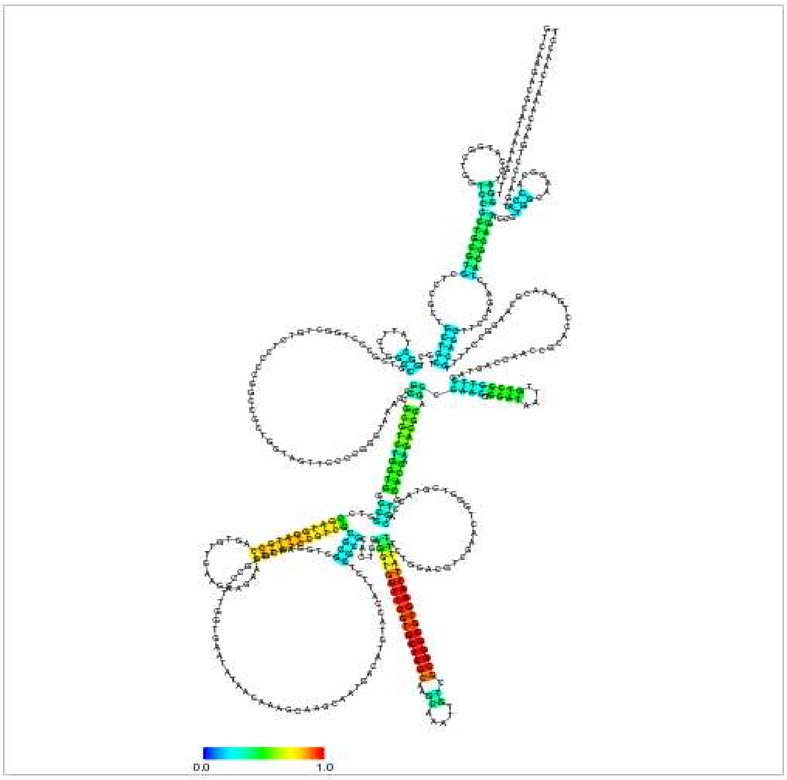

Fig. 3: Centroidfold server predicted mRNA structure with an energy of -93.43 Kcal/mol

Protein secondary structure prediction

Protparam computes various physico-chemical properties that can be deduced from a protein sequence including the molecular weight (15.7992 kDa), theoretical pI (9.00), half-life (>10 h in E. coli), amino acid composition (20.3% Ala, 29.1% Cys, 28.3% Gly, 22.3% Thr), atomic composition ($C_{698}H_{1115}N_{199}O_{203}S_8$), estimated half-life (>10h in e. coli), instability index (49.00), aliphatic index (84.25) and grand average of hydropathicity (GRAVY: -.112).consequently, Expasy's ProtParam classifies the Cys C as an unstable protein. The RPSP and SPP showed a 0.0 percent chance of solubility when overexpressed in *E. coli.* Total number of negatively and positively charged residues was also computed 5.5. Improbability of expression in inclusion bodies was also defined as 0.878 by Pepstats. The profile produced by Kyte & Doolittle amino acid scale by Pepstats and ProtScale calculates statistic also confirmed the hydrophobicity (Fig. 4).

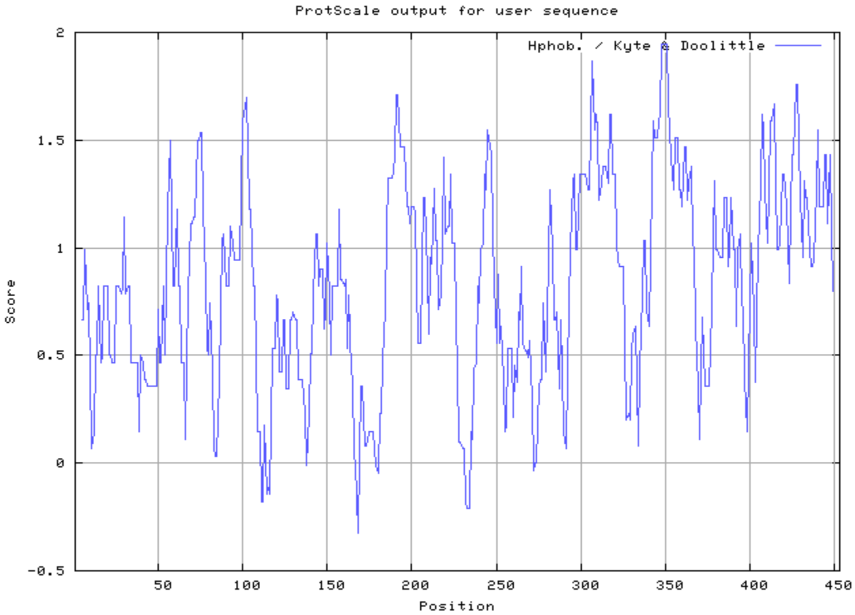

Fig. 4: The result of ProtScale confirmed the query as a hydrophobic protein

The total amino acid sequence length was 146, with an Alpha helix (Hh) motif accounting for 58 amino acids or about 39.73% of the protein. An extended strand (Ee) motif was 24 amino acids in length accounting for 16.44% , a Beta turn (Tt) motif of 9 amino acids represented 6.16% and a random coil (Cc) motif of 55 amino acids accounted for 37.67%. There were no 310 helix (Gg), Pi helix (Ii), Beta bridge (Bb), Bend region (Ss), Ambigous states or other states (fig. 5). The parameters were window width of 17 with a similarity threshold of 8 and number of states of 4.

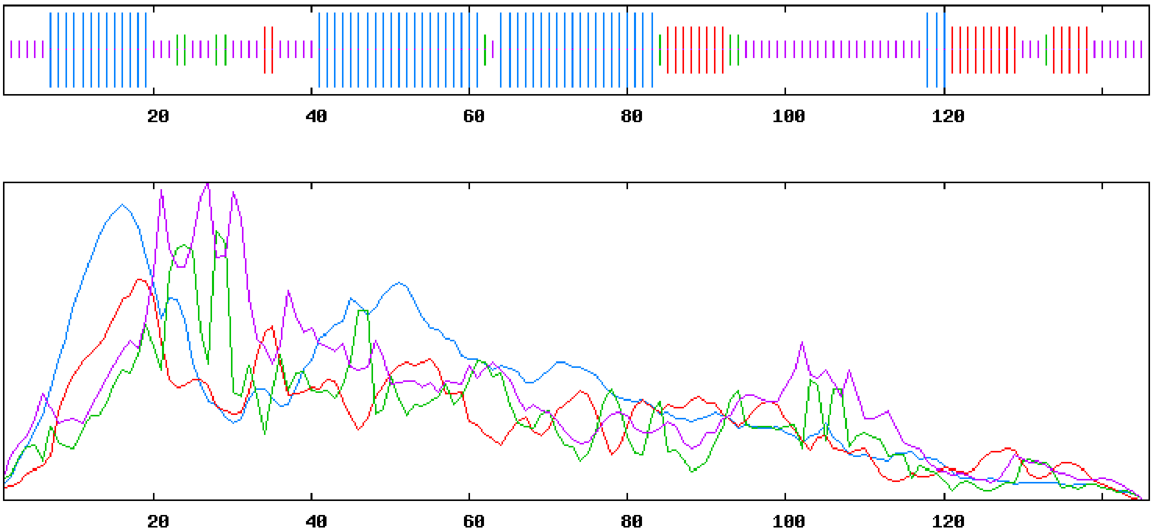

Fig. 5: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments (SOPMA), Graphical results for secondary structure prediction of chimeric protein. Extended strand: purple, Coil: red, Helix: blue.

Tertiary structure prediction

I-TASSER ab initio online software was used to draw the tertiary structural illustrations with Swiss-PdbViewer and DS Visualizer in order to determine the final structure of the protein (Fig. 6a). The quality of predicted models was evaluated by C-score which is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C score of higher value signifies a model with a high confidence and vice-versa. The model with the maximum C-score of -0.98 with estimated TM-score of 0.59±0.14 and estimated RMSD of 6.8±4.0Å obtained. The results showed several α-helices most in the N-terminal, which is in consistent with the results of our secondary structure analyses (Fgure 6b).

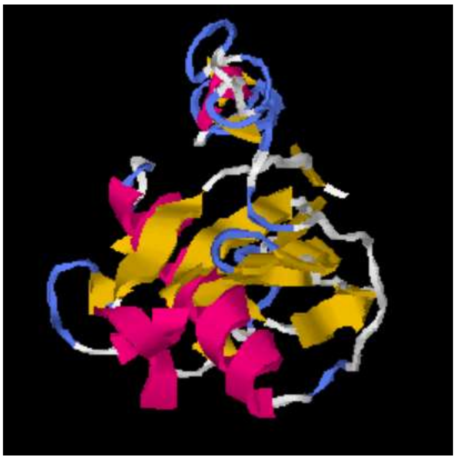

Fig. 6a: In silico 3-Dimensional Structure of Cys Cby Swiss-PdbViewer

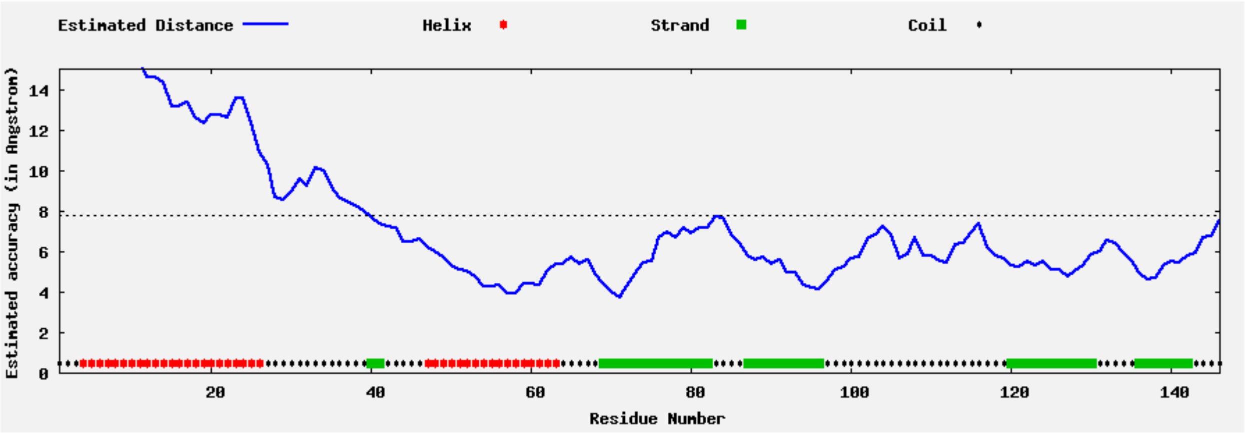

Fig. 6b: Several α-helices and Beta strands are shown in the N-terminal and C-terminal, respectively.

The highest homology was detected with cysteine protease inhibitor (3gax_A) with E-value of 4.6e-36 and Score of 208.58 (Fig. 7).

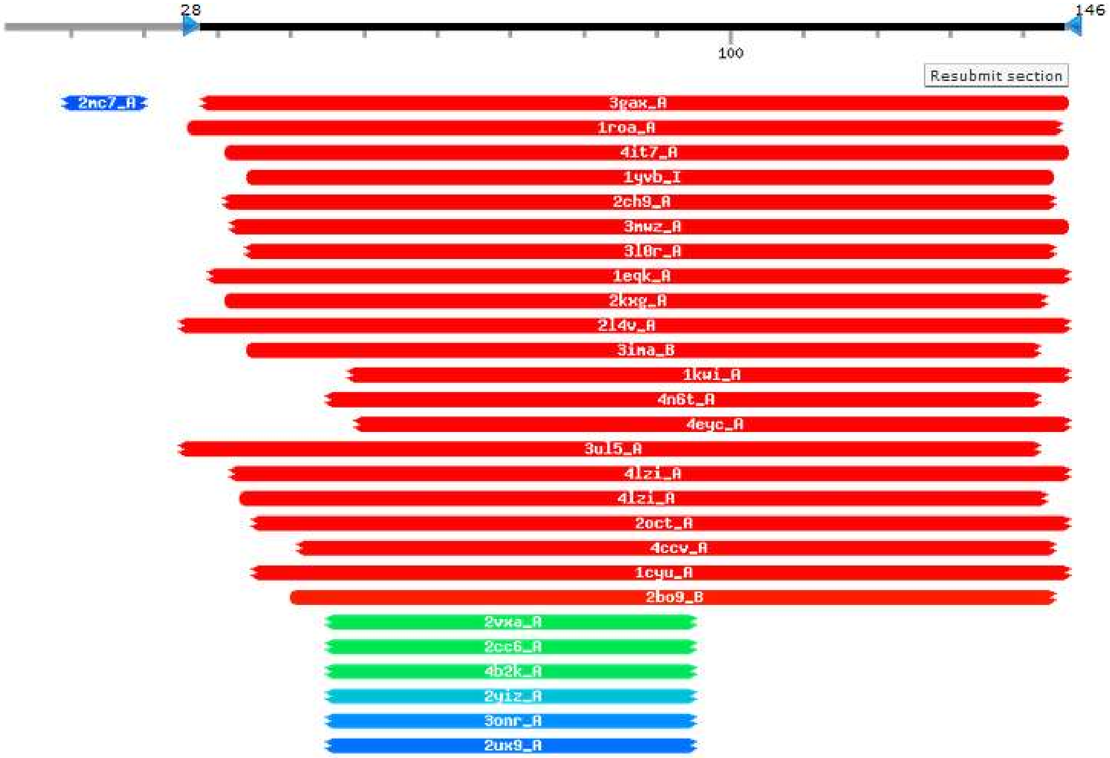

Fig. 7: HHpred detected hits to coiled coil-containing proteins. The highest homology was detected with cysteine protease inhibitor.

Prediction of B-cell epitopes

B-cell epitope analysis theoretically indicates that the DNA fragment involving the predicted putative antigens could be applied for antibody production. We evaluated the performance of existing linear B-cell epitope in the sequence using prediction methods based on physico-chemical properties such as Hydrophilicity method, Accessibility method, Antigenicity method, Flexibility method and secondary structure analysis on a non-redundant dataset (table 1). To obtain a more sufficient epitope and confirm the data, ABCpredis was used to predict linear B cell epitope with threshold setting of 0.51 and are ranked according to their score obtained by trained recurrent neural network (Table 2). All the peptides shown here are above the threshold value chosen. Prediction Servers predicted that overall Prediction for the Antigen by Discotope server was 0.6512 which means the sequence is probably antigen. The threshold for this model was 0.4.

Table 1: comparision of Linear B cell epitope based on single characters including hydrophilicity, Flexibility, antigenicity and exposed surface

Hydrophilicity	SPAAGSSPGKP, DASVEEEGVRR, GEYNKASNDM, GRTTCTKTQPN, SKSTCQDA
Flexibility	SPAAGSSPGKP, SVEEEGV, GEYNKAS, GTMTLSKST
Exposed surface	DQPHLKRKA
Antigenic Propensity	VNYFLDVE, CPFHDQP, FCSFQIY

Table 2: The predicted B cell epitopes by ABCpredis

Rank	Sequence	Start position	Score
1	GRTTCTKTQPNLDNCP	95	0.93
2	PFHDQPHLKRKAFCSF	110	0.86
3	DASVEEEGVRRALDFA	41	0.84
4	EGVRRALDFAVGEYNK	47	0.82
5	AVPWQGTMTLSKSTCQ	129	0.81
6	LQVVRARKQIVAGVNY	73	0.80
7	SSPGKPPRLVGGPMDA	27	0.79
8	AGPLRAPLLLLAILAV	2	0.78
8	HLKRKAFCSFQIYAVP	116	0.78
9	GEYNKASNDMYHSRAL	58	0.67
10	MYHSRALQVVRARKQI	67	0.65
10	ILAVALAVSPAAGSSP	14	0.65
10	QPNLDNCPFHDQPHLK	103	0.65

CONCLUSION

Bioinformatics tools may provide a valuable addition to traditional experimental methods in developing applicable protein production. The findings from the in silico study of Cystatin C protein were informative for the development of an ELISA kit for diagnosis of some of disorders especially renal disfunction. In conclusion, this is among the very few reports mapping the *Cystatin C* epitopes for designing an ELISA kit.

ACKNOWLEDGMENTS

The authors would like to thank all colleagues in the Applied Biotechnology Research Center of Baqiyatallah Medical Sciences University for their Kind Contributions to this research.

CONFLICT OF INTEREST DECLEARTION

No conflict of interest

REFERENCES

1. Aaron, K. J.,Kempf, M. C..,Christenson, R. H..,Wilson, C. M..,Muntner, P.&Shrestha, S. (2012). Prevalence of proteinuria and elevated serum cystatin C among HIV-Infected Adolescents in the Reaching for Excellence in Adolescent Care and Health (REACH) study. J Acquir Immune Defic Syndr. 61(4):499-506.
2. Dharnidharka, V. R..,Kwon, C.&Stevens, G. (2002). Serum cystatin C is superior to serum creatinine as a marker of kidney function: a meta-analysis. American Journal of Kidney Diseases. 40(2):221-226.
3. Doytchinova, I. A.&Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 8(4.
4. Gauthier, S.,Kaur, G..,Mi, W..,Tizon, B.&Levy, E. (2011). Protective mechanisms by cystatin C in neurodegenerative diseases. Front Biosci (Schol Ed). 3(541-54.
5. Geourjon, C.&Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 11(6):681-4.
6. Ghys, L.,Paepe, D..,Smets, P..,Lefebvre, H..,Delanghe, J.&Daminet, S. (2014). Cystatin C: a new renal marker and its potential use in small animal medicine. J Vet Intern Med. 28(4):1152-64.
7. Gilbert, D. (2003). Sequence file format conversion with command-line readseq. Current protocols in bioinformatics/editoral board, Andreas D. Baxevanis...[et al.]. Appendix 1E.
8. Iwasaki, M.,Taylor, G. W.,Sato, M.,Nakamura, K..,Yoshihara, A.&Miyazaki, H. (2014). Cystatin C-based estimated glomerular filtration rate and periodontitis. Gerodontology.
9. Kyte, J.&Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 157(1):105-132.
10. Newman, D. J..,Thakkar, H.,Edwards, R. G..,Wilkie, M..,White, T..,Grubb, A. O.&Price, C. P. (1995). Serum cystatin C measured by automated immunoassay: a more sensitive marker of changes in GFR than serum creatinine. Kidney Int. 47(1):312-8.

11. Panaich, S. S.,Veeranna, V.,Bavishi, C.,Zalawadiya, S. K.,Kottam, A.&Afonso, L. (2014). Association of Cystatin C with Measures of Obesity and Its Impact on Cardiovascular Events Among Healthy US Adults. Metab Syndr Relat Disord.
12. Reeder, J.,Hochsmann, M.,Rehmsmeier, M..,Voss, B.&Giegerich, R. (2006). Beyond Mfold: recent advances in RNA bioinformatics. J Biotechnol. 124(1):41-55.
13. Rice, P.,Longden, I.&Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends in Genetics. 16(6):276-277.
14. Roos, J. F.,Doust, J,Tett, S. E.&Kirkpatrick, C. M. (2007). Diagnostic accuracy of cystatin C compared to serum creatinine for the estimation of renal dysfunction in adults and children--a meta-analysis. Clin Biochem. 40(5-6):383-91.
15. Shaw, P. A..,Cox, J. L..,Barka, T.&Naito, Y. (1988). Cloning and sequencing of cDNA encoding a rat salivary cysteine proteinase inhibitor inducible by beta-adrenergic agonists. J Biol Chem. 263(34):18133-7.
16. Sundelof, J.Arnlov, J,Ingelsson, E.,Sundstrom, J..,Basu, S..,Zethelius, B..,Larsson, A..,Irizarry, M. C..,Giedraitis, V..,Ronnemaa, E,Degerman-Gunnarsson, M..,Hyman, B. T..,Basun, H..,Kilander, L.&Lannfelt, L. (2008). Serum cystatin C and the risk of Alzheimer disease in elderly men. Neurology. 71(14):1072-9.
17. Wald, R..,Liangos, O..,Perianayagam, M. C,Kolyada, A.,Herget-Rosenthal, S..,Mazer, C. D.&Jaber, B. L. (2010). Plasma cystatin C and acute kidney injury after cardiopulmonary bypass. Clin J Am Soc Nephrol. 5(8):1373-9.
18. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 9(40.
19. Zhang, Z..,Lu, B..,Sheng, X.&Jin, N. (2011). Cystatin C in prediction of acute kidney injury: a systemic review and meta-analysis. Am J Kidney Dis. 58(3):356-65.

CITATION OF THIS ARTICLE

Zahra A, Ali K, Jafar A, Samaneh K, Elaheh G . *In silico* study of Cystatin C protein to Develop an ELISA kit using Computational tools and servers. Bull. Env. Pharmacol. Life Sci., Vol 4 [7] June 2015: 87-94