20th IMEKO TC4 International Symposium and
18th International Workshop on ADC Modelling and Testing
Research on Electric and Electronic Measurement for the Economic Upturn
Benevento, Italy, September 15-17, 2014

Molecular dynamics simulations to study pathogenic mutations in proteins.

Valentina Mirabelli[1], Rocco Caliandro[1], Giovanni Nico[2], Giovanni Luca Tiscia[3], Giovanni Favuzzi[3], Elvira Grandone[3].
[1]*Institute of Crystallography - CNR, Via Amendola 122/O, 70126 Bari, Italy.*
[2]*Institute for Applied Mathematics "Mauro Picone"- CNR, Via Amendola 122/I, 70126 Bari, Italy.*
[3]*Atherosclerosis and Thrombosis Unit, I.R.C.C.S. Casa Sollievo della Sofferenza, S. Giovanni Rotondo (FG), Italy. valentina.mirabelli@gmail.com.*

Abstract – Here we present an example of application of molecular dynamics simulations, i.e. the study of pathogenic mutations in a protein, the FXI factor, that causes defects in blood coagulation. Starting from the measurement of the crystal structure of the protein, molecular dynamics simulations allow us to study changes in flexibility and movements of the mutated protein compared to the native form of the protein, and thus help to understand the molecular basis of disease. Molecular dynamics data have been extensively generated by using grid computing. This large amount of data has been analyzed by algorithms which capture structural features. In particular, the variations in the atomic distances between amino acids and in their exposure to the solvent , the formation or breaking of hydrogen bonds, and the distributions of the backbone dihedral angles, were calculated. Thus insights into the molecular basis of pathogenesis for each mutation found in population were obtained.

I. INTRODUCTION

Until short time ago, the study of biomolecules was limited to macroscopic observations, with in vitro or in vivo experiments, that gave clues to the function, trying to understand the wide variety of biological mechanisms. With the discovery of the X-ray diffraction from biomolecular crystals, macroscopic studies joined the microscopic study of the structure, at atomic level, which help to define in an ever more detailed and precise even function. Since always, structure-function binomial is, indeed, a key principle in the study of cellular components like proteins and nucleic acids. Therefore, the use of means such as X-ray Crystallography or Nuclear Magnetic Resonance (NMR), that allows us to "see" forbidden entities to the human eye, challenges more and more our knowledge in the field, opening us to new discoveries and questions. Because of its infinite potential and for the revolution that made in understanding the world around us, Crystallography is celebrated in 2014, one hundred years after its discovery and initial application by Laue, Bragg and Lawrence. [www.iycr2014.org].

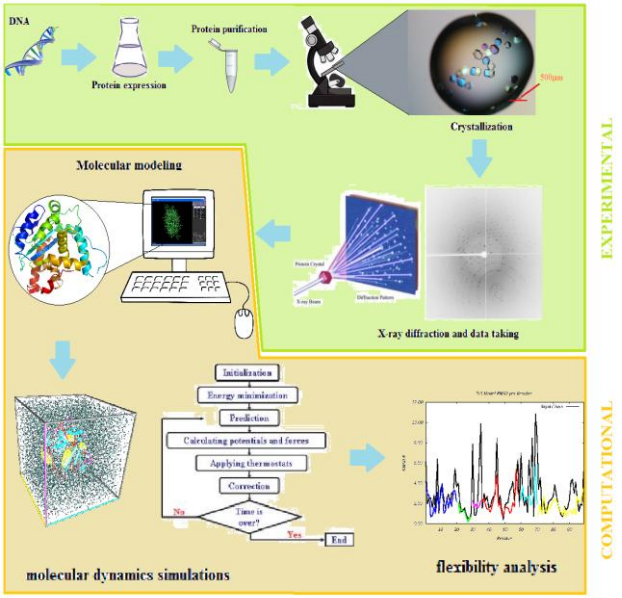

Fig. 1 – An overview image to give an idea of the workflow from experimental (green) to computational (orange) steps.

The X-ray diffraction measurement of protein structures requires a number of steps, which are outlined in Fig. 1 (experimental steps). The protein to be studied has to be produced in milligram amounts with high purity. The best way to achieve this, is the recombinant expression, i.e. using simple organisms such as bacteria and yeasts, capable, if properly treated, to produce proteins not normally present in their genetic heritage, belonging to other organisms, such as humans [1]. Protein samples are then used to set up crystallization trials, where a variety of crystallization conditions are sampled by varying protein concentration, precipitant

ISBN-14: 978-92-990073-2-7

solutions, pH, temperature etc. The crystallization set up is miniaturized, so that each trial requires nanolitres of protein samples. Crystallization trials are monitored by stereo-microscopes, to check the crystal growth. When/if protein crystals of dimensions of some micrometers grow, they can be used for an X-ray diffraction experiment. It is performed at synchrotron sources, which guarantee the high brilliance necessary to collect diffraction data from poorly diffracting objects as protein crystals (which are mainly composed of water). The diffraction pattern is detected by hybrid pixel detectors with fast readout, high sensitiveness and single photon counting capabilities [2]. A number of images produced by such detectors, each corresponding to a given orientation of the crystal with respect to the X-ray beam, are stored and processed for obtaining the crystal structure of the protein by means of complex algorithms known as phasing methods [3]. Protein crystal structures are validated by stereo-chemistry checks and deposited on a public database: the Protein Data Bank (PDB) [http://www.rcsb.org].

However, one more step has been made in recent decades, due to computational simulations of Molecular Dynamics (MD) [4], that starting from the atomic coordinates obtained from the crystal structure and applying the Newton's laws of motion of classical mechanics to every single atom, allows us to "see" also movements of the molecule, and thus reconstruct in silico, that is by the computer, a condition as similar as possible to that present in a biological organism, wherein the molecule is immersed in an aqueous environment and in which it interacts with other cellular components. The interaction between molecules needed to exert their own role often requires conformational changes, or simple oscillations of the atoms, that the static crystal structure is not able to reveal. The observed movement is then interpreted in a biological contest, to obtain useful information, in order to characterize the unknown function or cellular interactions of newly discovered molecules, for basic research, or lets us design drugs that act in a targeted and more effective way on the molecule, minimizing side effects, as an example for health applied research. All this requires the development of highly sophisticated algorithms, able to process a huge amount of data and perform very complex calculations in a reasonable time, along with the use of even more powerful computers able to take a number of measures at the same time. Thus, series of cross skills are required in the field of physics, chemistry, computer science, mathematics, in collaboration with physicians and biologists, at the service of human health and to improve the quality of life.

Here we present an example of application of MD, i.e. the study of pathogenic mutations in a protein, that causes defects in coagulation.

II. FACTOR XI

A. Factor XI structure and function.

Factor XI (FXI) [OMIM*264900] is a glycoprotein mainly synthesized by liver cells that circulates in plasma

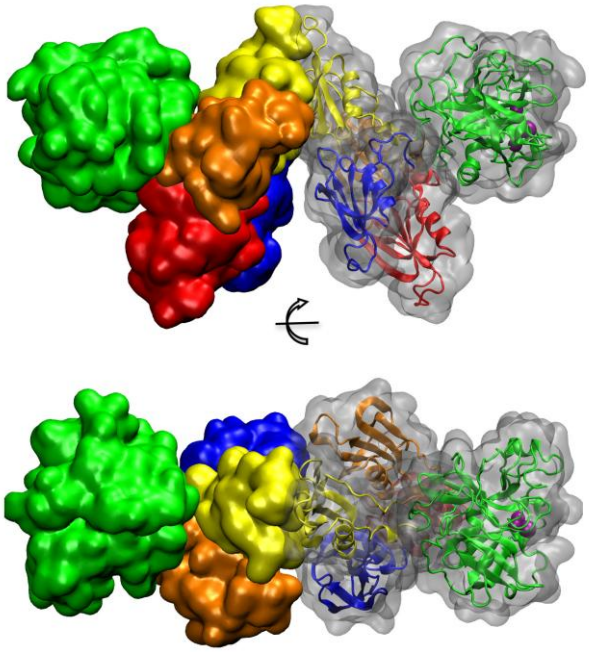

Fig. 2 - Factor XI homodimer structure viewed from two perspectives rotated 90 degrees. Apple domains are shown in blue (Ap1), red (Ap2), orange (Ap3) and yellow (Ap4), Catalytic domain (SP) is shown in green, with catalytic triad residues represented as purple beads. Monomers are shown in cartoon (right) and surface (left) representations.

as a complex with high molecular weight kininogen (HK) and participates in the propagation phase of blood coagulation as a catalyst in the conversion of factor IX to factor IXa, (Fig. 3) [5]. The F11 gene is located on chromosome 4 at position 4q35, has 15 exons and spans 23kb. Exon 1 encodes for the promoter region, exon 2 for the 18 amino acid long-signal peptide that is cleaved during FXI biosynthesis, and exons 3 to 15 encode for the 607 amino acid-long mature protein. FXI mature protein is a 80kDa molecule, containing 4 N-terminal 90- or 91-amino acid repeats called apple domains (Ap1 to Ap4), that form a disk structure, and a C-terminal serine protease domain (SP), with the catalytic triad at His413, Asp462 and Ser557 (Fig. 2). It is unique among coagulation proteases in that it circulates as a 160kDa homodimer with two identical FXI monomers linked by non-covalent interactions between the Ap4 domains and by a Cys321-Cys321 disulphide bond, although this bond is not essential for dimerisation. FXI zymogen is

converted to his active form (FXIa) by FXIIa, thrombin or may undergo auto-activation. Activation of FXI results in cleavage of the Arg369-Ile370 bond to form a heavy chain (369 amino acids), containing the Ap domains and a light chain (238 amino acids) containing the SP domain.

B. Inherited FXI deficiency and coagulation disorders.

Mutations within the F11 gene can cause FXI deficiency leading to a disorder with a variable clinical phenotype, also known as Haemophilia C, Plasma thromboplastin antecendent deficiency and Rosenthal syndrome. Congenital FXI deficiency is a rare bleeding disorder observed with a frequency of 1:1'000'000 worldwide. However, it is prevalent in some ethnic groups like Ashkenazi Jews, among whom severe FXI deficiency is frequent (1:450), representing one of the most common inherited disorders. Although four cases of dominant transmission have been reported, caused by the formation of non-secretable FXI heterodimers, severe FXI deficiency generally follows an autosomal recessive pattern of inheritance. More than 190 disease causing mutations have been described so far in the F11 gene, which are included in FXI mutation databases such as www.factorxi.org. The recent availability of crystal structures for FXI zymogen [PDB entry 2F83], (Fig. 2) and the FXIa catalytic domain have enhanced our understanding of structure-function relationships for this molecule, the activation mechanism, the interaction with substrates and inhibitors. Thus, structural analysis of pathogenic mutations associated with FXI deficiency today may help us to understand its physiopathology and to develop better therapeutic treatments for haemorrhagic disorders.

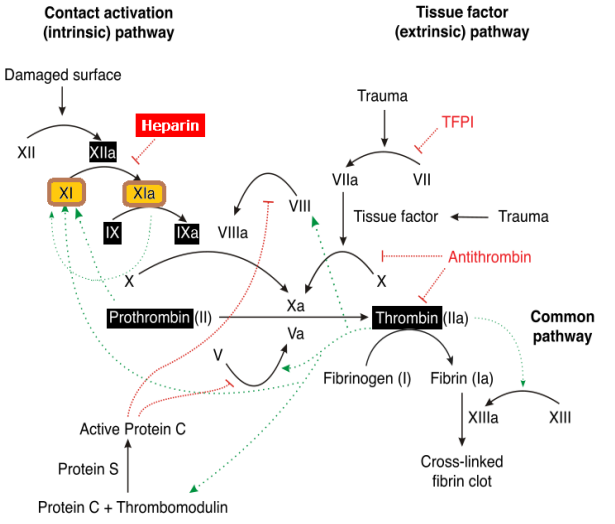

Fig. 3 – Coagulation cascade with positive (→) and negative (-|) feedback regulation. Factor XI is converted to its active form XIa by XIIa, thrombin and FXIa itself, with prothrombin as cofactor. It catalyzes conversion of IX to IXa and it can be inhibited by heparin.

III. MOLECULAR DYNAMICS SIMULATIONS

A. Generalities.

A PDB file containing the atomic coordinates of each atom in the protein, obtained by X-ray diffraction measurements (Fig. 4), represents the static three-dimensional model, to which the simulation will be applied.

ionized WT pdb file

```
CRYST1  143.306  143.306  116.195  90.00  90.00  90.00 P 1           1
.
.
ATOM   7645  N   TRP P 497      28.983  92.978 116.476  1.00  0.00      P1   N
ATOM   7646  HN  TRP P 497      28.660  92.139 116.045  1.00  0.00      P1   H
ATOM   7647  CA  TRP P 497      29.424  94.015 115.553  1.00  0.00      P1   C
ATOM   7648  HA  TRP P 497      29.226  94.984 115.995  1.00  0.00      P1   H
ATOM   7649  CB  TRP P 497      30.898  93.812 115.183  1.00  0.00      P1   C
ATOM   7650  HB1 TRP P 497      30.985  92.838 114.651  1.00  0.00      P1   H
ATOM   7651  HB2 TRP P 497      31.220  94.600 114.467  1.00  0.00      P1   H
ATOM   7652  CG  TRP P 497      31.842  93.760 116.352  1.00  0.00      P1   C
ATOM   7653  CD1 TRP P 497      31.838  92.851 117.371  1.00  0.00      P1   C
ATOM   7654  HD1 TRP P 497      31.146  92.022 117.444  1.00  0.00      P1   H
ATOM   7655  NE1 TRP P 497      32.854  93.118 118.251  1.00  0.00      P1   N
ATOM   7656  HE1 TRP P 497      33.081  92.597 119.046  1.00  0.00      P1   H
.
.
ATOM  18728  OG1 THR P 604      72.295  21.957 148.928  1.00  0.00      P2   O
ATOM  18729  HG1 THR P 604      72.850  22.671 148.595  1.00  0.00      P2   H
ATOM  18730  CG2 THR P 604      73.593  22.179 150.924  1.00  0.00      P2   C
ATOM  18731 HG21 THR P 604      74.182  23.042 150.548  1.00  0.00      P2   H
ATOM  18732 HG22 THR P 604      73.551  22.251 152.032  1.00  0.00      P2   H
ATOM  18733 HG23 THR P 604      74.134  21.246 150.656  1.00  0.00      P2   H
.
.
ATOM  37792  CLA CLA I  24       7.457  32.620 129.013  1.00  0.00      ION CL
ATOM  37793  CLA CLA I  25     121.788  59.457 133.532  1.00  0.00      ION CL
END
```

COMPUTATIONAL MUTAGENESIS

ionized TRP497GLY missense mutation pdb file

```
CRYST1  143.306  143.306  116.195  90.00  90.00  90.00 P 1           1
.
.
ATOM   7641  N   GLY P 497      28.983  92.978 116.476  1.00  0.00      P1   N
ATOM   7642  HN  GLY P 497      28.660  92.139 116.045  1.00  0.00      P1   H
ATOM   7643  CA  GLY P 497      29.424  94.015 115.553  1.00  0.00      P1   C
ATOM   7644  HA1 GLY P 497      30.453  93.827 115.277  1.00  0.00      P1   H
ATOM   7645  HA2 GLY P 497      29.254  94.986 115.999  1.00  0.00      P1   H
ATOM   7646  C   GLY P 497      28.575  93.921 114.296  1.00  0.00      P1   C
ATOM   7647  O   GLY P 497      27.994  92.877 114.013  1.00  0.00      P1   O
.
.
ATOM  18685  HB  THR P 604      71.626  23.142 150.487  1.00  0.00      P2   H
ATOM  18686  OG1 THR P 604      72.295  21.957 148.928  1.00  0.00      P2   O
ATOM  18687  HG1 THR P 604      72.850  22.671 148.595  1.00  0.00      P2   H
ATOM  18688  CG2 THR P 604      73.593  22.179 150.924  1.00  0.00      P2   C
ATOM  18689 HG21 THR P 604      74.182  23.042 150.548  1.00  0.00      P2   H
ATOM  18690 HG22 THR P 604      73.551  22.251 152.032  1.00  0.00      P2   H
ATOM  18691 HG23 THR P 604      74.134  21.246 150.656  1.00  0.00      P2   H
.
.
ATOM  37778  CLA CLA I  25      71.221   5.887 123.356  1.00  0.00      ION CL
END
```

Fig. 4 – An extract from the PDB file, showing the atomic coordinates for thousands of atoms in the simulated system. In this case, the amino acid tryptophan at the 497 position in WT protein sequence has been changed in a glycine, that is one of the pathogenic mutations considered.

The MD is based on Newton's second law:

$$F = ma \qquad (1)$$

where F is derived by proper force fields tuned for the study protein atoms. By integrating the equations of motion, one can derive the trajectory, namely positions r (t) as a function of time and velocity at (t). As the method is deterministic, the trajectory is determined by the initial conditions. The result of a simulation is a time series of molecular conformations corresponding to the trajectories for each single atom.

B. *MD simulations protocol.*

MD simulations were carried out by using the complete biological assembly constituted by the FXI zymogen available on the Protein Data Bank (PDB entry 2F83, Fig. 2). Mutated (MT) structures were generated replacing wild-type (WT) residues according to the pathogenic mutations considered, by computational mutagenesis (Fig. 4). Mutated (MT) models were generated by implementing the observed mutations by means of the program VMD [6]. MD calculations were performed by NAMD [7].

The simulation protocol was set as follows: the WT and MT systems underwent 2000 steps (i.e. 4 ps) of steepest-descent energy minimization. Then they were gradually heated from 0K up to 298K in 12 steps of 100ps simulation, during which harmonic position restraints were imposed on the protein atoms, by using a harmonic spring constant of 0.5 kcal mol-1Å-2. After that, a 200 ps equilibration was performed. Finally, 26 ns-long productive MD simulations were run in the NPT ensemble (298 K, 1 atm and 2 fs time-step). Coordinates were saved at regular time intervals of every 5 ps. The frames collected during the final MD simulations were considered for further analysis. MD calculations were performed on the grid datacenter "Bari Computer Center for Science, PON RECAS - INFN-Bari & UNIBA", and related data were analyzed by VMD [6].

C. *Simulation data analysis.*

Some properties that can be calculated from a trajectory are the Root Mean Square Deviation (RMSD), the Solvent Accessible Surface Area (SASA), the distances between selected amino acid residues, the presence of Hydrogen Bonds (HB) and the Protein Angular Dispersion (PAD).

The RMSD, calculated throughout the simulation, is an index of the system stability, which reveals deviations of the atomic positions with respect to their initial values.

The solvent accessible surface area (SASA) was calculated by using a probe sphere with radius of 0.14 nm.

Hydrogen bonds (HBs) were determined based on a cut-off donor-acceptor distance of 3.5Å and a cut-off acceptor-donor-hydrogen angle of 30°.

The PAD measures the spread of the backbone dihedral angles visited during the dynamics, and this is used to quantify the residue-by-residue flexibility in MT simulations [8]. It has proven effective in identifying anomalous protein plasticity upon pathogenic mutations [9,10], and in interpreting MD results to experimental observations [11].

IV. RESULTS

The structural deviation of the WT and the MT systems from their starting structures, expressed as the Root Mean Square Deviation (RMSD) of their Cα atoms is shown in Fig. 5. Following analysis has been carried out within time intervals where a relative stabilization of the RMSD values could be observed (Table 1). This ensures that the simulated system has reached its equilibrium state, hence the residual structural movements are due to the intrinsic protein flexibility.

The residue-by-residue PAD distributions determined for each mutant of the FXI protein (MT) have been compared with that obtained for the wild type protein (WT), i.e. the functional protein as it is normally present in nature (Fig. 6).

Some noticeable variation in distances, SASA and HBs for functionally important residues, are reported in Table 2

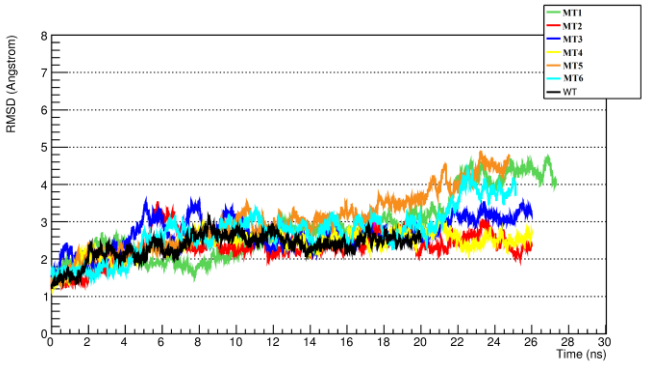

Fig. 5 – Root Mean Square Deviation (RMSD) plot for all systems considered, i.e. mutants (MT) and wild type (WT). Simulation time was about 26 ns for MT and 20 ns for WT.

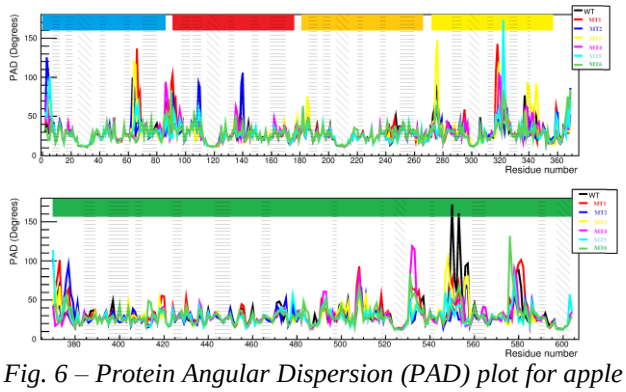

Fig. 6 – Protein Angular Dispersion (PAD) plot for apple domains (top) and catalytic domain (bottom) for the wild type and mutated FXI models. Apple domains residues are overlined in blue (Ap1), red (Ap2), orange (Ap3) and yellow (Ap4), catalytic domain (SP) residues are overlined in green.

Table 1 - Initial time (Tstart), temporal duration (Tstable) and average RMSD values (Å) of the selected part of the MD trajectory, for each simulated system.

System	T_{start} (ns)	T_{stable} (ns)	RMSD (Å)
WT	7.0	13.0	2.5 ± 0.2
MT1	8.0	18.0	2.4 ± 0.2
MT2	12.0	14.0	2.8 ± 0.3
MT3	8.0	18.0	2.6 ± 0.2
MT4	11.0	13.8	3.5 ± 0.6
MT5	8.0	17.0	3.0 ± 0.5
MT6	12.5	14.8	3.5 ± 0.7

Table 2 – An example for values of distances between C_α atoms (Å) (a), solvent accessibility surface areas ($Å^2$) (b), and hydrogen bond occupancies (%) (c) for residues involved in FIX recognition.

System	Arg184-Asn566 (a)	Asn566 (b)	Arg184-Asn566 (c)
WT	7.8 ± 0.4	57 ± 10	62.1
MT1	7.8 ± 0.3	68 ± 8	59.6
MT2	8.0 ± 0.3	63 ± 8	79.2
MT3	**8.5 ± 0.8**	**105 ± 14**	--
MT4	7.7 ± 0.3	69 ± 15	63.5
MT5	7.5 ± 0.4	59 ± 10	66.8
MT6	7.1± 0.3	70 ± 9	32.9

Results from PAD, distances, SASA and HBs variations have been finally merged and interpreted in terms of anomalous plasticity of mutated FXI with respect to that showed by the wild type FXI. As a consequence, hints about the mechanism by which each single mutation in the protein could impair the coagulation system in patients carrying that mutation have been outlined. For example, a critical region, in which the loss of a hydrogen bond, along with other abnormalities, has been observed, is the one containing the binding site of the FXI substrate, FIX (Fig.7). The loss of this hydrogen bond (Table 2) could alter the on-off regulation of the enzyme, switching it on and thus activating the coagulation cascade even in the absence of metabolic signal. This may be related to the role played by this protein in the phenomenon of thrombosis [12].

V. CONCLUSIONS

Here we have presented a modern computational method widely used in the study of biomolecules, i.e. molecular dynamics simulations, which can give very detailed information on phenomena such as conformational changes of proteins and nucleic acids, thus providing a valuable aid in the design of new drugs and in the determination of structures of complex molecules, in combination with experimental techniques such as X-ray diffraction. The cost can be high in terms of computing time: measurements of atomic trajectories and simulations of complex systems can be very expensive. However, the increased performances of computers on the one hand, the refinement of the method and algorithms on the other hand, have made possible calculations that were unthinkable a few years ago.

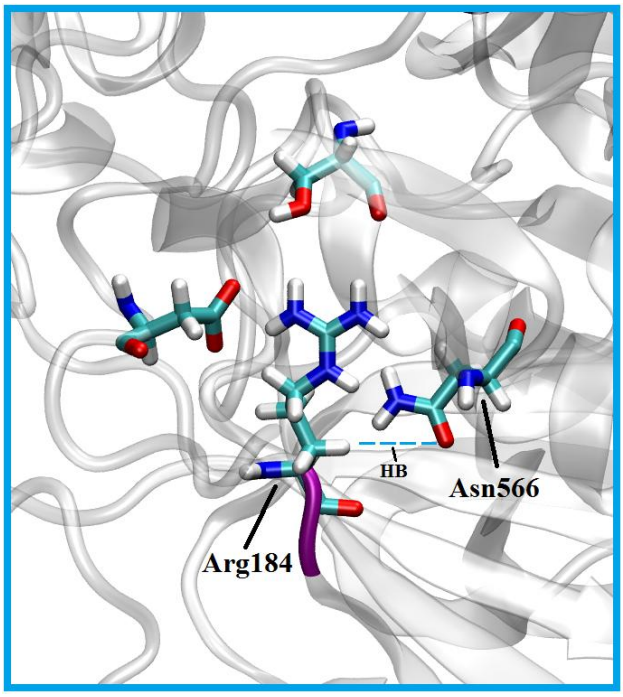

Fig. 7 – Focus on FIX binding site region in the FXI structural model, showing the HB between Arg184 and Asn566 residues. The FXI structure is shown in transparent cartoon, while noticeable regions involved in flexibility variations are shown in purple, and functionally important residues are shown in licorice representation.

Keywords: molecular dynamics simulations, protein crystal structure, X-ray diffraction, FXI, blood coagulation

REFERENCES

[1] R. Assenberg, PT Wan, S. Geisse, LM Mayr. "Advances in recombinant protein expression for use in pharmaceutical research." Curr Opin Struct Biol. 2013 Jun;23(3):393-402.

[2] M. Bech, O. Bunk, C. David, P. Kraft, C. Brönnimann, EF Eikenberry, F. Pfeiffer. "X-ray imaging with the PILATUS 100k detector." Appl Radiat Isot. 2008 Apr;66(4):474-8.

[3] C. Giacovazzo, HL. Monaco, G. Artioli and D. Viterbo “Fundamentals of Crystallography”. International Union of Crystallography Monographs on Crystallography.

[4] M. Karplus, GA Petsko Nature. “Molecular dynamics simulations in biology”. 1990 Oct 18;347(6294):631-9.

[5] J Emsley, PA McEwan, D Gailani. “Structure and function of factor XI.” Blood. 2010 Apr 1;115(13):2569-77.

[6] W. Humphrey, A. Dalke, K. Schulten. “VMD: visual molecular dynamics”. J Mol Graph 1996; 14: 33-38.

[7] JC Phillips, R. Braun, W. Wang, J. Gumbart, “Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD”. J Comp. Chem. 2005; 26:1781-1802.

[8] R. Caliandro, G. Rossetti, P. Carloni, “Local Fluctuations and Conformational Transitions in Proteins” J Chem Theory Comput 2012; 8: 4775-4785.

[9] G. Rossetti, X. Cong, R. Caliandro, G. Legname, P. Carloni. “Common structural traits across disease-linked variants of the human prion protein and their implications for familial prion diseases” J Mol Biol 2011; 411: 700-712.

[10] R. Caliandro, G. Nico, G. Tiscia, G. Favuzzi, V. De Stefano, E. Rossi, M. Margaglione, E. Grandone, “Structural Analysis of Protein Z gene variants in Patients with Foetal Losses”, Thromb. Haemost. 2013; 110.3: 534-542

[11] D. Di benedetto, G. Rossetti, R. Caliandro, P. A. Carloni “Molecular Dynamics Simulation-Based Interpretation of Nuclear Magnetic Resonance Multidimensional Heteronuclear Spectra of α-Synuclein•Dopamine Adducts” Biochemistry 2013 52, 6672-6683.

[12] R. He, D. Chen, S. He. “Factor XI: hemostasis, thrombosis, and antithrombosis.” Thromb Res. 2012 May;129(5):541-50