2012 4th International Conference on Bioinformatics and Biomedical Technology
IPCBEE vol.29 (2012) © (2012) IACSIT Press, Singapore

In silico structure characterization of Familial Mediterranean fever gene product (pyrin)

Lilit Nersisyan, Arsen Arakelyan [+]

Institute of Molecular Biology NAS RA

Abstract. Pyrin is responsible for the development of Familial Mediterranean fever. Using homology modeling combined with *ab initio* methods we've constructed full length structural models of pyrin-fl and pyrin-d2 isoforms and assessed their quality. These models can be used to further explore molecular functions of pyrin, as well as investigate the functional and structural changes caused by FMF-associated mutations.

Keywords: Familial Mediterranean fever, pyrin, protein modelling and assessment

1. Introduction

Familial Mediterranean fever (FMF, MIM 249100) is the prototypical recessively inherited autoinflammatory disease most commonly affecting the ethnic groups from Mediterranean basin. FMF is characterized by short recurrent bouts of fever and localized inflammation usually involving the peritoneum, pleura, joints, and skin [1]. Mutations in MEFV gene, which encodes a 781 amino acid (aa) long protein called pyrin (also named TRIM20) [2], have been shown to be responsible for FMF development [3]. Pyrin is mainly expressed in neutrophils, and monocytes/macrophages as well as in other cells, but at lower levels [4]. In addition to the 781 aa long product of MEFV (pyrin-fl), there are also other isoforms of this protein [4, 5]. The first identified and the most prevalent splice variant is the pyrin-d2, encoding a 570 aa long product, which results from an in-frame splice removal of exon 2 (aa residues 93-303) [5]. This and other isoforms can exist in the cell alongside the full-length pyrin, but may have different subcellular localizations [5]. The full length pyrin is mostly found in the cytoplasm and is reported to be associated with microtubules and actin filaments of the cytoskeleton, or distributed homogenously in the cytoplasm [4, 6]. Pyrin-d2 is usually located in the nucleus of polymorphonuclear cells and synovial fibroblasts [4, 5], however, the mechanisms which lead to localization of pyrin-d2 to the nucleus have not been deciphered so far. Exact biological functions of pyrin are yet to be elucidated: it is not known what role pyrin plays in immunity, particularly in inflammation. Association of pyrin with microtubules, led to the suggestion that the full-length pyrin could regulate migration, adhesion, phagocytosis and degranulation at the level of cytoskeleton [6]. The effects of pro- and anti-inflammatory mediators on MEFV mRNA expression suggest that pyrin may act as a downstream element in cytokine induced regulatory cascades [4].

Pyrin is a member of tripartite motif proteins' family (TRIM) and is also known as TRIM20 [7]. These proteins are typically characterized by the presence of N-terminal RING domain followed by one or two B-box type zinc finger domains followed by a coiled-coil domain linked to a C-terminal domain, which varies among different proteins of this family. Only three domains in pyrin have been well annotated so far: the N-terminal PYRIN domain (1-92 aa), type-2 B-box domain (375-412 aa) identified by the presence of corresponding homology or consensus sequences, and B30.2 (PRY/SPRY) domain isolated and characterized by X-Ray diffraction analysis [7]. The structure of the remaining regions of pyrin generally

[+] Corresponding author. Tel.: + 37410 282622; fax: +37410 282061.
E-mail address: aarakelyan@sci.am.

remains unknown, although a number of predictions by various computational approaches have been made. Particularly, presence of a coiled-coil has been predicted at the 413-577 region, a nuclear localization signal at the 420-437 aa region, high level of disorder has been assumed for the 93-303 aa region. The latter to a great extent limits the possibility to obtain crystal structure of full length pyrin. Moreover, it should be noted that predictions made are ambiguous, as they strongly depend on underlying algorithms and techniques.

Thus, limited data on secondary and tertiary structure of pyrin, the functional features of its isoforms and consequences of mutations greatly restrict our understanding of its function in normal and diseased conditions. In order to provide some clues into biological functions of pyrin, we have constructed 3D models of pyrin and its pyrin-d2 isoform through hybrid combination of homology based and *ab initio* methods. The models have been obtained through consensual combination of various predictions and simulations and simultaneous validation of the structures based on all the experimental information upon morphological and functional features of pyrin known so far. Thus, these models are valid for being used in further studies on the role of pyrin in FMF pathogenesis. Moreover, the strategy implemented in this study may be used for modelling of other poorly annotated proteins.

2. Methods

From dozens of available software and tools we have chosen those mentioned below that had the highest rank according to corresponding benchmark studies.

In order to predict the secondary structure content of pyrin we have used PHYRE2 [8] and PredictProtein web servers [9]. The GeneSilico MetaDisorder service (http://genesilico.pl/metadisorder/) was used to investigate the presence of disordered regions in pyrin. The presence of coiled-coil regions in the 413-577 sequence of pyrin was assessed through COILS server [10].

The search for structural templates for further use in homology modelling was done among resolved structures available in PDB (www.pdb.org) [11]. Sequence similarity search was performed with BLAST [12]. Threading was done with LOMETS [13] server. Since no template with significant homology was available for the 93-374 and 413-577 regions of pyrin, their initial structural models were obtained through I-TASSER [14] and Robetta [15] web servers, which make use of combination of homology based and *ab initio* modelling techniques. Assessment of the validity of the latter models was performed by hierarchical clustering based on pairwise RMSDs with MATLAB R2010a (Mathworks Inc, USA, www.mathworks.com).

Final models of pyrin-d2 and pyrin-fl based on obtained templates and initial structures generated by I-TASSER and Robetta servers were constructed with dope_loopmodel class implemented in MODELLER 9.9 [16].

Initial refinement of models was done with Chiron web server [17]. Further minimization was performed with NAMD 2.8 software [18]. Proteins with constraint backbone were subject to 20 ps (2 fs timestep) of conjugate gradient minimization in water box under periodic boundary conditions. Dielectric constant was set to 1.0, van der Waals interaction switching distance was set to 10A˚, and the cutoff value for electrostatic and van der Waals interactions was set to 12A˚. Visualization of the models was performed with VMD (http://www.ks.uiuc.edu/Research/vmd/) [19].

Parallel and final assessment of models was done based on DOPE potential, protein free energy calculated with CHARMM force field (by NAMD), PROCHECK (validates stereochemical quality of protein structures) [20] and QMEAN (a composite scoring function, which is able to derive both global and local error estimates of a protein model) [21] web-servers. The threshold for model or structure selection was chosen to be the 90^{th} percentile of the distribution of the corresponding assessment parameter among the models.

3. Results and discussion

3.1. Sequence based structural analysis of pyrin

The pyrin PRY/SPRY domain structure file was obtained from PDB (id: 2wl1A). Sequence similarity search and threading revealed two homologous structures with significant identity to N-terminal 1-92 aa (PDB ID: 2hm2, ASC2, pyrin domain-only protein from Homo sapiens, sequence identity: 36%, e-value:

3.30E-09) and 375-412 aa region of pyrin (PDB ID: 2yrg, B-box domain of human tripartite protein 5, sequence identity: 44%, e-value: 0.0000015).

The sequences spanning the 93-374 and 413-577 aa regions had no relevant homologues and were subject to additional analyses. Secondary structure predictions of pyrin 93-374 aa segment showed high degree of disorder (1-7% of helices), that implicates lack of well-defined secondary and tertiary structure. This was further confirmed by the results of Genesilico Metadisorder service (Fig. 1). In contrast, the region 408-594 aa was enriched with helices (84 %) suggesting possible presence of coiled-coils. Further analysis through COILS server confirmed the presence of coiled-coils with high significance (Fig. 1).

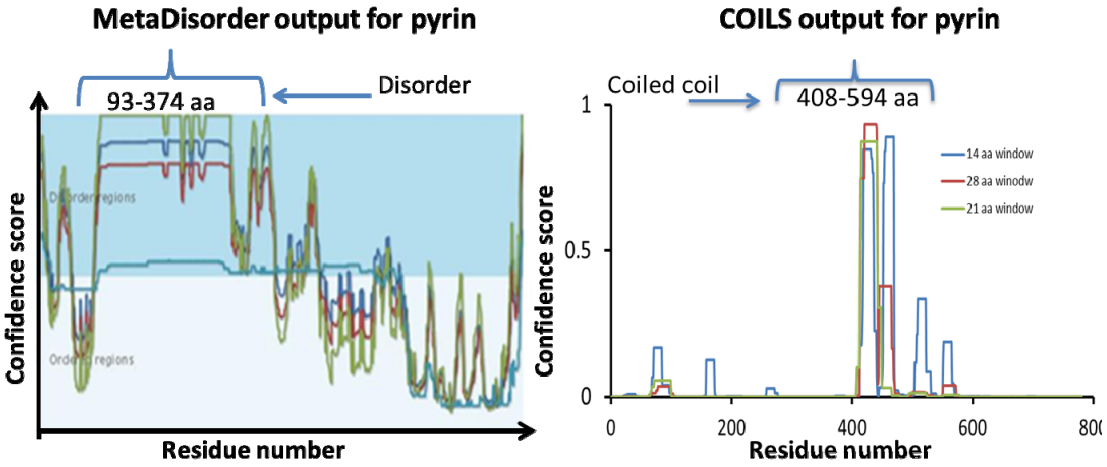

Fig. 1: Results of predictions of presence of disordered (left) and coiled-coil (right) regions in pyrin.

3.2. Model assembly

Since the disordered region encoded by exon-2 of MEFV creates difficulties for model construction, we have firstly constructed the structure of pyrin-d2 isoform. Later, this structure has been used as a template for structure prediction of the full length pyrin.

Models of 413-577 aa coiled-coil domain were constructed using I-TASSER and Robetta web servers. Overall 30 out of 200 models with highest scores have been chosen (10 from I-TASSER and 20 from Robetta). In order to assess the validity of these models, we have clustered them based on pairwise RMSDs, and obtained 8 separate clusters. Correspondingly, 8 representative models (one from each cluster closest to the cluster center) were chosen.

Having template homologues for PYRIN and B-box domains, the solved structure of pyrin B30.2 domain and 8 initial structure predictions of the coiled-coil domain, we obtained 8 possible structural combinations. Based on these combinations 8 models of pyrin-d2 were constructed with MODELLER dope_loopmodel class, treating the 70 aa gap between PYRIN and B-box domains as a loop (according to secondary structure predictions). Based on normalized DOPE scores and free energy estimations only one model was chosen and further refined with MODELLER, CHIRON and NAMD (Fig 1). In order to model the structure of full length pyrin based on the model obtained for pyrin-d2, we've tried to identify the spatial arrangement of the 93-374 region. This sequence was again submitted to I-TASSER and Robetta servers, and 23 models with high QMEAN scores were used for full length pyrin construction with MODELLER. The best 3 models of pyrin-fl were chosen based on normalized DOPE scores and refined as in the case of pyrin-d2 (Fig. 2).

3.3. Model assessment

Structure validation with PROCHECK showed that the percentage of residues lying within allowed regions of Ramachandran plot is in the range of 96.3-97.8% for all the models. Therefore, the models may be considered acceptable from stereochemical point of view.

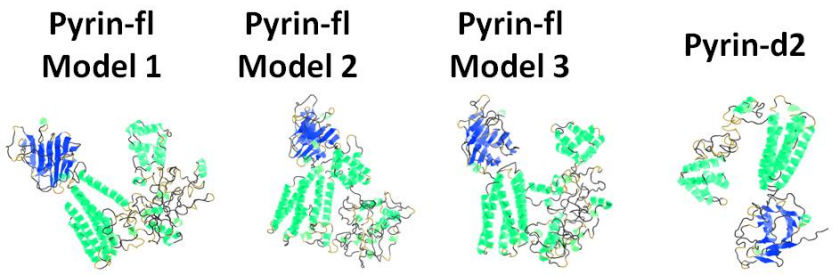

Fig. 2: Predicted models for pyrin-fl and pyrin-d2.

The free energy values and QMEAN scores for all the models are presented in figure 3. Since disordered regions are very flexible and natively disordered proteins are able to appear in several different conformations, we considered all the three models as possible candidates for the native fold of pyrin-fl.

Model	E, kcal/mol, CHARMM 27	QMEAN score	Z-score
Pyrin-d2	-11000	0.598	-1.84
Pyrin-fl_1	-15300	0.513	-2.77
Pyrin-fl_2	-15000	0.486	-3.06
Pyrin-fl_3	-15500	0.430	-3.68

Fig. 3: Free energy and QMEAN scores for computed models of pyrin.

The QMEAN score of the pyrin-d2 model is within 95% range of the scores for high resolution protein structures; whereas pyrin-fl models had lower scores (fig. 4). In order to compare our results with other theoretically predicted models we submitted 30 theoretical models of lengths ranging from 500 to 1000 aa obtained from ftp repository of PDB and constructed the distribution of their QMEAN scores (fig. 4). All the four models lye within 25th and 50th percentiles of this distribution indicating that our models are of good quality in comparison with other structures obtained through bioinformatics approaches. The fact that the models of pyrin-fl have lower QMEAN scores may be explained by the presence of the disordered region, since this region is the main source of residue error.

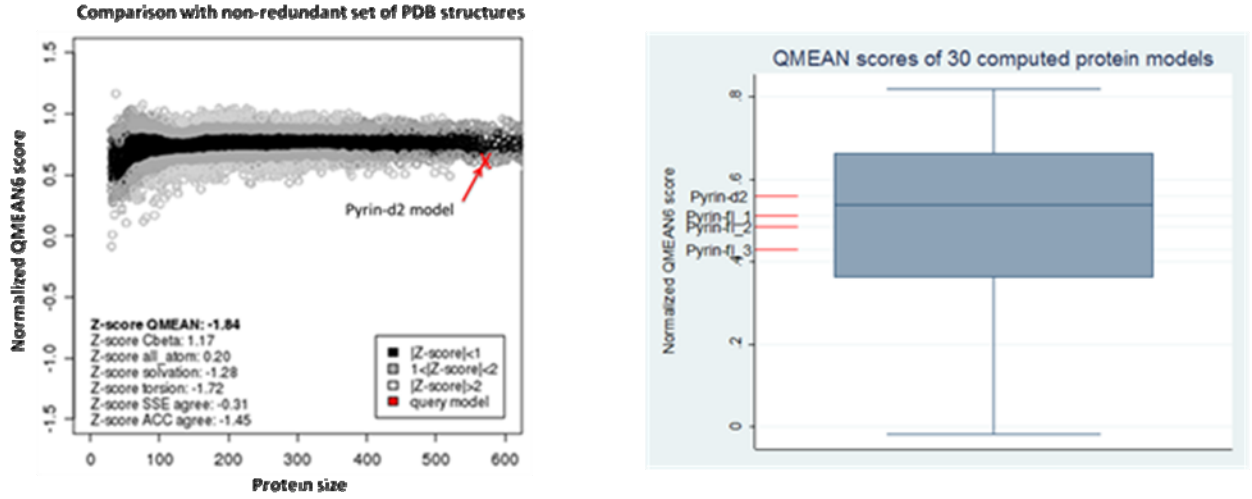

Fig. 4. Comparison of pyrin models' QMEAN scores with experimental (left) and theoretical (right) structures from PDB repositories.

In conclusion, we have obtained complete structural models of pyrin-fl and pyrin-d2 isoforms that can be used to further explore the nature of their interactions with other biomolecules as well as investigate the functional and structural changes caused by FMF-associated mutations. We believe that insights into pyrin structure-functional features will serve as a valuable clue for our understanding of FMF pathogenesis as well as other hereditary autoinflammatory disorders.

4. Aknowledgments

This study was supported by State Committee of Sciences of Ministry of Education and Science of the Republic of Armenia (YSSP grant 11B-1F014).

5. References

[1] Schaner, P.E. and D.L. Gumucio, Familial Mediterranean fever in the post-genomic era: how an ancient disease is providing new insights into inflammatory pathways. Curr Drug Targets Inflamm Allergy, 2005. 4(1): p. 67-76.

[2] A candidate gene for familial Mediterranean fever. Nat Genet, 1997. 17(1): p. 25-31.

[3] Touitou, I., Standardized testing for mutations in familial Mediterranean fever. Clin Chem, 2003. 49(11): p. 1781-2.

[4] Diaz, A., et al., Lipopolysaccharide-induced expression of multiple altenatively spliced MEFV transcripts in human synovial fibroblasts: a prominent splice isoform lacks the C-terminal domain that is highly mutated in familial Mediterranean fever. Arthritis Rheum, 2004. 50(11): p. 3679-89.

[5] Papin, S., et al., Alternative splicing at the MEFV locus involved in familial Mediterranean fever regulates translocation of the marenostrin/pyrin protein to the nucleus. Hum Mol Genet, 2000. 9(20): p. 3001-9.

[6] Mansfield, E., et al., The familial Mediterranean fever protein, pyrin, associates with microtubules and colocalizes with actin filaments. Blood, 2001. 98(3): p. 851-9.

[7] Weinert, C., et al., The crystal structure of human pyrin b30.2 domain: implications for mutations associated with familial Mediterranean fever. J Mol Biol, 2009. 394(2): p. 226-36.

[8] Kelley, L.A. and M.J. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc, 2009. 4(3): p. 363-71.

[9] Rost, B., G. Yachdav, and J. Liu, The PredictProtein server. Nucleic Acids Res, 2004. 32(Web Server issue): p. W321-6.

[10] Lupas, A., M. Van Dyke, and J. Stock, Predicting coiled coils from protein sequences. Science, 1991. 252(5009): p. 1162-4.

[11] Berman, H.M., et al., The Protein Data Bank. Acta Crystallogr D Biol Crystallogr, 2002. 58(Pt 6 No 1): p. 899-907.

[12] Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. 215(3): p. 403-10.

[13] Wu, S. and Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res, 2007. 35(10): p. 3375-82.

[14] Roy, A., A. Kucukural, and Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc, 2010. 5(4): p. 725-38.

[15] Kim, D.E., D. Chivian, and D. Baker, Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res, 2004. 32(Web Server issue): p. W526-31.

[16] Eswar, N., et al., Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics, 2006. Chapter 5: p. Unit 5 6.

[17] Ramachandran, S., et al., Automated minimization of steric clashes in protein structures. Proteins, 2011. 79(1): p. 261-70.

[18] Phillips, J.C., et al., Scalable molecular dynamics with NAMD. J Comput Chem, 2005. 26(16): p. 1781-802.

[19] Humphrey, W., A. Dalke, and K. Schulten, VMD: visual molecular dynamics. J Mol Graph, 1996. 14(1): p. 33-8, 27-8.

[20] Laskowski, R.A., et al., AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR, 1996. 8(4): p. 477-86.

[21] Benkert, P., M. Kunzli, and T. Schwede, QMEAN server for protein model quality estimation. Nucleic Acids Res, 2009. 37(Web Server issue): p. W510-4.