Structural and functional analysis of perforin mutations in association with clinical

data of familial hemophagocytic lymphohistiocytosis type 2 (FHL2) patients

by

Ömer AN

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Chemical and Biological Engineering

Koc University

December 2011

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ömer AN

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____
Prof. Özlem Keskin (Advisor)

_____
Prof. Attila Gürsoy

_____
Asst. Prof. Dr. Mehmet Sayar

Date: _____06.12.2011_____

ii

ABSTRACT

The immune system is the defense mechanism of the body against pathogenic agents and tumor cells. Perforin, a multi-domain pore-forming protein, plays a key role in the immune system as a cytotoxic effector molecule secreted by T-lymphocytes and Natural killer (NK) cells within granules, which also contain granzyme B that induces apoptosis. In synergy with granzyme B, perforin acts via pore formation at the cell membrane of virus-infected and transformed cells that are targeted to be eliminated. A vast number of observed mutations in perforin impairs this mechanism resulting in a rare but fatal disease, familial hemophagocytic lymphohistiocytosis type 2 (FHL2). FHL2 is an autosomal recessive disorder characterized by fever, hepatosplenomegaly, cytopenia, hyperferritinemia, hypertriglyceridemia and/or hypofibrinogenemia, decreased NK cell activity, increased CD25 level and hemophagocytosis. Here we report a comprehensive structural analysis of a collection of 76 missense perforin mutations associated with FHL2 based on a proposed pre-pore model. In our model, perforin monomers oligomerize having cyclic symmetry in consistent with previously found experimental constraints yet having flexibility in the size of the pore and the number of monomers involved. Clusters of the mutations on the model map to three distinct functional regions of the perforin. Stability change ($\Delta\Delta G$) calculations show that the mutations mainly destabilize the protein structure, interestingly however, A91V change, often suggested as a polymorphism, leads to a more stable one. Structural characteristics of mutations help explain the severe functional consequences on perforin deficient patients. Our study provides a structural approach to the mutation effects on the perforin oligomerization and impaired cytotoxic function in FHL2 patients.

ÖZET

Bağışıklık sistemi, vücudun hastalık yapıcı etkenlere ve tümör hücrelerine karşı savunma mekanizmasıdır. Çoklu işlevsel alt gruba sahip por oluşturucu bir protein olan perforin, bağışıklık sisteminde sitotoksik etkiye sahip efektör bir molekül olarak anahtar bir rol oynar ve T-lenfositleri ile doğal öldürücü hücreler tarafından granüller içerisinde salgılanır, bu granüller aynı zamanda programlı hücre ölümüne neden olan granzim B'yi de içerir. Granzim B ile sinerji oluşturan perforin, yok edilmek üzere hedeflenmiş virüsle enfekte olmuş ya da transformasyona uğramış hücrelerin hücre zarında por oluşturacak şekilde görev yapar. Perforinde gözlenmiş çok sayıda mutasyon bu mekanizmayı bozarak nadir görülen fakat ölümcül olan Ailevi Hemofagositik Lenfohistiositoz tip 2 (FHL2) hastalığına yol açar. FHL2 şu karakteristik özellikleriyle bilinen otozomal çekinik bir hastalıktır: ateş, karaciğer ve dalağın büyümesi, kan hücrelerinin sayıca azalması, kandaki ferritin miktarının artması, trigliserid miktarının artması ve/veya fibrinojen miktarının azalması, düşük doğal öldürücü hücre aktivitesi, yüksek CD25 seviyesi ve hemofagositoz. Biz bu çalışmada, tasarladığımız bir öncül-por modeline dayanarak FHL2 hastalığına ilişkin 76 yanlış anlam mutasyonunun kapsamlı bir yapısal analizini rapor ediyoruz. Modelimizde, perforin monomerleri daha önce bulunmuş deneysel kısıtlarla tutarlı ve çevrimsel simetriye sahip bir şekilde oligomer oluşturmaktadır, üstelik model por çapında ve içerdiği monomer sayısında esnekliğe sahiptir. Mutasyonların model üzerinde kümelenmesi perforinin üç farklı işlevsel bölgesinin haritasını çizmektedir. Kararlılık değişimi hesaplamaları mutasyonların çoğunlukla protein yapısının kararlılığını bozduğunu göstermektedir, bununla birlikte ilginçtir ki, sıkça polimorfizm olarak öne sürülen A91V değişimi, daha kararlı bir protein yapısına yol açmaktadır. Mutasyonların yapısal özellikleri perforin eksikliği görülen hastalarda şiddetli işlevsel bozukluklar görülmesini açıklamaya yardımcı olmaktadır. Bizim bu çalışmamız, perforinin oligomer oluşturması üzerindeki mutasyon etkilerine ve FHL2 hastalarının bozulmuş sitotoksik işlevlerine yapısal bir yaklaşım sağlamaktadır.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

vii

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

HLH	Hemophagocytic lymphohistiocytosis
FHL2	Familial hemophagocytic lymphohistiocytosis type 2
CTL	Cytotoxic T-lymphocyte
NK	Natural killer
CNS	Central nervous system
TNF-α	Tumor necrosis factor-alpha
IL	Interleukin
BMT	Bone marrow transplantation
MACPF	Membrane Attack Complex / Perforin
C2	Calcium-dependent membrane binding
EGF	Epidermal growth factor
RASA	Relative Accessible Surface Area
ΔΔG	Stability change upon mutation
HGMD	The Human Gene Mutation Database
LOVD	Leiden Open Variation Database
PMID	Pubmed Identifier
CDC	Cholesterol-dependent cytolysin

Chapter 1

1   INTRODUCTION

Rare diseases have always been a great challenge for diagnosis and treatment since their behaviour at molecular level are not well-understood. Hence the genetic cause or the underlying mechanism of the defective system are desirable to be known. The increasing clinical data about patients and the biological knowledge on the responsible genes or proteins serve for a better approach. FHL is only one of those rare diseases which still has mysteries although many promising steps have been taken since it was first described in 1952[1]. In this thesis, we study by structural means the mutations of a particular protein called perforin which is one of the genetic markers of FHL disease. We aim to contribute to the understanding of how perforin mutations lead to such a lethal disease at protein level.

The explosion of genetic era led to enormous accumulation of biological information thanks to the functional genomics and proteomics. As a result, computational studies have been gaining more and more importance in the recent years due to the increasing demand of dealing with large scale data. A specific field among them, the study of protein structure and function, provides an invaluable medium to have insight into many biological processes and cellular mechanisms at molecular level. In fact, the major reason of that human being suffer from diseases today underlies this biological background. In this thesis, we employ a number of computational methods to reveal the relationship between mutations and the resulting defect regarding a particular disease, FHL2.

In Chapter 2, an extensive literature review is presented. In the first part, the biological track of perforin from the general to the specific is given in the order: An overview of the immune system, cell-mediated cytotoxicity, perforin/granzyme B cell death pathway and structure and function of perforin. In the second part, FHL is introduced starting with an

overview followed by its clinical symptoms, diagnostic criteria, genetic basis, patients and treatment. The links between perforin and the disease are given where considered necessary.

Chapter 3 introduces in detail a number of computational methods used throughout the study. The chapter starts with a brief statement on that computational methods are integrated with the collected experimental data. The rationale of each method is explained at the beginning, and the technical details follow.

Chapter 4 includes a combined presentation of the results and the discussion. The characteristics of the perforin mutations are initially evaluated. Then the created perforin pre-pore model is shown and several aspects of the mutations are discussed on the model such as functional regions, interfaces etc. The mutations are analysed according to their effects on the stability and the structure of the protein. Based on relevant findings, the mutations are grouped with respect to their locations on the protein, energy changes and mutation types. The significance of our results is demonstrated with a case study. Finally, the special case of a particular mutation (A91V) is further discussed.

The final chapter states the conclusion of the study, followed by appendix, which includes the supplementary tables, and bibliography.

Chapter 2

2   LITERATURE REVIEW

2.1   Immune System and Perforin

2.1.1   Overview of the Immune System

The immune system is the defense of the body against infectious organisms, tumor cells and other invaders. Through collective action of a network of cells, tissues and organs called the immune response, the immune system attacks disease-causing agents and substances. We are safe from the continuous exposure to the bacteria, viruses and other pathogens thanks to our body's effective and quick immune response. The failure at one of the steps of the immune response cause serious problems, often leading to life-threatening diseases.

The main cells involved in the immune system are called leukocytes, also known as white blood cells, all arising from the stem cells in the bone marrow (**Figure 1**). Leukocytes are produced and stored in a continuous manner in the body parts such as thymus, spleen and bone marrow. There are many types of leukocytes based on their function and cell type, which can be basically classified into two; phagocytes and lymphocytes. Phagocytes are specialized to engulf/eat bacteria, viruses and dead cells, which are important for the subsequent immunity. On the other hand, lymphocytes have more specific mechanisms which allow to fight and destroy the foreign substances and organisms, and clear away infected host cells, as well as creating memory to remember and recognize previous invaders.
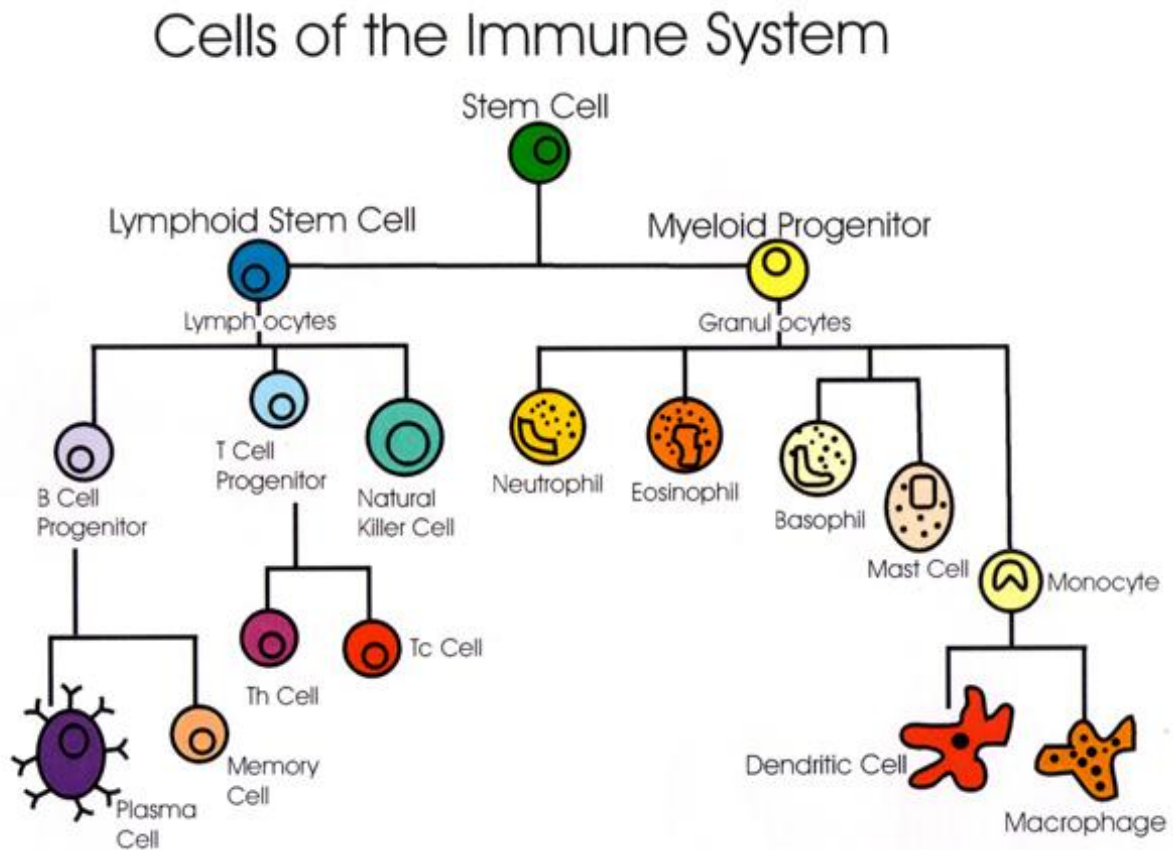
Figure 1: Cells of the immune system
(Source: http://www.textbookofbacteriology.net/cellsindefenses75.jpg)

There are two types of immunity; innate and adaptive immunity. Innate immunity comes with birth and provides a generic protection. Although an immediate immune response is created, innate immunity is non-specific to pathogens. Cells of innate immunity that are already circulating in the blood and body fluids destroy the invading organisms as soon as they are encountered. Among these cells, natural killer (NK) cells play an important role which do not directly attack the microbes but eliminates the virus-infected or transformed cells of the host, generating cytotoxicity. Adaptive immunity, on the other hand, is distinctive as being highly pathogen-specific and creating memory for quick and stronger response for the re-infection with the same pathogen. Because of its life-long development via passed diseases or immunization through vaccination, it is called as "adaptive". Adaptive immune system employs B lymphocytes as the major antibody producers and memory developers whereas T lymphocytes as the main effectors of cell-mediated immunity.

Upon detection of antigens (foreign substances that enter the body), B lymphocytes are initially triggered to produce antibodies, proteins which specifically bind to the antigens. Antibodies are kept within the body after first production, which makes them ready for the next challenge of the same infection. However, the destruction of the recognized antigens or the infected cells are provided by the T cells, that's why some of the T cells are called "killer cells" (**Figure 2**).
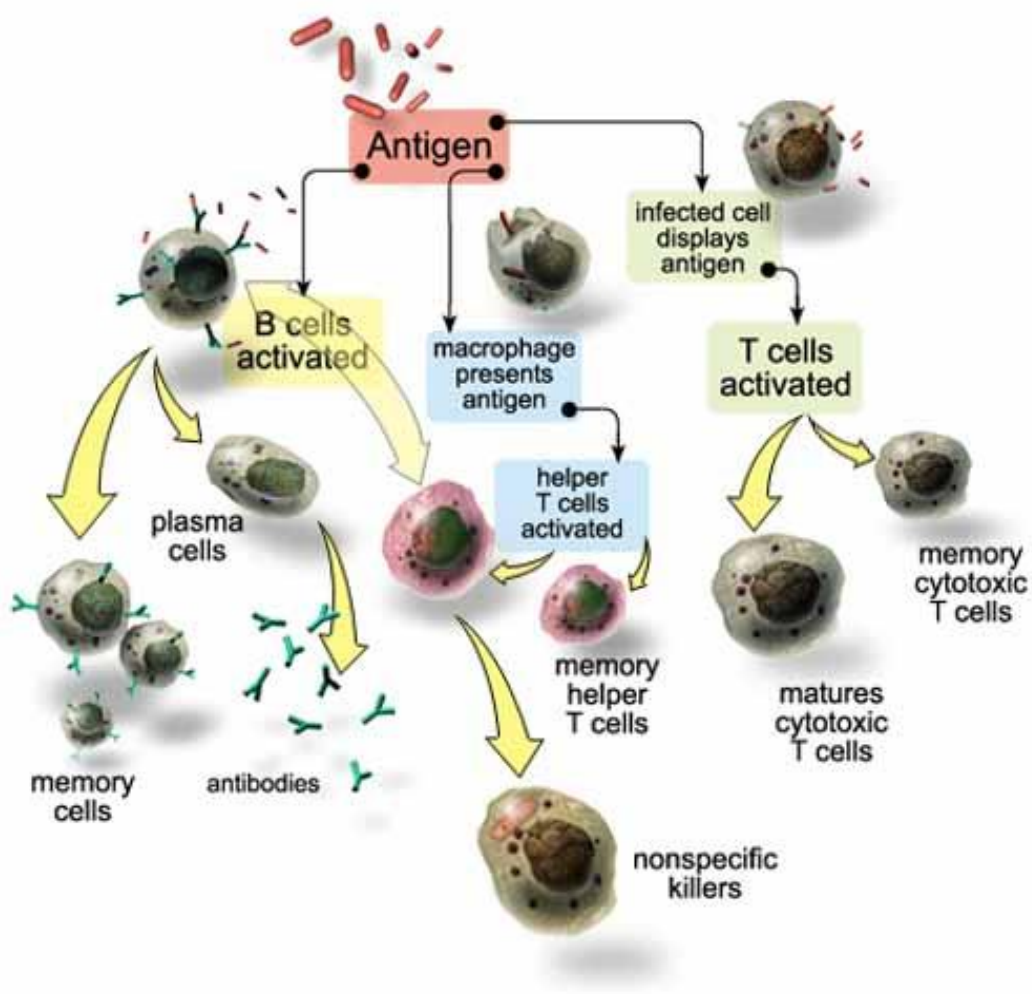
Figure 2: Activation of immune cells upon detection of antigen
(Source: http://www.uta.edu/chagas/images/immunSys.jpg)

All of these specialized cells and mechanisms protect the body against disease, which is called immunity. However, improper functioning of the immune system due to inborn errors or environmental factors result in several immunologic diseases. Infections are the most

common cause of human diseases, and the most common disease-causing microbes are bacteria, viruses and parasites. Disorders of the immune system can be categorized as:

1. Immunodeficiency (absence or dysfunction of a part of the immune system)
2. Autoimmune disorders (self-attack of the immune system)
3. Allergic disorders (overreaction to an antigen)
4. Cancers (out-of-control growth of cells of the immune system)

2.1.2 Cell-mediated Cytotoxicity

Among a variety of mechanisms that the immune system use to destroy target cells or microorganisms, two of them are in charge of providing cell-mediated cytotoxicity (direct cell-cell contact) in which apoptosis is eventually induced: Fas mediated and perforin/granzyme cell death pathways. NK cells (innate immunity) and cytotoxic T lymphocytes (adaptive immunity) directly take part in these pathways. Foreign organ transplants and autoimmune diseases are also prevented by the same mechanisms.

Fas mediated pathway is mainly responsible for activation-induced cell death[2]. Fas ligand (FasL) is expressed on the surface of cytotoxic T lymphocytes (CTLs) upon activation and binds to the Fas receptor (Fas, CD95) on the target cell. This binding induces the trimerization of the Fas molecules on the cell surface, which associates with a transducing molecule (Fas-associated death domain, FADD) to create a complex. This complex triggers apoptosis through the classical caspase cascade[3]. Fas mediated pathway is primarily used by CTLs, however perforin/granzyme pathway is dominant in both CTLs (CD8+) and NK cells playing a major role, which we will explain in detail next.

2.1.3 Perforin/Granzyme Cell Death Pathway

Cytotoxic lymphocytes (CTLs and NK cells) are the main effectors charged in killing of virus-infected and malignant cells. These cells contain granules in their cytoplasm secreted by the lysosome. NK cells have the granules readily available for a rapid response which can be initiated within minutes (typical to innate immunity), whereas CTLs require 1–3 days for

maximal activity (typical to adaptive immunity). Upon activation, these cells contact with the target cells where an immunological synapse is formed[4]. Secretory granules which include perforin and granzymes are released into the synaptic cleft. Due to the presence of Ca^{2+} ions at the intercellular space, the perforin is activated by assuming a conformational change. Activated perforin attaches to the target cell membrane where it polymerizes into pores (transmembrane channels). Granzymes enter into the cytosol effectively synergizing with these pores, where apoptosis is induced via several pathways[5]. In the absence of perforin, granzymes can still enter into the target cell, but adequate toxicity is not produced[6] The several models proposed for the mechanisms of killing via perforin/granzyme cell death pathway are given in **Figure 3**.

The indispensable role of perforin for cytotoxicity was shown in experiments performed with perforin-knockout mice[7-10]. In these experiments, the absence of perforin significantly reduced the cytotoxicity although the granzymes were present and the level of cytotoxic lymphocytes was unchanged. However, endogeneous injection of granzymes still trigger apoptosis showing that perforin is required for the uptake of the granzymes into the cell. Although perforin/granzyme B cell death pathway is not the only mechanism for cell-mediated cytotoxicity, it accounts for the majority.
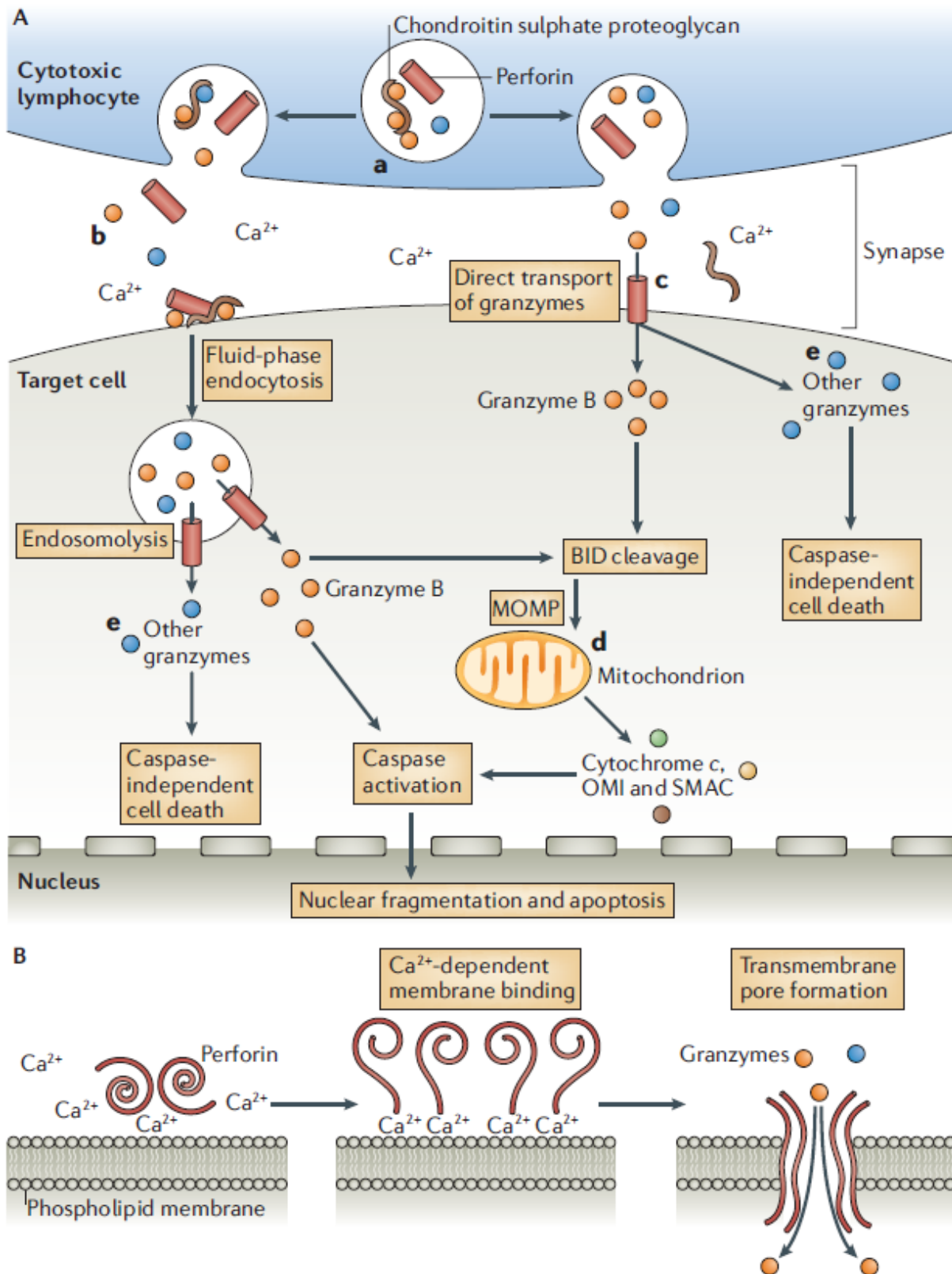
Figure 3: Models of perforin and granzyme synergy in target-cell death.
(Source: Voskoboinik et al. 2006[5])

2.1.4   Perforin: Protein Structure

Perforin is a multi-domain pore-forming protein that is known with its structural and functional resemblence to the complement component, C9. The sequence length of human perforin is 555 amino acids including three domains: MACPF (membrane attack complex / peforin), EGF-like (epidermal growth factor-like) and C2 (Ca-dependent lipid binding). MACPF is the largest conserved domain with 349 amino acids which is also found in bacterial cytolysins and complement componenets C6 to C9. Playing a major role in pore formation, MACPF accounts for the cytotoxic activity of perforin. It inserts into the cell membrane assuming large conformational changes and oligomerize to form transmembrane channels. Ca-dependent lipid binding domain is required for attachment of the perforin to the target cell membrane in the presence of Ca. Some Aspartate residues within this domain were shown to be important for Ca binding[11]. The exact function of the third domain is not known. The crystal structure of human perforin has not been solved so far.

2.2   Familial Hemophagocytic Lymphohistiocytosis (FHL)

2.2.1   Overview

Hemophagocytic lymphohistiocytosis (HLH) is a hematologic disorder analyzed in two broad forms, familial (primary) HLH and secondary HLH. Secondary HLH (sHLH) is a sporadic form associated with systemic infection, immunodeficiency or malignancy ranging in all age groups. Familial HLH (FHL), the inherited form, is autosomal recessive and genetically heterogeneous disorder of immune dysfunction, usually associated with early childhood. The disease was first described by James Farquhar and Albert Claireaux in 1952 in siblings[1]. Since then, increasing number of patients were reported as the accumulated knowledge led to better diagnosis. The annual incidence of the disease is given as 1/50000 live births[12], although various incidences were published depending on the prevalence in the ethnic groups studied: 1.2/1000000 in Sweden[12], 3.42/1000000 in Japan[13], 7.5/10000 in Turkey[14]. FHL is more common in some populations due to the consanguineous marriages. The disease is triggered by uncontrolled activation of T cells and macrophages and

overproduction of inflammatory cytokines[15]. Up to now, the only curative treatment is bone marrow transplantation (BMT).

2.2.2 Clinical Symptoms

FHL is characterized by fever, hepatosplenomegaly, cytopenia, hyperferritinemia, hypertriglyceridemia and/or hypofibrinogenemia, decreased NK cell activity, increased CD25 level and hemophagocytosis, as well as central nervous system (CNS) involvement and neurologic symptoms reported in some cases associated with late onset. The symptoms are caused by an increase of inflammatory cytokines, as a result of massive proliferation of activated macrophages and lymphocytes, leading multiorgan failure and eventually death.

The symptoms of FHL can be associated directly with the overexpression of cytokines. The fever is caused by high levels of inflammatory cytokines. The hypertriglyceridemia is caused by a decrease in lipoprotein lipase activity[16]. Activated macrophages secrete ferritin but also plasminogen activator which leads to high levels of plasmin that cleaves fibrinogen. Cytopenias are attributed to high levels of TNF-α and IFN-γ, but also to hemophagocytosis. However, hemophagocytosis play a secondary role since it is not found in all patients at onset. The neurological symptoms and the hepatosplenomegaly are most likely due to the infiltration of activated macrophages and lymphocytes in these organ[17].

2.2.3 Diagnosis

In the past years, HLH was difficult to diagnose, and sometimes even misdiagnosed since its presenting symptoms were very similar to some other genetic immunodeficiencies. The barely differentiated patients were sporadic and the lethal character of the disease led to rapid death of them. In the last decade, the identification of marker genes made it possible to diagnose FHL patients more accurately. The patients with presenting symptoms suspected for FHL were sequenced for a genetic variation in these genes and often found positive. Today

the diagnosis of HLH is made in accordance with diagnostic guidelines set up by the Histiocyte Society (**Table 1**).

Table 1: Diagnostic criteria for HLH as outlined in the HLH–2004 protocol by Histiocyte Society

The diagnosis HLH can be established if one of either 1 or 2 below is fulfilled.

1. A molecular diagnosis consistent with HLH.
2. Diagnostic criteria for HLH fulfilled (5 out of the 8 criteria* below).

A) Initial diagnostic criteria (<u>to be evaluated in all patients with HLH</u>).
<u>**Clinical criteria**</u>
* Fever
* Splenomegaly
<u>**Laboratory criteria**</u>
* Cytopenias (affecting ≥ 2 of 3 lineages in the peripheral blood:
Hemoglobin (<90 g/L), Platelets (<100 x 109/L), Neutrophils (<1.0 x 109/L)
(In infants <4 weeks: Hemoglobin <100 g/L)
* Hypertriglyceridemia and/or hypofibrinogenemia
(fasting triglycerides ≥3.0 mmol/L (i e ≥265 mg/dL), fibrinogen ≤1.5 g/L)
<u>**Histopathologic criteria**</u>
* Hemophagocytosis in bone marrow or spleen or lymph nodes.
No evidence of malignancy

B) New diagnostic criteria.
* Low or absent NK-cell activity (according to local laboratory reference)
* Ferritin ≥500 microgram/L
* Soluble CD25 (i.e. soluble IL–2 receptor) ≥2400 U/ml

2.2.4   Genetic Basis

The first reports on genetic defects causing FHL disease came in 1999[15, 18]. Two loci were identified although one of them still is not associated with a particular gene (FHL1). The second was the perforin, which accounts for the 20–40% of total FHL cases. In the following years three more loci each associated with a specific gene were identified that are all involved in one of the steps of perforin/granzyme cell death pathway. A small percentage of FHL cases still cannot be explained today by genetic mutations in one of above loci. According to the genes responsible for the disease, FHL is divided into five subtypes (**Table 2**).

Table 2: Chromosome location, genes and gene functions of currently known to cause FHL

Disease	Chromosome	Gene	Gene Function	Reference
FHL 1	9q21.3-q22	Unknown	Unknown	Ohadi et al 1999
FHL 2	10q22.1-q22	*PRF1*	Pore formation, induction of apoptosis	Stepp et al 1999
FHL 3	17q25.1	*UNC13D*	Vesicle maturation, docking and priming	Feldmann et al 2003
FHL 4	6q24.2	*STX11*	Vesicle transport, t-SNARE	zur Stadt et al 2004
FHL 5	19p13.3-p13.2	*STXBP2*	Vesicle trafficking and fusion with membrane	zur Stadt et al 2009

2.2.5 Patients

In the last decades, several hundreds of FHL patients were reported in the literature despite the given rarity of the disease. The number of reported FHL cases increased recently due to the better understanding of the disease at molecular level thanks to the defined marker genes (**see 2.2.4**), which improved its diagnosis in company with the employment of sequencing facilities. Several large scale studies presented the mutational spectrum, clinical findings and phenotypes of FHL patients[19-21], whereas many other research articles presented single cases or smaller groups of patients in detail.

FHL patients were diagnosed in different parts of the world including USA, Europe, Africa and Asia, having a higher incidence on Mediterranean, Turkish and Japanese origins. The family history of the patients show that the disease is closely associated with consanguineous marriages, a social problem observed in some territories of these countries. Another hyphothesis suggests the presence of a founder effect where the recessive mutated allele causative for FHL has been accumulated in the newly-established smaller population.

FHL patients suffer from the mutations in the genes given in the previous section, in which *PRF1* mutations have a higher percentage. Although the majority of the patients present the symptoms at early ages in the childhood, there are also examples of late onset cases associated with missense mutations. The patients initially receive a chemotherapy following bone marrow transplantation, which will be described next. The survival rate of the

diagnosed patients is currently very low, however, successful treatment together with BMT has been shown as curative.

2.2.6   Treatment

In 1994, an international colloborative study presented a treatment protocol called HLH–94 by Histiocyte Society which was consulted for long years in the treatment of HLH patients[22]. Before this protocol, the treatment of the patient was up to the treating physician since there was no guideline to follow. The protocol included chemotherapy and immunotherapy (etoposide, corticosteroids, cyclosporin A, and, in selected patients, intrathecal methotrexate) which induced remission in FHL[23]. Ten years later, a second revised version of the protocol was published as HLH–2004[24]. The patients were treated according to these protocols to keep them survive until a suitable donor is found for bone marrow transplantation (BMT). Several cases were reported with successful BMT in patients who showed no sign of FHL afterwards. Together with HLH–2004, BMT is the only curative treatment so far.

Chapter 3

3   METHODS

This study was entirely conducted by computational means, for this reason, a variety of bioinformatics tools, packages and webserves were extensively used. Python programming language was the main medium for data handling and file operations. VMD served as a molecule visualization and analysis tool at many steps. However, clinical and experimental findings from other studies were integrated, where the resulting combined knowledge was interpreted from a biological perspective. In fact, our study is the first example at such large scale that associates the present clinical data with a novel structural analysis on the topic covered.

3.1   Homology Modelling

Homology modelling is the construction of a model of the target protein from its amino acid sequence based on a related homologous protein as the template, whose three-dimensional structure is experimentally known. If the structure of the protein of interest is not known, homology modelling is the first step, as in our case. We obtained a homology model of human perforin from Swiss-Model Webserver[25] based on mouse perforin as the template, which was minimized by the webserver at the end of the modelling procedure (**Figure 4**). The mouse perforin is deposited in the Protein Data Bank (PDB) with the ID: 3NSJ and a resolution of 2.75Å. There is a 68% sequence identity between the human and the mouse perforin which makes our model highly reliable. The root mean square deviation (RMSD) between two structures was calculated as 2.28, which shows high structural similarity. The model includes the residues 22–552 of the human perforin which actually has 555 amino acids, still fully covering all 3 domains of perforin (MACPF, EGF-like and C2). All residue numbers throughout this study are shown for human perforin. UniProtKB[26] was referred for sequence and domain assessment of the human perforin (ID: P14222).
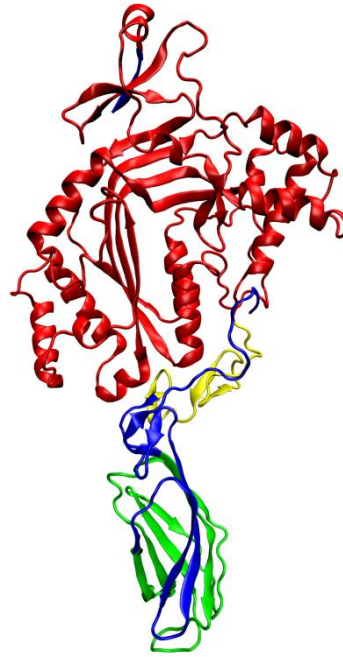
Figure 4: A homology model for human perforin
The model is constructed based on mouse perforin as the template with PDB ID: 3NSJ, 2.75 Å resolution, 68% sequence indentity, 2.28 RMSD. Perforin domains MACPF (349aa), EGF-like (33aa) and C2 (83aa) are colored as red, yellow and green, respectively. The residues in blue show N (top) and C (bottom) termini as well as some loop region.

3.2   Data Collection

In order to see the functional effects of the mutations, we needed as much information as possible about the mutations and the patients. Perforin is known to have more than hundred mutations of all types: missense, nonsense and frameshift caused by insertion or deletion (**Table 3**). Since nonsense and frameshift mutations result in truncated proteins which have no use for structural study, we are interested in only the missense mutations. We collected a total of 76 observed missense perforin mutations from the public mutation databases (HGMD[27], LOVD[28]) and the reported cases in the literature. Since there was more than one mutation for some residues among them, the number of total different mutation sites dropped to 65. Similarly, clinical data of 89 FHL2 patients harboring perforin mutations were manually curated from the literature. The clinical data included Patient/Case ID, Nucleotide Sequence Alteration, Country of Origin, Consanguinity/Family History, Age at Diognasis, Sex, Central Nervous System Involvement, Symptoms, Treatment, BMT Status and its Outcome. The complete list of 76 missense mutations with first-reported references and the

curated clinical data with Pubmed IDs (PMID) are given as **Appendix Table 9** and **Appendix Table 10,** respectively.

Table 3: Mutation types and corresponding number of mutations in perforin as currently given in HGMD
(http://www.hgmd.cf.ac.uk/ac/gene.php?gene=PRF1)

Mutation type	Number of mutations
Missense/nonsense	86
Small deletions	13
Small insertions	4
Total	**103**

3.3 Clustering Mutations

Perforin has a vast number of missense mutations observed at 65 different sites, but not all of them are expected to cause the same structural impact on the protein. Therefore a link between groups of mutations and the functional regions of the perforin might be useful for predicting the structural cause of the resulting perforin deficiency. For this reason, we clustered the mutation sites according to the distances between them on the model structure by using NetworkX package of Python. A distance matrix (65x65) was created from the Cα coordinates of 65 residues where the residues closer to each other than 10Å were marked as 'connected'. Each continuous group of connected residues formed a different cluster. The graph of connected residues was plotted with matplotlib (**Figure 5**).
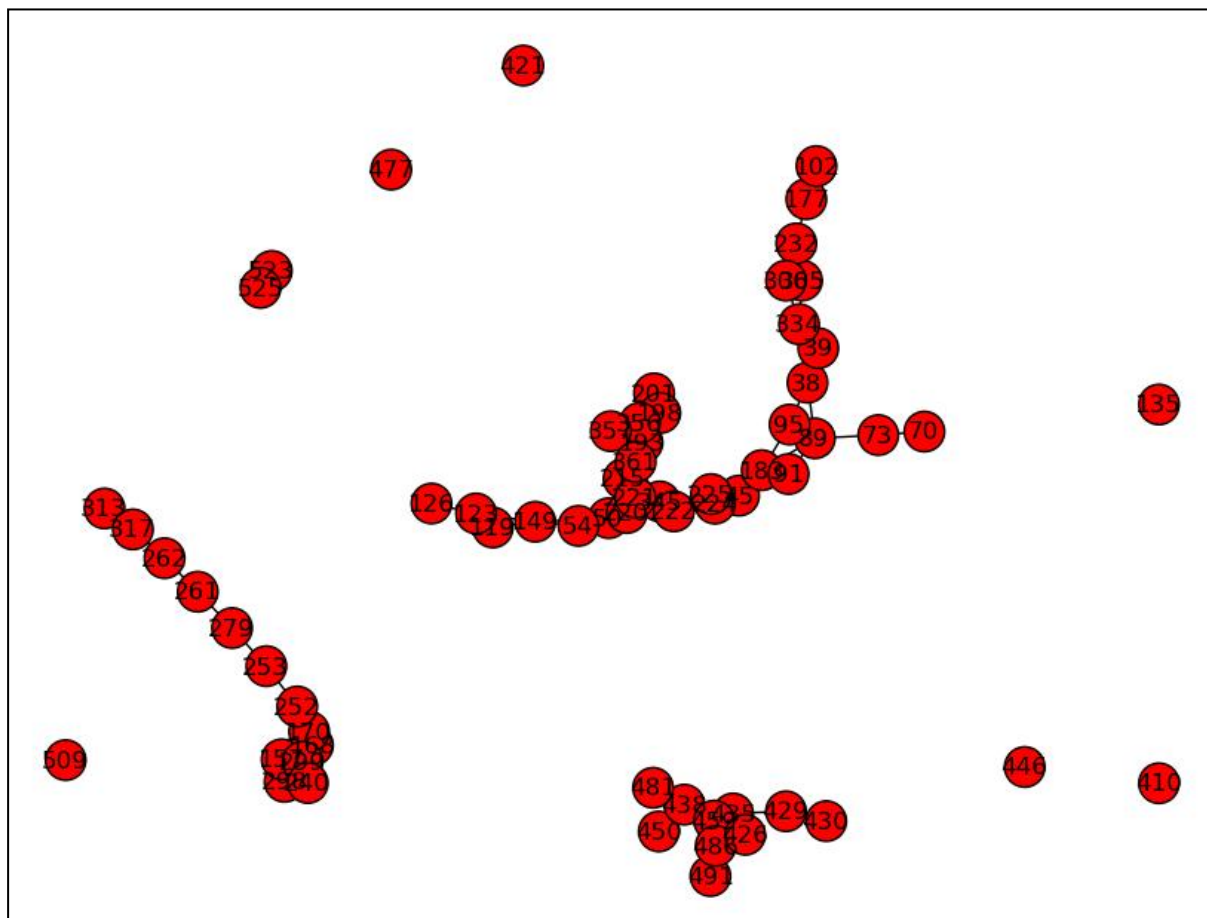
Figure 5: Graph of perforin mutations highlighting 3 main clusters
Numbers on the nodes represent the residue numbers. Edges are drawn between 'connected' mutations where the residues closer to each other than 10Å according to their Cα distances were marked as 'connected'.

3.4   Stability Analysis

The stability of a protein is closely related to its function, and the structural conformation of a protein determines its stability. Mutations are known to alter the stability of the protein structure, hence rendering them non-functional. To have a further insight on the relationship between perforin mutations and its stability, we calculated the change in protein stability upon mutation by <PositionScan> command of FoldX[29] version 3.0 beta 5.1 (http://foldx.crg.es/). Prior to energy calculation, the model structure was minimized using the <RepairPDB> command. Each of 531 residues in the model was mutated to other 20 aminoacids (including itself). Then ΔΔG values were extracted from the FoldX output files. Any mutation resulting in ΔΔG < 0kcal/mol and ΔΔG > 2kcal/mol was considered as

stabilizing and destabilizing, respectively, referring to the calculated correlations of $\Delta\Delta G_{exp}$ and $\Delta\Delta G_{calc}$ values tested on 1088 mutants[30].

3.5   Statistical Analysis

So many mutations dispersed on the protein structure brings the question to mind that if they occupy random sites that any other possible mutation elsewhere would have the same impact on the stability or they rather deploy selective positions where they result in the largest damage, so that perforin function is maximally disrupted. To check the significance of the observed mutations leading to the instability of the protein, we applied two-tailed Student's t-Test in Excel, assuming unequal variance between measured samples. Three samples were created from the $\Delta\Delta G$ table such that:

1. **First sample:** <u>The observed mutation sites</u> with the total number of mutations per site which resulted in total energy changes greater than 2 kcal/mol according to the FoldX calculations (n = 65 residues, mean = 8,91)
2. **Second sample:** <u>The entire sequence</u> with the total number of mutations per site which resulted in total energy changes greater than 2 kcal/mol according to the FoldX calculations (n = 531 residues, mean = 6,36)
3. **Third sample:** <u>The non-mutation sites</u> with the total number of mutations per site which resulted in total energy changes greater than 2 kcal/mol according to the FoldX calculations (n = 466 residues, mean = 6,00)

3.6   Perforin Pre-pore Model

The hallmark of perforin's function is the pore formation at the membrane, so it is vital to understand how this step is interrupted by the mutations where this is the case. Although the mechanism of action of pore formation is poorly understood, it was proposed that perforin monomers undergo a large conformational change during membrane insertion. However, they were observed to have two states in oligomeric form, that are pre-pore and pore states, where the former state assumes rather a rigid complex prior to membrane insertion. A model which

explains the molecular dynamics of the pore at atomic level would serve perfect to understand the pore state, however, due to the time limitation we created a model only for the pre-pore state at macroscopic level.

A pre-pore model was created in VMD[31] from the monomeric homology model as the starting structure. An identical copy of the homology model was loaded to form a homodimer complex (currently on top of each other). The initial and the new-loaded monomer are renamed as left partner and right partner of the complex from now on, respectively. First, the three principal axes of the left partner were drawn originating from its center of mass (**Figure 6-a**). Second, the complex was oriented so that the third principal axes was aligned to the z-axis, hence orthogonality of the complex to the xy plane was achieved (**Figure 6-b**). Third, the right partner of the z-aligned complex was rotated about the z-axis by an arbitrary degree, which was used as a parameter to determine the number of monomers to be included in the pre-pore, and translated along the y-axis by an arbitrary distance, which was used as a parameter to determine the size of the pre-pore (**Figure 6-c,d**). Finally, the right partner of this z-aligned complex was superimposed with the left partner of another loaded complex by using <measure fit> command in VMD, this particular step was repeated until the desired number of monomers came together to obtain the pre-pore model structure bearing cyclic symmetry. The redundant monomers, i.e. all right partners, are finally deleted. In the reference pre-pore model (**Figure 7**), 20 monomers are shown (C20 symmetry) having ~14nm diameter, indeed the size of the pore was observed by cryo-electron microscopy to vary from 10 to 25 nm[32]. More models with different number of monomers and pore sizes are given in **Appendix Table 11**).
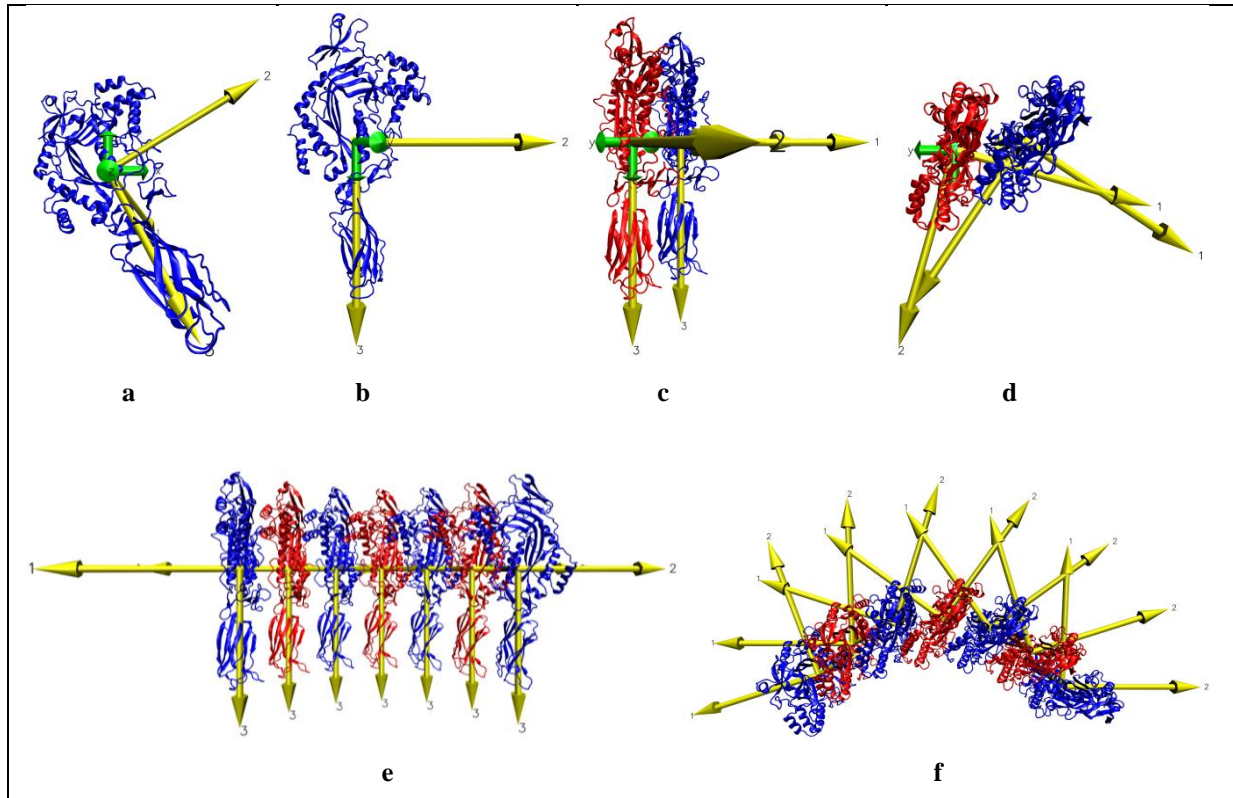
Figure 6: Steps of perforin pre-pore modelling

First row shows the alignment of the homodimer model complex to z-axis along its 3. principal axes. Second row shows the symmetrical growing of the monomers towards a complete pre-pore. **a)** The principal axes are drawn originating from the center of the mass of the protein. **b)** The protein is oriented so that the third principal axis is aligned to the z axis (both pointing downwards). **c)** The second monomer (the red one, initially on top of the blue one) is rotated about the z axis by -18 degree and translated along the y axis by 35Å so that the homodimer model complex is created. **d)** Top view of the aligned homodimer model complex in **c**. **e)** More monomers are added up by superimposing the homodimer model complexes. **f)** Top view of **e**.

3.7   Structural Analysis

The regions where proteins interact with other molecules to form a complex are called interfaces. In most of the cases, proteins exert their function via these interfaces. In agreement with this, perforin monomers come together to have interfaces between each other during oligomerization. Recent experimental evidence already revealed three charged residues that contribute to the oligomerization by creating ionic interactions. On the other hand, mutations occurring at the interfaces are likely to disrupt the protein function much more severely than the ones anywhere else. Therefore interface residues are no doubt very important for protein interactions. In order to see if some of the perforin mutations affect its interaction, we performed an interface analysis.

The interface between perforin monomers for the homodimer model complex (**Figure 9**) was extracted by using Hotpoint Webserver[33] with the default distance threshold, i.e. the sum of van der Waals radii of two atoms + 0.5Å.). For the protein interfaces, some residues are attributed to be more important because of greater contribution to the binding than the others, which are called hot spots. The hot spots between perforin monomers were predicted by the same server (**Figure 10**). The mutations observed at the hot spots are expected to have more dramatic changes for perforin oligomerization. We also calculated the relative surface accessibility of the residues using PSAIA[34] to check the distribution of the mutated residues on the protein structure.

Chapter 4

4   RESULTS and DISCUSSION

4.1   Perforin Mutations

Perforin is a remarkable protein having many lethal mutations mostly spread on its functional domains. In HGMD[27], currently a total of 103 reported mutations are given including missense, nonsense, deletion and insertion types of mutations (**Table 3**). Since the last three types result in truncated proteins that are eventually degraded, which becomes out of the structural scope of this study, here only the missense mutations are included. We analysed 76 missense perforin mutations collected from the public databases and the reported cases in the literature (**Appendix Table 9**). Among these, the majority (58) are found in the large MACPF domain (349 residues) whereas another significant portion (13) reside in the Ca-dependent lipid binding domain C2 (83 residues). However, the total number of different mutation sites is 65, indicating that more than one mutation occurred at some residues.

Perforin mutations obviously have a heterogeneous character. Three types of perforin mutations in terms of genetic form are observed: homozygous, heterozygous and compound heterozygous. Although the significance of homozygous mutations is relatively straightforward to interpret, heterozygous ones contribute to the complexity of the disease, such as late onset cases. Since an important question about perforin mutations is whether they are distinguishable in their resulting effects (i.e. perforin expression, NK cell activity, cytotoxicity, severity, onset of the disease etc.), different genetic forms of the same mutation might be useful to solve it. Indeed, there are few examples in which mutations are found both in homozygous and compound heterozygous form (V50M[20, 35, 36], G149S[21, 35-37], G220S[38, 39], R225W[36, 38, 39], H222Q[21, 40]). Although it is possible to study the structural effects of these mutations separately, the lack of the sufficient knowledge about the underlying mechanisms of perforin function keeps our question alive. Nevertheless our

detailed analysis of the missense perforin mutations reveal important clues about the perforin deficiency in FHL2 patients. More clinical reports about FHL2 patients and functional studies about perforin mutations will further clarify this question, as well as the one that is why some patients show late onset of the disease.

4.2   Perforin Pre-pore Model

The perforin pre-pore model we created is consistent with the previously proposed classical pore models; differently, our model is based on human perforin where the homology model is used, additionally, suggesting flexibility for the size of the pore and the number of monomers involved, which are indeed observed to vary in vivo by cryo-electron microscopy[32, 41, 42] (**Figure 7**). However, perforin undergoes a large conformational change during its re-organization from pre-pore to pore state to insert into the membrane, forming a large amphipathic transmembrane β-barrel[43]. Since the contact regions between monomers are subject to change during this transition, our interface analysis is mainly valid for the pre-pore state. As a result, our functional analysis primarily address the pre-synaptic defects of perforin mutations. Further experimental studies revealing structure information of the complete pore state will help us better understand the post-synaptic defects as well.

side view   top view
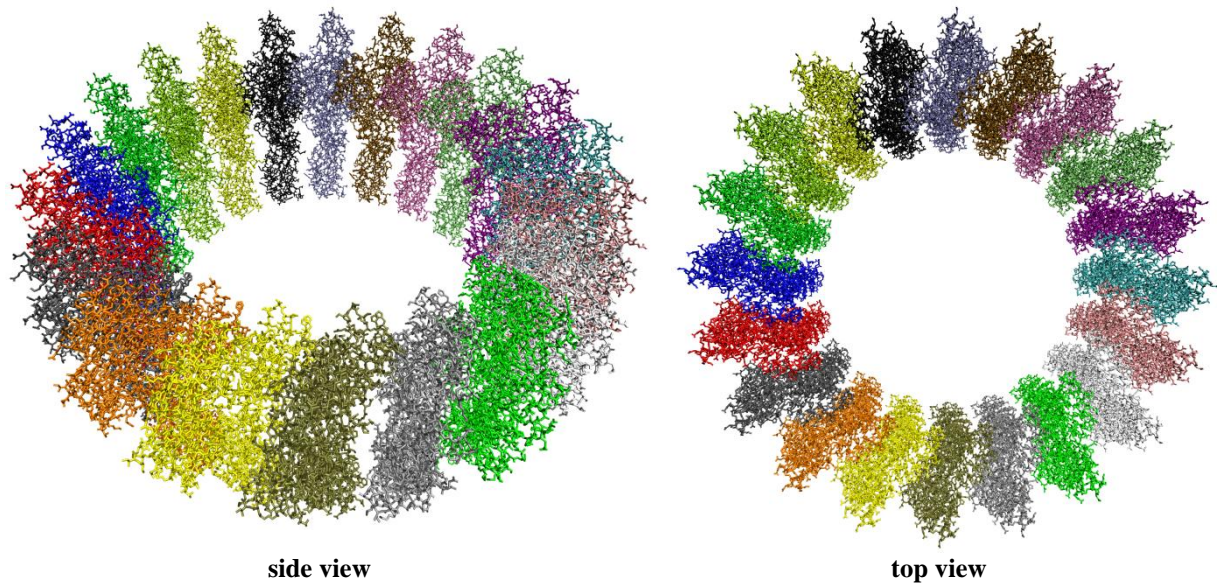
Figure 7: A model for perforin pre-pore
Each monomer is shown in a different color (20 monomers are in the model). The model is consistent with the three experimentals constraints: **1)** Perforin monomers interact via the opposite flat faces of MACPF domain[42]. **2)** Arg214(positive) and Glu344(negative) undergo a direct ionic interaction between two adjacent monomers, as well as Asp192(negative) is important for oligomerization[42]. **3)** The pore is symmetric where orientation of MACPF domain is opposite to that of in CDCs[41] (rotational cyclic symmetry of C20 is shown). More models from C12 to C30 symmetry are given in **Appendix Table 11**.

4.3   Clusters of Mutations

Clustering mutations according to the distances between their Cα atoms resulted in 3 large clusters where each includes at least 2 pairs of 'connected' mutations (**Figure 5**). Cluster 1, Cluster 2 and Cluster 3 consist of 34, 13 and 10 mutations, respectively (**Table 4**). Other than these, one cluster including 1 pair of connected mutations and six clusters including single mutations were found, which were disregarded to be discussed further since the number of mutations included are not sufficient to be associated with a common feature. Cluster 1 and Cluster 2 are located in the MACPF domain whereas Cluster 3 sits at the tip of the C2 domain.

Table 4: Clusters of missense perforin mutations and their functional assignment on the protein structure

Cluster No	Domain	Region/function
Cluster 1	MACPF	Interface of perforin monomers/oligomerization
Cluster 2	MACPF	Small clusters of α helices/membrane-spanning
Cluster 3	C2	Ca binding domain/membrane attachment

Residue numbers in clusters:
Cluster 1: 38, 39, 45, 50, 54, 70, 73, 89, 91, 95, 102, 119, 123, 126, 149, 177, 183, 193, 198, 201, 215, 220, 221, 222, 224, 225, 232, 305, 306, 334, 345, 356, 357, 361
Cluster 2: 157, 168, 170, 240, 252, 253, 261, 262, 279, 298, 299, 313, 317
Cluster 3: 426, 429, 430, 435, 438, 450, 459, 481, 486, 491

Observation of clusters on the protein structure immediately reveals 3 distinct regions where each of them depicts a different functional group of perforin (**Figure 8**). Cluster 1 maps the region where perforin monomers interact with each other during oligomerization, including nearby mutations to residues 192, 214 and 344 which have been previously shown to undergo ionic interactions required for the pore assembly[42]. Cluster 2 maps the region where the two small clusters of α helices (CH1 and CH2) are located that they refold into membrane-spanning amphipathic β strands which is a crucial step for the membrane insertion. Similarly, Cluster 3 maps the region where C2 domain exerts its function, Ca-binding membrane attachment which is a prerequisite for the pore formation.
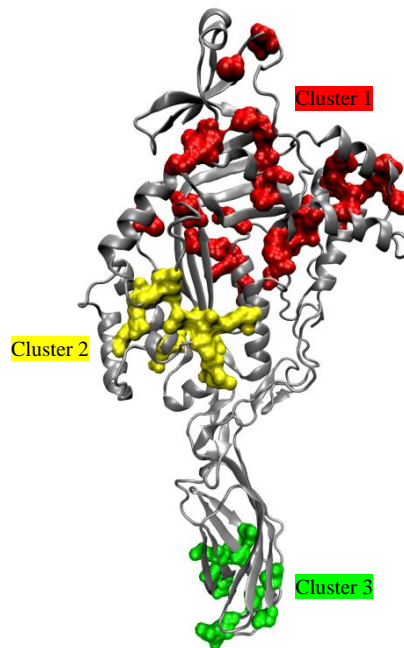
Figure 8: Three main clusters of perforin mutations shown on the model structure

4.4   Stability Analysis

Mutations are often associated with the stability of protein structure, resulting in free energy changes. As a result, the native state of the protein may be totally lost as in the case of nonsense or frameshift mutations, or may be partially altered as in our case of missense mutations. To have further insight on these mutations, we analysed in detail the total energy changes on the perforin caused by the 76 observed missense mutations. To assess the specifity, not only the 76 observed mutations but all the possible missense mutations over the entire sequence (531x20=10620) were computationally studied. A statistical comparison between three samples was performed, where the first sample is the 65 clinically observed different mutation sites with the total number of destabilizing ($\Delta\Delta G > 2$ kcal/mol) mutations per site. The second and the third samples are the entire sequence (n = 531) and the non-mutation sites (n = 466) with the same criteria, respectively. We concluded that observed mutations occupy critical positions for perforin stability, having significantly greater energy changes than that in the random case (p = 0,008 between sample 1 and sample 2, p = 0,002 between sample 1 and sample 3).

Further stability analysis was carried on the types of mutations and the relative accessible surface areas of the corresponding residues. The mutations were classified according to the residue types into 4 groups as charged → charged (8%), charged → neutral (30%), neutral → charged (20%) and neutral → neutral (42%) (**Table 5**). Among these ΔΔG values of the third group (neutral → charged) were particularly high showing that charge contribution of the mutated residues greatly lowers the overall stability of the protein as expected. It also points that the mutations of this group are located at such functionally critical regions of the perforin that addition of charge groups cannot be tolerated. Moreover, individual analysis of the mutations revealed that there is a good inverse correlation between the ΔΔG and the total relative accessible surface area (RASA). The residues having very low RASA resulted in significantly high ΔΔG values, suggesting that the mutations in the core region cause large conformational changes or displacement of charge groups that stabilize the protein (**Table 6**). On the other hand, the majority of the stabilizing mutations have large relative accessible surface areas falling into the 'exposed' category.

Table 5: Classification of the perforin mutations according to the residue types

Type of mutation	Number of mutations (percentage)	Mutations	Comment on total energy changes
Charged → charged	6 (8%)	R123H, H222R, R232H, E253K, E261K, R509K	Generally low
Charged → neutral	23 (30%)	R54C, D70V, D70Y, R119W, R126C, R177C, H222Q, R225P, R225Q, R225W, R232C, R240G, E298G, R299C, D313V, R356W, R357W, R361W, R410P, R410W, D430Y, D486G, D491N	Generally low
Neutral → charged	15 (20%)	P39H, G45R, G45E, C73R, W95R, G149R, S170R, G198R, G220R, A262D, G305D, G317R, G429E, Q446R, A523D	Generally high
Neutral → neutral	32 (42%)	V38L, V38M, V50M, C73Y, P89T, A91V, C102F, V135M, G149S, F157V, S168N, V183G, F193L, P201T, L215I, G220S, T221I, I224N, N252S, C279Y, G306C, G334S, P345L, F421C, G426S, T435M, Y438C, T450M, P459L, P477A, Q481P, C525S,	Generally low
Acidic → neutral	7	D70V, D70Y, E298G, D313V, D430Y, D486G, D491N	
Acidic → basic	2	E253K, E261K	

Acidic → acidic	0	none	
Basic → neutral	16	R54C, R119W, R126C, R177C, H222Q, R225P, R225Q, R225W, R232C, R240G, R299C, R356W, R357W, R361W, R410P, R410W	
Basic → acidic	0	none	
Basic → basic	4	R123H, H222R, R232H, R509K	

charged residues : E,D,R,K,H
neutral residues : A,C,F,G,I,L,M,N,P,Q,S,T,V,W,Y
acidic residues   : E,D
basic residues   : R,K,H

Table 6: The correlation between total RASA and ΔΔG of perforin mutations
Numbers show the total number of mutations for a given RASA and ΔΔG range.

		ΔΔG (in kcal/mol)			
		< 0	**between 0 and 2**	**> 2**	**Total**
Total RASA	**Buried (< 5%)**	1*	4	21	**26**
	Intermediate exposure (between 5% and 30%)	2	14	15	**31**
	Exposed (> 30%)	8	9	2	**19**
	Total	**11**	**27**	**38**	**76**

*: A91V change (a possible polymorphism)

Mutations are usually expected to destabilize the protein structure. Among the 76 missense mutations we studied, half of them fall into this category having a total energy change greater than 2 kcal/mol, yet other 27 have less than 2 kcal/mol but still positive values. However, a group of 11 mutations resulted in negative numbers, which by definition shows that these mutations contribute to the stability of the protein. Two more webservers (CUPSAT[44], PoPMuSiC[45]) were used to eliminate the dependency on a single energy function, which gave consistent results. One mutation with the greatest negative energy change among the last group was remarkable, A91V, which is further discussed in a later section.

4.5   Interface Analysis

Perforin monomers have a large interface along two alpha helices protruding from the MACPF domain through which they oligomerize (**Figure 9**). Based on the distance criteria 38 interface residues were predicted at this region where 6 of them are the mutation sites on either face of the MACPF domain (**Table 7**). Among 16 predicted hot spots (the residues

which account for the majority of the binding energy at the interfaces[46]) out of 38 interface residues (**Figure 10**), 3 of them are the mutation sites, which might show that the absent or reduced cytotoxicity in particular patients carrying those mutations is due to the disruption of perforin oligomerization by them. Corroboratively, in a patient with homozygous R225W mutation (one of our predicted hot spot residues), significantly reduced natural killer cell function was reported although perforin expression was at its normal range (91%). However, more experimental setups are required to verify the characteristics of interactions between individual residues at the interface.
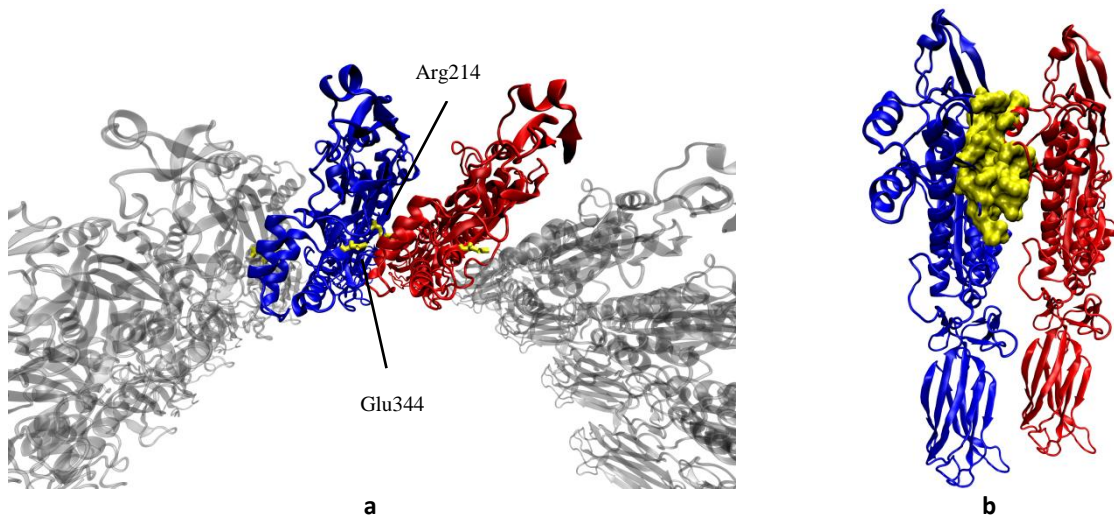
Figure 9: Interface between perforin monomers
a) The ionic interaction between Arg214 and Glu344[42] shown on our model. **b)** The interface residues are highlighted in yellow that are predicted by Hotpoint Webserver[33].

Table 7: Predicted interface residues and hot spots between perforin monomers
Residues 192, 214 and 344 were experimentally shown to be at the interface in a previous study[1]. Hotspots are highlighted in red and observed missense mutations are denoted with the mark(*).

Residue Number	Residue Name	Chain	Prediction
191	P	X	NH
192	D	X	NH
195	R	X	NH
198*	G	X	NH
199	D	X	H
201*	P	X	NH
205	N	X	NH

206	A	X	NH
207	S	X	H
208	T	X	NH
209	Q	X	NH
210	P	X	NH
211	A	X	H
212	Y	X	H
213	L	X	NH
214	R	X	H
215*	L	X	H
218	N	X	NH
93	T	Y	NH
94	N	Y	NH
124	S	Y	NH
125	I	Y	H
126*	R	Y	NH
127	N	Y	H
128	D	Y	H
129	W	Y	H
130	K	Y	NH
131	V	Y	H
132	G	Y	NH
133	L	Y	H
135*	V	Y	H
136	T	Y	NH
182	H	Y	H
185	H	Y	NH
225*	R	Y	H
341	Y	Y	H
342	T	Y	NH
344	E	Y	NH

[1]   :Baran, K., Dunstone, M., Chia, J., Ciccone, A., Browne, K. A., Clarke, C. J., Lukoyanova, N., Saibil, H., Whisstock, J. C., Voskoboinik, I. & Trapani, J. A. (2009). The molecular basis for perforin oligomerization and transmembrane pore assembly. Immunity 30, 684–95.

Proteins interact with each other via interface residues. Having a large interface between perforin monomers during oligomerization prior to the pore formation may help explain the fatal consequences of perforin mutations, particulary those which are found at or nearby the interface (**Cluster 1**). As a matter of fact, it was previously shown that charge

reversal or substitution to different type of residues of three interface residues dramatically decreased cytotoxicity (D192K, R214E, E344R)[42].

chain X   chain Y

Figure 10: The predicted hot spots and interface between perforin monomers for the homodimer model complex

Interface residues on chain X   Interface residues on chain Y
Predicted hot spots in interface (chain X)   Predicted hot spots in interface (chain Y)

4.6   Structural Analysis

The structural effects of mutations are strongly associated with their location on protein structure. Mutations which occur on the surface of the protein may disrupt the binding site whereas those occur in the core region distort the protein conformation. We classified the mutated residues in perforin as buried, exposed and residues with intermediate exposure according to their relative solvent-accessible surface areas (**Table 8**). Our analysis show that perforin mutations are uniformly distributed to the surface and core regions of the protein.

Table 8: The number of mutations in three groups classified according to relative accessible surface area

The numbers in brackets show the total number of residues with the given RASA thresholds for each group. <5%:buried, between 5% and 30%:intermediate exposure, >30%:exposed

	RASA*		
	<5%	**between 5% and 30%**	**>30%**
Number of mutations*	20 (113)	28 (184)	17 (234)

* Relative accessible-surface area. RASA values for each mutated residue are given in **Appendix Table 12**.

Mapping of the mutations onto the protein structure shows valuable information about the potential characteristics of the mutations. Perforin is functional only in the presence of high concentration of Ca^{+2} ions which binds to the C2 domain thereby initiate the membrane attachment. It was shown that mutagenesis of a number of conserved aspartate residues at this region including two of the mutated residues in our **Cluster 3** (D430A and D486A) caused total loss of cytotoxic activity of perforin[11]. In the same study, two more mutations in this cluster (G429E and T435M) that are close to critical aspartate residues including above two for Ca binding were characterized and the former was found to significantly reduce the cytotoxicity whereas the latter, surprisingly, had no effect on perforin activity, as opposed to the prediction in its original report[47]. Hence according to this case, mutations occurring nearby the functionally important residues may or may not disrupt the overall perforin function, whereas mutations occurring at the critical residues themselves do so.

4.7   Patients Analysis

Analysis of the clinical data regarding the patients reveal important remarks on the characteristics of the FHL2 disease. That the specific mutations are strongly associated with certain ethnic groups[48] was confirmed once again in a larger cohort in our study. Accordingly, W374X nonsense mutation is most common in Turkish population, L17Xfs22 might be a founder mutation in African Americans, whereas L364fsX is confined to Japanese origin (**Table 10**).

A case study is investigated below:

4.7.1   A Case Study:

Molleran Lee *et al.* and Kogawa *et al.*[36, 37] reported a three years old female patient carrying compound heterozygous mutations for perforin, G45R and R54C. The father had the former mutation in additional to A91V and the mother had the latter mutation. The father had 4% NK cell function, whereas mother had 14%, both having the normal range which is above 3,2%. On the other hand, the father had decreased perforin activity in NK cells (68%, the normal range is above 81%) whereas the perforin activity in CD8$^+$ T and CD56$^+$ T cells were at its normal range. The situation in the mother was vice versa for perforin activity, which was at its normal range in NK cells (95%) whereas decreased in CD8$^+$ T and CD56$^+$ T cells. However, the parents were both healthy. This can be explained for the mother easily since she had only single mutated allele (heterozygous mutation), so she could still produce functional perforin in sufficient amount via the second wild type allele. The father, on the other hand, had both alleles with mutation (compound heterozygosity), but it should be noted that one of these is A91V which is considered as a polymorphism in many cases (discussed in detail in **4.8**). So it behaved as a healthy allele in the father, or at least it did not affect his survival. Unlike the parents, the child was diagnosed as having FHL2 due to the both alleles inherited with a pathogenic mutation (compound heterozygosity). As a result, the child had decreased perforin activity in all immune cell types. Although it was extremely unique to see that she had quite high perforin activity in NK cells (61%) which is unusul for FHL2 patients, it was still unsufficient to cope with infections. With the given information so far, it can be said that NK function and perforin activity are the direct indicators of the disease state, provided with the characteristics of the mutations, so are among the diagnostic criteria.

From the structural point of view used in this study, the particular mutations G45R, R54C and A91V are all within our **Cluster 1**, which maps the interface region of perforin monomers for the oligomerization. Interestingly, there are remarkable differences between the individual characteristics of these three mutations. Both G45R and R54C cause change of charge at the specified residue (G45R → change from neutral aa to charged aa and R54S →

change from charged aa to neutral aa) whereas A91V does not (neutral aa to neutral aa change) (**Table 5**). G45R has a very high $\Delta\Delta G$ value (10,24 kcal/mol) which probably destabilizes the protein structure, and R54C has a very high RASA (50,243%) which probably affects the interaction of the protein at the surface; whereas A91V has both negative $\Delta\Delta G$ value (-1,07 kcal/mol, stabilizing) and very low RASA (far from the interface) (**Table 12**). All together, our findings on the structural characteristics of the mutations support the disease state of the child and the healthy state of the parents, consistent with the clinical data.

4.8   A Special Case: A91V

FHL2 patients carrying A91V change show variety in phenotype which makes it difficult to determine if it is a neutral polymorphism or a pathogenic mutation. Previously it was reported that A91V change was found in healthy populations with high allele frequency presenting no signs of the disease[49]. Later it was shown that A91V change causes decreased level of perforin expression and partial loss of lytic capacity[39]. Some other studies revealed late onset cases associated with it[35, 50]. Recently an experimental study demonstrated that A91V change causes conformational change and reduces the production of the active perforin, resulting in defective cytotoxic function[51]. Moreover, A91V change was observed in a FHL2 patient together with tuberculosis, suggesting its contribution to the pathophysiology and possible synergistic role in the late onset of FHL2[52]. All of these diverse findings bring A91V change to the focus of perforin mutations, which actually might be important to understand the structure-phenotype relationship between perforin and FHL2 if studied in detail. In our stability analysis, we found that A91V change decreased the total energy of the protein structure significantly ($\Delta\Delta G$ = -1,07 kcal/mol) although it was buried in the core region of the MACPF domain (RASA = 2,994%). Therefore A91V change particularly stabilizes the protein structure. Furthermore, the conserved Alanine residue at position 91 does not directly take place at the interface of perforin monomers in our pre-pore model, also is far from the small clusters of membrane-spanning α helices. In addition, both having small neutral side chains, change of Alanine to Valine might be tolerated for the sake of chemical resemblance. Hence our results support that A91V change is rather a polymorphism or a 'milder' class mutation[53] from the structural point of view. However, severe pathogenic

cases of A91V change primarily suggests that pathogenecity might arise from the second allele (in compound heterozygous cases). Alternatively, other factors such as epigenetic regulation[54] or temperature sensitivity[55] might be involved which makes the link between the mutation and the perforin deficiency more complex than the assumption of a direct cause-and-effect relationship, such as differential levels of protein expression or cytotoxicity. Moreover, the presence of a second genetic aberration at the upstream or downstream proteins of the secretory granule-dependent pathway that are not scanned for the given patient might result in a phenotype same as in the FHL2 that creates ambiguity in the interpretence of the disease, since still approximately 10% of FHL cases could not be associated with a specific protein. A network-based study may be helpful to reveal these relationships.

Chapter 5

5   CONCLUSION

Perforin has been a mystery for decades after its role in cytotoxicity has been first described in 1985[56, 57]. Nevertheless since then it was well known that perforin acts via pore formation in synergy with granzyme B, however, the exact mechanism of the pore formation remained elusive because of the lack of sufficient evidence. For decades, an analogy was made between perforin and CDCs. Recently, the determination of the X-ray crystal structure of the murine perforin elucidated the mechanism of perforin pore formation. Nevertheless, the number of functional studies on perforin mutations is very few. In this study we investigated the human perforin (homology model) with a structural perspective, redounding the understanding of severe consequences of perforin deficiency due to mutations observed on FHL2 patients. Up to now, numerous clinical studies have reported the clinical manifestations of FHL2 patients carrying a variety of perforin mutations (**Appendix Table 10**). On the other hand, several studies addressed the structure of homologous partners of perforin or MACPF domain[58-60], where a huge step was taken by the determination of the crystal structure of monomeric murine perforin. Our study combines these two methods, providing a comprehensive analysis of perforin mutations from the structural point of view.

It was interesting to see that although there are many mutation sites observed on the perforin, they are not just randomly distributed. Rather they reside in where mutations change the protein stability significantly when compared to non-mutation sites. This may help explain the lack of cytotoxicity in FHL2 patients, because the perforin monomers may not oligomerize properly due to the conformational changes introduced by the mutations to form pore. Although the exact mechanism is unknown, pore formation is required for perforin function, which bears the cytotoxicity. Furthermore, cluster analysis showed that the distribution of mutations match with its functional regions, which further confirmes our conclusion.

The cause-and-effect relationship between perforin mutations and FHL2 disease may be ultimately solved. However, a more important question from the evolutionary point of view is why such a vital conserved protein has so many lethal mutations. Although was among one of the goals of this study, unfortunately, our results are not sufficient to answer this question.

The increasing number of the reported mutations and the clinical data about FHL2 is going to help better understand the genetic basis of the disease undoubtedly. Similarly, the determination of the crystal structure of human perforin will be a milestone for perforin biology and FHL2 disease. Additionally, molecular dynamics studies of the complete pore structure have potential to clarify the mechanism of pore formation. Finally, more experimental setups which are designed to study the individual mutations and molecular interactions will significantly contribute to the functional basis of perforin. Taken together, perforin and FHL2 seem to remain to be a hot topic in immunology in close future.

APPENDIX

List of Mutations

Table 9: List of 76 missense perforin mutations analysed in this study

Nucleotide change	Amino acid change	Residue number	Phenotype*	Reference
cGTG-ATG	Val-Met	38	FHL2	Zur Stadt (2006) Hum Mutat 27, 62
	Val-Leu	38	FHL2	Horne (2008) Br J Haematol 143(1):75–83
CCT-CAT	Pro-His	39	FHL2	Kogawa (2002) Blood 99, 61
cGGG-AGG	Gly-Arg	45	FHL2	Kogawa (2002) Blood 99, 61
GGG-GAG	Gly-Glu	45	FHL2	Molleran Lee (2004) J Med Genet 41, 137
cGTG-ATG	Val-Met	50	FHL2	Goransdotter Ericson (2001) Am J Hum Genet 68, 590
cCGC-TGC	Arg-Cys	54	FHL2	Molleran Lee (2004) J Med Genet 41, 137
cGAC-TAC	Asp-Tyr	70	FHL2	Molleran Lee (2004) J Med Genet 41, 137
GAC-GTC	Asp-Val	70	FHL2	Trizzino (2008) J Med Genet 45, 15
cTGC-CGC	Cys-Arg	73	FHL2	Molleran Lee (2004) J Med Genet 41, 137
	Cys-Tyr	73	FHL2	Trizzino (2008) J Med Genet 45, 15
gCCT-ACT	Pro-Thr	89	FHL2	Al-Lamki (2003) Pediatr Hematol Oncol 20, 603
GCG-GTG	Ala-Val	91	Childhood acute lymphoblastic leukemia, association with	Santoro (2005) Haematologica 90, 697
cTGG-CGG	Trp-Arg	95	FHL2	Clementi (2001) J Med Genet 38, 643
c.305G>T	Cys-Phe	102	EBV-HLH CAEBV	Lu G. (2009) Chin Med J 122(23):2851–5
cCGG-TGG	Arg-Trp	119	Perforin deficiency ?	Moshous (2007) Arthritis Rheum 56, 995
CGT-CAT	Arg-His	123	Anaplastic	Cannella (2007)

			large cell lymphoma	Cancer 109, 2566
c376C>T	Arg-Cys	126	FHL2 polymorphism	LOVD
	Val-Met	135	FHL2	SM2PH-kb Knowledgebase
cGGC-AGC	Gly-Ser	149	FHL2	Kogawa (2002) Blood 99, 61
cGGC-CGC	Gly-Arg	149	FHL2	Trizzino (2008) J Med Genet 45, 15
cTTT-GTT	Phe-Val	157	FHL2	Molleran Lee (2004) J Med Genet 41, 137
c.503G>A	Ser-Asn	168	NK cell lymphoma	Lu G. (2009) Chin Med J 122(23):2851–5
cAGC-CGC	Ser-Arg	170	FHL2	Turtzo (2007) J Child Neurol 22, 863
cCGC-TGC	Arg-Cys	177	Anaplastic large cell lymphoma	Cannella (2007) Cancer 109, 2566
GTG-GGG	Val-Gly	183	FHL2	Stepp (1999) Science 286, 1957
cTTC-CTC	Phe-Leu	193	Chronic active Epstein-Barr virus infection	Katano (2004) Blood 103, 1244
c592G>A	Gly-Arg	198	B-ALL, Ph+	Liyun Yang (2011) Leuk Res. 35(2):196–199
gCCC-ACC	Pro-Thr	201	FHL2	Zur Stadt (2006) Hum Mutat 27, 62
gCTT-ATT	Leu-Ile	215	FHL2	Astigarraga (2004) Pediatr Neurol 30, 361
cGGC-AGC	Gly-Ser	220	FHL2	Clementi (2001) J Med Genet 38, 643
cGGC-CGC	Gly-Arg	220	FHL2	Spisek (2006) Cas Lek Cesk 145, 50
ACC-ATC	Thr-Ile	221	FHL2	Clementi (2001) J Med Genet 38, 643
CAC-CGC	His-Arg	222	FHL2	Molleran Lee (2004) J Med Genet 41, 137
CACt-CAA	His-Gln	222	FHL2	Molleran Lee (2004) J Med Genet 41, 137
ATC-AAC	Ile-Asn	224	FHL2	Goransdotter Ericson (2001) Am J Hum Genet 68, 590
cCGG-TGG	Arg-Trp	225	FHL2	Stepp (1999) Science 286, 1957
CGG-CCG	Arg-Pro	225	FHL2	Muralitharan (2007) Am J Hematol 82, 1099
c674G>A	Arg-Gln	225	B-ALL, Ph+	Liyun Yang (2011) Leuk Res. 35(2):196–199
cCGC-TGC	Arg-Cys	232	FHL2	Clementi (2001) J Med Genet 38, 643

CGC-CAC	Arg-His	232	FHL2	Feldmann (2002) Br J Haematol 117, 965
gCGC-GGC	Arg-Gly	240	FHL2	Zur Stadt (2006) Hum Mutat 27, 62
AAC-AGC	Asn-Ser	252	FHL2 ?	Stepp (1999) Science 286, 1957
cGAG-AAG	Glu-Lys	253	FHL2	Kobayashi (2007) J Pediatr Hematol Oncol 29, 178
cGAG-AAG	Glu-Lys	261	FHL2	Feldmann (2002) Br J Haematol 117, 965
GCC-GAC	Ala-Asp	262	FHL2	Astigarraga (2004) Pediatr Neurol 30, 361
TGT-TAT	Cys-Tyr	279	FHL2	Stepp (1999) Science 286, 1957
GAG-GGG	Glu-Gly	298	FHL2	Trizzino (2008) J Med Genet 45, 15
gCGC-TGC	Arg-Cys	299	FHL2	Molleran Lee (2004) J Med Genet 41, 137
c914G>A	Gly-Arg	305	Lymphoma	Clementi (2005) Blood 105(11):4424–8
cGGC-TGC	Gly-Cys	306	FHL2	Grossman (2005) Blood 106, 1203
GAC-GTC	Asp-Val	313	FHL2	Molleran Lee (2004) J Med Genet 41, 137
cGGG-AGG	Gly-Arg	317	FHL2	Ueda (2003) Br J Haematol 121, 503
c1000G>A	Gly-Ser	334	FHL2	LOVD
CCC-CTC	Pro-Leu	345	FHL2	Stepp (1999) Science 286, 1957
gCGG-TGG	Arg-Trp	356	FHL2	Trizzino (2008) J Med Genet 45, 15
c1069C>T	Arg-Trp	357	FHL2	LOVD
gAGG-TGG	Arg-Trp	361	FHL2	Molleran Lee (2004) J Med Genet 41, 137
CGG-CCG	Arg-Pro	410	Chronic active Epstein-Barr virus infection	Katano (2004) Blood 103, 1244
tCGG-TGG	Arg-Trp	410	FHL2	Ueda (2007) Am J Hematol 82, 427
TTC-TGC	Phe-Cys	421	Anaplastic large cell lymphoma	Cannella (2007) Cancer 109, 2566
c1276G>A	Gly-Ser	426	FHL2 polymorphism	LOVD
GGG-GAG	Gly-Glu	429	FHL2	Stepp (1999) Science 286, 1957
c1288G>T	Asp-Tyr	430	FHL2	Yenan (2007) Blood 110(6):1906–15
ACG-ATG	Thr-Met	435	FHL2	McCormick (2003) Am J Med Genet 117A, 255
c1313A>G	Tyr-Cys	438	FHL2	LOVD

CAG-CGG	Gln-Arg	446	FHL2	Trizzino (2008) J Med Genet 45, 15
ACG-ATG	Thr-Met	450	FHL2	Ueda (2007) Am J Hematol 82, 427
1376C>T	Pro-Leu	459	FHL2	Feldmann (2005) Blood 105(7):2658–63
c1429C>G	Pro-Ala	477	FHL2	Orilieri (2008) Diabetes 57(4):1078–83
CAG-CCG	Gln-Pro	481	FHL2	Molleran Lee (2004) J Med Genet 41, 137
c1457A>G	Asp-Gly	486	B-ALL, Ph-	Liyun Yang (2011) Leuk Res. 35(2):196–199
cGAT-AAT	Asp-Asn	491	Anaplastic large cell lymphoma	Cannella (2007) Cancer 109, 2566
c1526G>A	Arg-Lys	509	B-ALL, Ph+	Liyun Yang (2011) Leuk Res. 35(2):196–199
1568C>A	Ala-Asp	523	FHL2	Hamza Okur (2008) Leuk Res. 32(6):972–5
TGC-TCC	Cys-Ser	525	FHL2	Trizzino (2008) J Med Genet 45, 15

* FHL2: Familial hemophagocytic lymphohistiocytosis type 2,
? : unclear

Clinical Data

Table 10: Clinical data collection of 89 FHL2 patients harboring perforin mutations reported in the literature

Mutation*	Patient/case ID as in the Reference	Nucleotide Sequence Alteration	Country of Origin	Consanguinity/Family History	Age at Diagnosis/Sex	CNS	F	S	C	H	H/H	Treatment	BMT Status/Outcome/Survival after diagnosis	Reference (PMID)
Trp374Stop	1	1122 G→A	Turkey	+/+	2m/M	-	+	+	+	+	+	Vp/Cs/CsA	No BMT/dead	11179007
	7	1122 G→A	Turkey	-/+	1m/F		+	+	+		+	Vp/Cs	No BMT/dead	11179007
	11	1122 G→A	Turkey	+/+	39m/M		+	+	+	+	+	HLH–94	BMT/alive and well	11179007
	17	1122 G→A	Turkey	+/+	10m/F		+	+	+	+	+	No treatment	No BMT/dead	11179007
	Case 1		Turkey	1st/-	7d/M	-						No treatment	2d	18190960
	Case 2		Turkey	2nd/+	4m/M	-						Steroid	7d	18190960
	Case 3		Turkey	1st/+	6m/F	+						HLH–94	1m	18190960
	Patient 21	1122 G→A	Turkey	+/+										12060139
	Patient 34	1122 G→A	Turkey	+/+										12060139
	Patient 62	1123 G→A	Turkey	+/+										12060139
	Case 7	1123 G→A	Turkey	+/+	3m/M	-	+	+	+				BMT/alive	11565555
	Case 8	1123 G→A	Turkey	+/+	2,5m/F	+	+	+	+				BMT/alive	11565555
	Case 10	1123 G→A	Turkey	+/-	2m/M	-	+	+	+				BMT/alive	11565555
Ile224Asp	3	671 T→A	Sweden	+/-	58m/F		+	+	+	+	+	Vp/Cs/CsA	BMT/alive, mild retardation	11179007
Tyr219Stop/Val50Met	9	657 C→A/148 G→A	Turkey	-/-	4m/M		+	+	+	+	+	HLH–94	No BMT/dead	11179007
Lys285Del	33	853-855 del	Turkey	-/-	3m/F		+	+	+	+	+	No treatment	No BMT/dead	11179007

		AAG									
Gly149Ser	Case 4		Turkey	2nd/+	26d/M	+			HLH–94	10m	18190960
	Case 5		Turkey	3rd/+	14m/F	+			HLH–2004	8m	18190960
Val50Met	Case 6		Turkey	1st/-	16y/M	+			HLH–2004	Alive	18190960
Ala91Val	Case 7		Turkey	1st/-	13y/F	+			HLH–2004	Alive	18190960
	Case 8		Turkey	3rd/+	12y/M	-			HLH–2004	1,5y	18190960
	II-2	272 C→T	Spain	-/	49y/M	-	+	+			16956828
Ala523Asp	Case 9		Turkey	1st/+	36m/F	+			HLH–2004	2m	18190960
Pro89Thr	Family 1-Case A	265 C→A	Oman	+/+	1m/F	-	+ +	+ +	HLH–94	BMT/alive	14578030
	Family 7-Case A	265 C→A	Oman	+/-	2m/F	-	+ +	+ +	HLH–94	BMT/alive	14578030
Gly45Arg/Arg54Ser	P10	133 G→A/160 C→T	Hispanic		3y/F						11756153 , 14757862
Pro39His/Gly149Ser	P1	116 C→A/445 G→A	White		10y/F						11756153 , 14757862
Ser150Stop/ Arg225Trp	P12	449 C→A/673 C→T	White		5y/F						11756153 , 14757862
Leu17Stop	P9	50delT	African/American		3m/F						11756153 , 14757862
	P11	50delT	African/American		8m/M						11756153 , 14757862
	Patient 4	50delT	Algeria	+/+							12060139
	Patient 92	50delT	Haiti	+/+							12060139
	P21	50delT	African/American		2m/M						14757862
	P30	50delT	African/American		2m/M						14757862
	P59	50delT	African/American		3m/M						14757862
	P26	50delT	African/American		6m/F						14757862

	P31	50delT	African/American		6m/F									14757862
	P56	50delT	African/American											14757862
	Case 1	50delT	Ghana	+/-	2m/M	-	+	+		+			BMT/alive	11565555
Val38Leu/???			Middle East											18710388
Glu46Stop/Ala91Val+Arg119Trp	4-year-old girl		Russia (German descent)	-/-	4y/F							HLH–94	BMT/alive	17328077
Tyr450Met	Case 1	1349 C→T	Japan		7y/F	+	+	+	+	+	+	HLH–94,CHOP	BMT/alive	12716377 , 15632205 , 17266056
Met1Val/Gly317Arg	Case 2	1 A→G/949 G→A	Japan	/+	11y/F	+	+	+	+	+	-	HLH–94	No BMT/alive	12716377 , 15632205 , 17266056
Arg410Trp/Leu364fs	Case 3	1228 C→T/1090.91 delCT	Japan		12/F	-	+	+	+	+	+	PSL, CsA		12716377 , 15632205 , 17266056
Met1I/Asp62fsX12	Patient 2	3 G→A/185_195del11	Germany	-/	2m/									16278825
Val38Met/Ala91Val	Patient 19	112 G→A/272 C→T	Morocco	+/	6m/									16278825
Gly149Ser/Ala262fsX22	Patient 3	445 G→A/786_801del16	Germany	-/	2m/									16278825
Gly149Ser	Patient 16	445 G→A	Turkey	+/	3m/									16278825
	P50	445 G→A	Hispanic		6m/M									14757862
Pro201Thr/Lys285Del	Patient 17	601 C→A/853_855delAAG	Turkey	-/	10y/									16278825
His222Gln/Arg232His	Patient 18	666 C→A/695 G→A	Italy	-/	>10y/									16278825

Arg240Gly	Patient 20	718 C→G	Greece	+/	12m/								16278825
Leu526fsX87	Patient 1	1576delT	Germany	+/	10m/								16278825
Arg225Trp/ Gly429Glu	Patient 5	673 C→T/1286 G→A	Africa	-/+									12060139
Pro345Leu	Patient 6	1034 C→T	Africa	+/+									12060139
Cys279Tyr/Val183Gly	Patient 11	836 G→A/548 T→G	France	-/-									12060139
Asn252Ser/?	Patient 25	755 A→G/?	Italy	-/+									12060139
Arg232His/?	Patient 27	695 G→A/?	Italy	-/+									12060139
Gln64Stop	Patient 29	190 C→T	Algeria	+/-									12060139
Met1Ile	Patient 30	3 G→A	Lebanon	+/+									12060139
Gln261Lys	Patient 82	781 G→A	Algeria	+/-									12060139
Gly220Ser/Leu17Stop	Patient 95	658 G→A/50delT	Algeria	-/-									12060139
Trp94Arg	Case 2	283 T→C	Italy	+/-	3m/M	-	+	+	+			BMT/alive	11565555
Tyr219Stop	Case 3	657 C→A	Italy	+/+	1,5m/M	-	+	+	+			BMT/alive	11565555
Gly220Ser	Case 4	658 G→A	Italy	+/-	1,5m/F	+	+	+	+			dead	11565555
Thr221Ile/Arg225Trp	Case 5	662 C→T/673 C→T	Italy	-/-	6m/F	+	+	+	+			dead	11565555
Arg232Cys/Gly394Fs	Case 6	694 C→T/1182insT	Italy	-/-	6y/M	-	+	+	+			dead	11565555
Phe193Leu/Arg410Pro		577 T→C/1229 G→C	Caucasian	/-	7y/M	-	+	+	+	+	+		14576041
Val50Met	P35	148 G→A	Russia		7y/M								14757862
Met1Ile/Asp70Tyr	P58	3 G→A/208 G→T	White		8y/M								14757862
His222Arg/	P22	665 A→G	White/Asian		1m/M								14757862

Phe157Val/fs	P23	469 T→G/1628ins T	White		2m/F									14757862
Leu17Stop/Gln481Pro	P33	50delT/1442 A→C	Hispanic		2m/F									14757862
Leu17Stop/Cys73Arg	P32	50delT/217 T→C	African/American		4m/M									14757862
Leu17Stop/Gly45Glu	P34	50delT/134 G→A	African/American		8m/F									14757862
His222Gln/fs	P28	666 C→A/1636del C	White		3m/M									14757862
Arg225Trp	P57	673 C→T	Filipino		2y/M									14757862
Gly149Ser/Arg299Cys	P25	445 G→A/895 C→T	Portugal		2y/M									14757862
Gly149Ser/Arg361Trp	P24	445 G→A/1081 A→T	Hispanic		4y/F									14757862
Arg225Trp	P29	673 C→T	White		5y/F									14757862
Met1Ile/Asp313Val	P27	3 G→A/938 A→T	White											14757862
Ala91Val-Arg231His/Ala91Val	Proband(twin)	272 C→T-695 G→A/272 C→T	?	/-	13/F	-	+	+	+	+	+			14739222
Ala91Val/Trp374Stop	sibling 1	272 C→T/1122 G→A	Italy		27y/M	+	+	+	+	+		No BMT/alive		12229880
	sibling 2	272 C→T/1122 G→A	Italy		25,5/F	+	+	+	+	+		BMT/alive		12229880
Pro459Leu	Patient 1	1376 C→T	Algeria	+/-	9m/M	+	-	+		-		dead		15598808
	Patient 2	1376 C→T	Algeria	+/-	33m/M	+	+	+	+		+	dead		15598808
Thr435Met	Patient W.	1304 C→T												12599189

| **His222Gln** | Patient 1 | 666 C→A | Netherlands | +/ | 2m/F | + | + | + | + | + | SCT/alive and well | 17525286 |
| **Glu317Arg/Asp430Tyr** | Patient 2 | 949 G→A/1288 G→T | Sweden | | 0m/M | + | + | + | + | + | dead | 17525286 |

*   : if single mutation is shown, it is homozygous; otherwise mutations are compound heterozygous

+   : finding is present
-          : finding is not present
M   : male
F   : female
d   : day
m   : month
y   : year
fs   : frameshift
ins   : insertion
BMT   : bone marrow transplantation
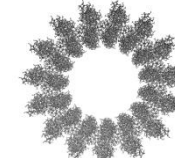SCT   : stem cell transplantation
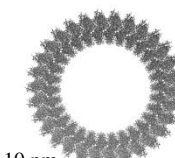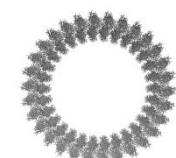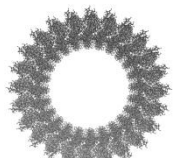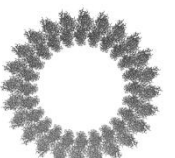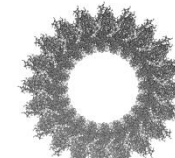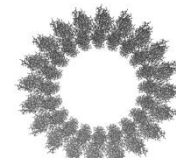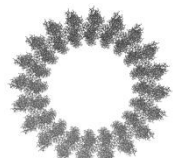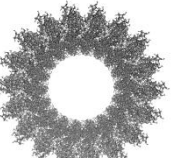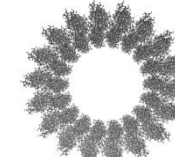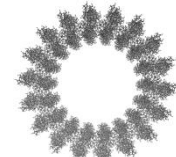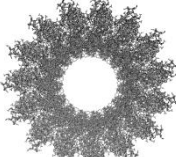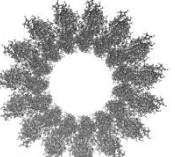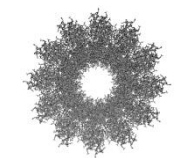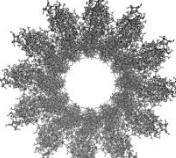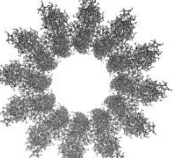
Symptoms

CNS   : central nervous system involvement
F   : fever
S   : splenomegaly
C   : cytopenia
H/H   : hypertriglyceridemia and/or hypofibrinogenemia
H   : hemophagocytosis

Pore Models

Table 11: Perforin pore models

Pore models differ from each other in the size and the number of monomers involved. All models were created in VMD by following the same procedure explained as in the Methods part.

 10 nm			
n=30	n=30	n=24	n=24
n=20	n=20*	n=20	n=18
n=18	n=18	n=15	n=15
n=15	n=12	n=12	n=12

n : number of identical monomers

* : reference pore model given in the article

Structure-Phenotype Relationship

Table 12: Structure-phenotype relationship of missense perforin mutations on FHL2 patients

Mutations are given in the first two columns. Third column shows the stability change upon mutation (ΔΔG) values calculated by FoldX. In the fourth column, the total relative accessible surface areas (RASA) are given for the original residue before mutation in the monomeric model structure. The fifth and the sixth columns show the clinical phenotypes of the patients carrying the corresponding mutations, natural killer cell activity and the percentage of perforin expressing natural killer cells, respectively. The last column shows if the mutation is observed in heterozygous state (yes), homozygous state (no) or compound heterozygous state (with the other mutation name). See **Appendix Table 9** for references.

Mutation		Stability Change (ΔΔG in kcal/mol)	Total RASA (%)	NK function[1]	Perforin expressing NK cells (%)[2]	Heterozygous (with)
1-letter code	3-letter code					
V38L	VAL38LEU	2,24	1,497	NA	NA	yes
V38M	VAL38MET	2,70	1,497	NA	NA	no
P39H	PRO39HIS	2,08	9,751	0.01	1	G149S
G45R	GLY45ARG	10,24	0,116	0.12	61	R54C
G45E	GLY45GLU	9,62	0,116	0.01	1	L17X
V50M	VAL50MET	2,66	0,000	X	X	X
R54C	ARG54CYS	0,16	50,243	0.12	61	G45R
D70V	ASP70VAL	2,50	53,344	NA	NA	Y219X
D70Y	ASP70TYR	1,86	53,344	0.01	1	M1I
C73R	CYS73ARG	13,57	0,308	0.01	1	L17X
C73Y	CYS73TYR	25,09	0,308	NA	NA	L17X
P89T	PRO89THR	5,02	0,636	NA	reduced	NA
A91V[*]	ALA91VAL[*]	-1,07[*]	2,994[*]	X[*]	X[*]	X[*]
W95R	TRP95ARG	5,79	3,178	absent	NA	no
C102F	CYS102PHE	7,88	23,388	NA	NA	yes
R119W	ARG119TRP	-0,34	58,547	significantly reduced	NA	E46X & A91V
R123H	ARG123HIS	0,68	57,741	NA	NA	NA
R126C	ARG126CYS	0,76	70,943	NA	NA	NA
V135M	VAL135MET	-0,67	11,020	NA	NA	NA
G149R	GLY149ARG	16,23	0,000	NA	NA	NA
G149S	GLY149SER	4,32	0,000	X	X	X
F157V	PHE157VAL	2,05	14,511	0.01	1	1628insT(fs)
S168N	SER168ASN	-0,03	16,006	NA	NA	yes
S170R	SER170ARG	0,46	0,258	absent	NA	E243X
R177C	ARG177CYS	-0,16	46,484	NA	NA	NA
V183G	VAL183GLY	3,14	12,588	NA	NA	yes
F193L	PHE193LEU	1,89	0,000	NA	NA	R410P
G198R	GLY198ARG	-0,69	51,310	NA	NA	yes

P201T	PRO201THR	3,57	34,886	NA	NA	L285X
L215I	LEU215ILE	0,90	5,408	NA	absent	A262D
G220R	GLY220ARG	20,99	0,655	absent	absent	NA
G220S	GLY220SER	13,17	0,655	NA	NA	No
T221I	THR221ILE	1,04	0,056	NA	NA	R225W
H222R	HIS222ARG	7,72	0,002	0.01	1	yes
H222Q	HIS222GLN	3,21	0,002	0.01	1	1636delC (fs)
I224N	ILE224ASN	4,36	0,372	significantly reduced	absent	No
R225P	ARG225PRO	4,37	26,072	NA	absent	No
R225Q	ARG225GLN	0,98	26,072	NA	NA	yes
R225W	ARG225TRP	2,84	26,072	0.01/0.01/0.01	91/54/NA	no/no/S150X
R232C	ARG232CYS	1,64	12,879	absent	NA	1182insT(G394fs & stop)
R232H	ARG232HIS	30,21	12,879	significantly reduced	absent	NA
R240G	ARG240GLY	2,29	20,629	NA	NA	No
N252S	ASN252SER	0,17	52,647	X	X	X
E253K	GLU253LYS	0,17	29,238	absent	absent	285delLys
E261K	GLU261LYS	1,29	8,665	significantly reduced	absent	No
A262D	ALA262ASP	11,54	0,000	NA	absent	L215I
C279Y	CYS279TYR	7,14	5,386	NA	NA	yes
E298G	GLU298GLY	1,69	28,521	absent	NA	NA
R299C	ARG299CYS	1,25	27,350	0.01	NA	G149S
G305D	GLY305ASP	23,70	13,117	NA	NA	R356W
G306C	GLY306CYS	5,77	18,000	significantly reduced(2)/normal(1) (triplet)	absent(2)/normal	1190-1191insTG(fs & stop)(2)/yes
D313V	ASP313VAL	2,15	16,085	NA	NA	M1I
G317R	GLY317ARG	2,59	21,354	NA	NA	M1V
G334S	GLY334SER	-0,32	55,949	NA	NA	NA
P345L	PRO345LEU	1,90	3,454	NA	NA	NA
R356W	ARG356TRP	1,62	27,695	absent	absent	NA
R357W	ARG357TRP	0,43	29,976	NA	NA	NA
R361W	ARG361TRP	2,32	23,118	0.01	18	G149S
R410P	ARG410PRO	-0,56	53,102	NA	NA	F193L
R410W	ARG410TRP	0,79	53,102	2	NA	L364fs
F421C	PHE421CYS	0,55	39,681	NA	NA	NA
G426S	GLY426SER	0,10	33,852	NA	NA	NA
G429E	GLY429GLU	5,65	14,953	NA	NA	NA
D430Y	ASP430TYR	-0,47	85,615	NA	absent	G317R
T435M	THR435MET	2,68	0,196	NA	NA	NA
Y438C	TRY438CYS	3,48	3,738	NA	NA	NA

Q446R	GLN446ARG	**-0,15**	39,388	NA	NA	NA
T450M	THR450MET	**1,51**	10,558	absent	NA	NA
P459L	PRO459LEU	**4,33**	0,000	NA	NA	NA
P477A	PRO477ALA	**0,21**	45,191	absent	absent	no
Q481P	GLN481PRO	**1,94**	16,770	0.01	1	L17X
D486G	ASP486GLY	**0,80**	25,115	NA	NA	yes
D491N	ASP491ASN	**1,49**	24,204	NA	NA	NA
R509K	ARG509LYS	**-0,17**	77,683	NA	NA	yes
A523D	ALA523ASP	**2,71**	0,101	NA	NA	no
C525S	CYS525SER	**0,89**	26,868	NA	NA	NA

 > 2 (kcal/mol)   (destabilizing)   buried (< 5%)

 between 0 and 2 (kcal/mol)   intermediate exposure (between 5% and 30%)

 < 0 (kcal/mol)   (stabilizing)   exposed (> 30%)

* : discussed further in the article
NA : not available
NK : natural killer
[1] : normal values: > 3,2
[2] : normal values: 91–97 (0–1 year), 81–91 (1–15 years)
X: reported in many cases with diverse clinical phenotype and/or in different genetic forms (homozygous, heterozygous or compound heterozygous)

BIBLIOGRAPHY

[1]   J. W. Farquhar and A. E. Claireaux, "Familial haemophagocytic reticulosis," *Arch Dis Child,* vol. 27, pp. 519-25, Dec 1952.

[2]   J. H. Russell and T. J. Ley, "Lymphocyte-mediated cytotoxicity," *Annu Rev Immunol,* vol. 20, pp. 323-70, 2002.

[3]   S. Nagata, "Fas-mediated apoptosis," *Adv Exp Med Biol,* vol. 406, pp. 119-24, 1996.

[4]   J. C. Stinchcombe and G. M. Griffiths, "Secretory mechanisms in cell-mediated cytotoxicity," *Annu Rev Cell Dev Biol,* vol. 23, pp. 495-517, 2007.

[5]   I. Voskoboinik*, et al.*, "Perforin-mediated target-cell death and immune homeostasis," *Nat Rev Immunol,* vol. 6, pp. 940-52, Dec 2006.

[6]   C. Gholam*, et al.*, "Familial haemophagocytic lymphohistiocytosis: advances in the genetic basis, diagnosis and management," *Clin Exp Immunol,* vol. 163, pp. 271-83, Mar 2011.

[7]   D. Kagi*, et al.*, "Cytotoxicity mediated by T cells and natural killer cells is greatly impaired in perforin-deficient mice," *Nature,* vol. 369, pp. 31-7, May 5 1994.

[8]   B. Lowin*, et al.*, "A null mutation in the perforin gene impairs cytolytic T lymphocyte- and natural killer cell-mediated cytotoxicity," *Proc Natl Acad Sci U S A,* vol. 91, pp. 11571-5, Nov 22 1994.

[9]   H. Kojima*, et al.*, "Two distinct pathways of specific killing revealed by perforin mutant cytotoxic T lymphocytes," *Immunity,* vol. 1, pp. 357-64, Aug 1994.

[10]   C. M. Walsh*, et al.*, "Immune function in mice lacking the perforin gene," *Proc Natl Acad Sci U S A,* vol. 91, pp. 10854-8, Nov 8 1994.

[11]   I. Voskoboinik*, et al.*, "Calcium-dependent plasma membrane binding and cell lysis by perforin are mediated through its C2 domain: A critical role for aspartate residues 429, 435, 483, and 485 but not 491," *J Biol Chem,* vol. 280, pp. 8426-34, Mar 4 2005.

[12]   J. I. Henter*, et al.*, "Incidence in Sweden and clinical features of familial hemophagocytic lymphohistiocytosis," *Acta Paediatr Scand,* vol. 80, pp. 428-35, Apr 1991.

[13]   E. Ishii*, et al.*, "Clinical and epidemiologic studies of familial hemophagocytic lymphohistiocytosis in Japan. Japan LCH Study Group," *Med Pediatr Oncol,* vol. 30, pp. 276-83, May 1998.

[14]   A. Gurgey*, et al.*, "Primary hemophagocytic lymphohistiocytosis in Turkish children," *Pediatr Hematol Oncol,* vol. 20, pp. 367-71, Jul-Aug 2003.

[15]   S. E. Stepp*, et al.*, "Perforin gene defects in familial hemophagocytic lymphohistiocytosis," *Science,* vol. 286, pp. 1957-9, Dec 3 1999.

[16]   J. I. Henter*, et al.*, "Hypercytokinemia in familial hemophagocytic lymphohistiocytosis," *Blood,* vol. 78, pp. 2918-22, Dec 1 1991.

[17]   J. I. Henter and I. Nennesmo, "Neuropathologic findings and neurologic symptoms in twenty-three children with hemophagocytic lymphohistiocytosis," *J Pediatr,* vol. 130, pp. 358-65, Mar 1997.

[18]   M. Ohadi*, et al.*, "Localization of a gene for familial hemophagocytic lymphohistiocytosis at chromosome 9q21.3-22 by homozygosity mapping," *Am J Hum Genet,* vol. 64, pp. 165-71, Jan 1999.

[19]   M. Arico*, et al.*, "Hemophagocytic lymphohistiocytosis. Report of 122 children from the International Registry. FHL Study Group of the Histiocyte Society," *Leukemia,* vol. 10, pp. 197-203, Feb 1996.

[20]   K. Goransdotter Ericson*, et al.*, "Spectrum of perforin gene mutations in familial hemophagocytic lymphohistiocytosis," *Am J Hum Genet,* vol. 68, pp. 590-7, Mar 2001.

[21]   U. Zur Stadt*, et al.*, "Mutation spectrum in children with primary hemophagocytic lymphohistiocytosis: molecular and functional analyses of PRF1, UNC13D, STX11, and RAB27A," *Hum Mutat,* vol. 27, pp. 62-8, Jan 2006.

[22]   J. I. Henter*, et al.*, "HLH-94: a treatment protocol for hemophagocytic lymphohistiocytosis. HLH study Group of the Histiocyte Society," *Med Pediatr Oncol,* vol. 28, pp. 342-7, May 1997.

[23]   J. I. Henter*, et al.*, "Treatment of hemophagocytic lymphohistiocytosis with HLH-94 immunochemotherapy and bone marrow transplantation," *Blood,* vol. 100, pp. 2367-73, Oct 1 2002.

[24]   J. I. Henter*, et al.*, "HLH-2004: Diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis," *Pediatr Blood Cancer,* vol. 48, pp. 124-31, Feb 2007.

[25]   L. Bordoli*, et al.*, "Protein structure homology modeling using SWISS-MODEL workspace," *Nat Protoc,* vol. 4, pp. 1-13, 2009.

[26]   M. Magrane and U. Consortium, "UniProt Knowledgebase: a hub of integrated protein data," *Database (Oxford),* vol. 2011, p. bar009, 2011.

[27]   P. D. Stenson*, et al.*, "The Human Gene Mutation Database: 2008 update," *Genome Med,* vol. 1, p. 13, 2009.

[28]   I. F. Fokkema*, et al.*, "LOVD v.2.0: the next generation in gene variant databases," *Hum Mutat,* vol. 32, pp. 557-63, May 2011.

[29]   J. Schymkowitz*, et al.*, "The FoldX web server: an online force field," *Nucleic Acids Res,* vol. 33, pp. W382-8, Jul 1 2005.

[30]   R. Guerois*, et al.*, "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations," *J Mol Biol,* vol. 320, pp. 369-87, Jul 5 2002.

[31]   W. Humphrey*, et al.*, "VMD: visual molecular dynamics," *J Mol Graph,* vol. 14, pp. 33-8, 27-8, Feb 1996.

[32]   T. Praper*, et al.*, "Human perforin employs different avenues to damage membranes," *J Biol Chem,* vol. 286, pp. 2946-55, Jan 28 2011.

[33]   E. Cukuroglu*, et al.*, "Analysis of hot region organization in hub proteins," *Ann Biomed Eng,* vol. 38, pp. 2068-78, Jun 2010.

[34]   J. Mihel*, et al.*, "PSAIA - protein structure and interaction analyzer," *BMC Struct Biol,* vol. 8, p. 21, 2008.

[35]   H. Okur*, et al.*, "Clinical and molecular aspects of Turkish familial hemophagocytic lymphohistiocytosis patients with perforin mutations," *Leuk Res,* vol. 32, pp. 972-5, Jun 2008.

[36]   S. Molleran Lee, *et al.*, "Characterisation of diverse PRF1 mutations leading to decreased natural killer cell activity in North American families with haemophagocytic lymphohistiocytosis," *J Med Genet,* vol. 41, pp. 137-44, Feb 2004.

[37]   K. Kogawa, *et al.*, "Perforin expression in cytotoxic lymphocytes from patients with hemophagocytic lymphohistiocytosis and their family members," *Blood,* vol. 99, pp. 61-6, Jan 1 2002.

[38]   R. Clementi, *et al.*, "Six novel mutations in the PRF1 gene in children with haemophagocytic lymphohistiocytosis," *J Med Genet,* vol. 38, pp. 643-6, Sep 2001.

[39]   J. Feldmann, *et al.*, "Functional consequences of perforin gene mutations in 22 patients with familial haemophagocytic lymphohistiocytosis," *Br J Haematol,* vol. 117, pp. 965-72, Jun 2002.

[40]   Y. T. Bryceson, *et al.*, "Defective cytotoxic lymphocyte degranulation in syntaxin-11 deficient familial hemophagocytic lymphohistiocytosis 4 (FHL4) patients," *Blood,* vol. 110, pp. 1906-15, Sep 15 2007.

[41]   R. H. Law, *et al.*, "The structural basis for membrane binding and pore formation by lymphocyte perforin," *Nature,* vol. 468, pp. 447-51, Nov 18 2010.

[42]   K. Baran, *et al.*, "The molecular basis for perforin oligomerization and transmembrane pore assembly," *Immunity,* vol. 30, pp. 684-95, May 2009.

[43]   C. J. Rosado, *et al.*, "The MACPF/CDC family of pore-forming toxins," *Cell Microbiol,* vol. 10, pp. 1765-74, Sep 2008.

[44]   V. Parthiban, *et al.*, "CUPSAT: prediction of protein stability upon point mutations," *Nucleic Acids Res,* vol. 34, pp. W239-42, Jul 1 2006.

[45]   Y. Dehouck, *et al.*, "PoPMuSiC 2.1 : a web server for the estimation of protein stability changes upon mutation and sequence optimality," *BMC Bioinformatics,* vol. 12, p. 151, May 13 2011.

[46]   N. Tuncbag, *et al.*, "Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy," *Bioinformatics,* vol. 25, pp. 1513-20, Jun 15 2009.

[47]   J. McCormick, *et al.*, "Novel perforin mutation in a patient with hemophagocytic lymphohistiocytosis and CD45 abnormal splicing," *Am J Med Genet A,* vol. 117A, pp. 255-60, Mar 15 2003.

[48]   A. Trizzino, *et al.*, "Genotype-phenotype study of familial haemophagocytic lymphohistiocytosis due to perforin mutations," *J Med Genet,* vol. 45, pp. 15-21, Jan 2008.

[49]   U. Zur Stadt, *et al.*, "A91V is a polymorphism in the perforin gene not causative of an FHLH phenotype," *Blood,* vol. 104, p. 1909; author reply 1910, Sep 15 2004.

[50]   R. Busiello, *et al.*, "Atypical features of familial hemophagocytic lymphohistiocytosis," *Blood,* vol. 103, pp. 4610-2, Jun 15 2004.

[51]   C. Trambas, *et al.*, "A single amino acid change, A91V, leads to conformational changes that can impair processing to the active form of perforin," *Blood,* vol. 106, pp. 932-7, Aug 1 2005.

[52]   E. Mancebo, *et al.*, "Familial hemophagocytic lymphohistiocytosis in an adult patient homozygous for A91V in the perforin gene, with tuberculosis infection," *Haematologica,* vol. 91, pp. 1257-60, Sep 2006.

[53]   K. A. Risma, *et al.*, "Aberrant maturation of mutant perforin underlies the clinical diversity of hemophagocytic lymphohistiocytosis," *J Clin Invest,* vol. 116, pp. 182-92, Jan 2006.

[54]   Q. Lu, *et al.*, "DNA methylation and chromatin structure regulate T cell perforin gene expression," *J Immunol,* vol. 170, pp. 5124-32, May 15 2003.

[55]   A. J. Brennan, *et al.*, "Perforin deficiency and susceptibility to cancer," *Cell Death Differ,* vol. 17, pp. 607-15, Apr 2010.

[56]   M. E. Masson and R. B. May, "Identification of words and letters during reading: a sentence inferiority effect for letter detection," *Can J Psychol,* vol. 39, pp. 449-59, Sep 1985.

[57]   E. R. Podack, *et al.*, "Isolation and biochemical and functional characterization of perforin 1 from cytolytic T-cell granules," *Proc Natl Acad Sci U S A,* vol. 82, pp. 8629-33, Dec 1985.

[58]   C. J. Rosado, *et al.*, "A common fold mediates vertebrate defense and bacterial attack," *Science,* vol. 317, pp. 1548-51, Sep 14 2007.

[59]   Q. Xu, *et al.*, "Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont Bacteroides thetaiotaomicron," *Acta Crystallogr Sect F Struct Biol Cryst Commun,* vol. 66, pp. 1297-305, Oct 1 2010.

[60]   M. A. Hadders, *et al.*, "Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense," *Science,* vol. 317, pp. 1552-4, Sep 14 2007.