On the early events of the calcium induced activation of coagulation factor XIII-A_2 and tissue transglutaminase – an in silico study.

Attila Fekete[1,2,3 *], István Komáromi[2,4 †], Dániel Mucs[2,5,6]

[1]Governmental Information-Technology Development Agency, Budapest, Hungary
[2]Division of Clinical Laboratory Science, Department of Laboratory Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary
[3]Department of Organic Chemistry, Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary
[4]Research Group of Vascular Biology, Haemostasis and Thrombosis, Hungarian Academy of Sciences, University of Debrecen, Debrecen, Hungary
[5]Swetox, Karolinska Institutet, Unit of Toxicology Sciences, Södertälje, Sweden
[6]Unit of Work Environment Toxicology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[†]Passed away on 4th August, 2017.

***Author to whom correspondence should be addressed.**
Attila Fekete
Governmental Information-Technology Development Agency, Budapest, Hungary
Email: attila.fekete@kifu.gov.hu

Abstract

Both coagulation factor XIII-A_2 (FXIII-A_2) and tissue transglutaminase (TG2) play distinctive and important roles in homeostasis by crosslinking proteins or peptides via isopeptide bonds. In this present study, a series of microsecond long all-atom molecular dynamics (MD) simulations were carried out in order to reveal the dynamic, atomic-level events which may contribute to the activation of these proteins via the binding of calcium ions. In addition to previously conducted *in vitro* and crystallographic studies, further suggestions have been made concerning the calcium binding features of these enzymes. The different systems used for running the simulations were based on the zymogen, computationally cleaved and even the activation peptide (AP-FXIII) free FXIII-A_2' homodimer form. The effects of various ionic environments have also been explored in the simulations of FXIII-A_2. Our results suggest that the presence of calcium ions can cause increased AP fluctuations, which ultimately could lead to their relocation on the homodimer surface. The release of these APs seem to be crucial for rotation of the A subunits based on equilibrium MD simulations. The primary evidence for this assumption comes from the predicted principal component eigenvector what is considered as the first, large scale event of the overall activation process. To get a more accurate atomic-level description, the calcium binding sites of TG2 have also been investigated in our extensive *in silico* experiments, which suggests the presence of a previously unidentified binding sites as well.

KEYWORDS: blood coagulation factor XIII, FXIII-A_2, tissue transglutaminase, TG2, calcium binding, molecular dynamics simulation

List of Abbreviations

AP, AP-FXIII	Activation peptide
APs	Activation peptides
Ca^{2+}	Calcium ion
cFXIII	Cellular coagulation factor XIII
DCCM	Dynamic cross-correlation matrix
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
FXIII	Coagulation factor XIII
FXIII-A	Coagulation factor XIII, A subunit
FXIII-A_2	Coagulation factor XIII, A subunit homodimer

FXIII-A_2’	Inactive intermediate A subunits homodimer of coagulation factor XIII
FXIII-A_2*	Active A subunit homodimer of factor XIII, activated by thrombin plus Ca^{2+}
FXIII-A^O	Active A subunit of coagulation factor XIII, activated by Ca^{2+}
FXIII-A_2^O	Active A subunit homodimer of coagulation factor XIII, activated by Ca^{2+}
FXIII-A_2B_2	Coagulation factor XIII, heterotetramer
FXIII-B_2	Coagulation factor XIII, B subunit homodimer
KCl	Potassium-chloride
MD	Molecular dynamics
NaCl	Sodium-chloride
NMR	Nuclear magnetic resonance
NpT	Constant temperature, constant pressure, constant particle number ensemble
PCA	Principal component analysis
pFXIII	Plasma coagulation factor XIII
RMSD	Root mean square deviation
RMSF	Root mean square fluctuations
r_{gyr}	Radius of gyration
Sr^{2+}	Strontium(II) ion
TG2	Transglutaminase 2, tissue transglutaminase
Yb^{3+}	Ytterbium(III) ion

Introduction

Both blood coagulation factor XIII subunit A (FXIII-A) and tissue transglutaminase (TG2) are protein-glutamine gamma-glutamyltransferase enzymes (EC 2.3.2.13). In the plasma, FXIII (pFXIII) is circulating as a heterotetramer with two carrier B subunits (FXIII-A_2B_2), while its cellular form (cFXIII) exists as a homodimer (FXIII-A_2) in the cytoplasm of platelets and monocytes. The pFXIII requires proteolytic activation by thrombin (Lorand & Konishi, 1964) which cleaves the peptide bond between Arg37 and Gly38 of FXIII-A_2. This cleavage then provides the activation peptides (here denoted as: APs) and the FXIII-A_2' (Takagi & Doolittle, 1974). Crystallographic studies suggest that the cleaved N-terminal peptides remain attached to the surface of the bovine α-thrombin activated homodimer form (Yee, Pedersen, Bishop, Stenkamp, & Teller, 1995), blocking the entrance to the cavities where the catalytic residues are located (Cys314, His373, Asp396) (Pedersen et al., 1994). It has been previously noted that pFXIII can be activated without proteolytic cleavage under non-physiological conditions using high Ca^{2+} concentration (≥ 50 mM) (Credo, Curtis, & Lorand, 1978). Whereas the cellular FXIII (cFXIII), derived from placenta, can be activated either proteolitically by thrombin and low Ca^{2+} level or with high sodium-chloride (NaCl) or potassium-chloride (KCl) concentration in the presence of 2 mM Ca^{2+}. Even in the absence of any Ca^{2+} a slow progressive activation can still be observed (Polgár, Hidasi, & Muszbek, 1990). Based on resolved active TG2 structure (Pinkas, Strop, Brunger, & Khosla, 2007) large conformational changes were predicted to be able to obtain the enzymatically active FXIII-A_2* form (Komáromi, Bagoly, & Muszbek, 2011). Such major domain motions were published by Stieler et al. where the recombinant cellular FXIII (rcFXIII) was activated without proteolytic cleavage (FXIII-A^o) using 300 mM sodium-chloride and 3 mM Ca^{2+} concentration. The displacement of β-barrel 1 and β-barrel 2 domains caused no major changes in the structure of the other two domains, however the β-barrels were significantly displaced providing an accessible catalytic cavity for substrate molecules (Stieler et al., 2013). Nevertheless, domain movements in this structure differs from those that were suggested based on the analogy with TG2 (Komáromi et al., 2011). In summary, calcium ions play a role in the dissociation of FXIII-B_2 subunits from FXIII-A_2 and also to induce conformational activation of FXIII-A_2. It was suggested that FXIII-A dimers are responsible for Ca^{2+} binding (K_m= 0.12 mM, based on fluorescence measurement) and were estimated to bind 1.2-1.5 calcium ions based on equilibrium dialysis experiments (Lewis, Freyssinet, & Holbrook, 1978). The hypothesis also proposed the existence of two strong binding sites per symmetrical dimer, along with some other weaker binding sites. Crystallographic information can be found in the literature for the location of Ca^{2+} binding sites and even for binding Sr^{2+} or Yb^{3+}. Two symmetrical Ca^{2+} and Sr^{2+} binding sites can be found near residues Ala457,

Asn436, Asp438, Glu485 and Glu490. It is worth noting, that in the crystallographic structure published by Fox et al., there are eight Yb^{3+} ions present (Fox et al., 1999) with a total of four Yb^{3+} bound between the Asp270 and Asp271 residues of both monomer units. Interestingly, this is not one of the known calcium binding sites, however a recently published active FXIII-A structure also appears to contain a single Ca^{2+} bound to this site (Stieler et al., 2013). Site-directed mutagenesis studies were carried out in order to validate the importance of the main binding site Glu485 and Glu490 in the activation process and whether residues Asp476 and Asp479 play a role in the conformational changes indirectly via a switch between the protease-sensitive and protease-resistant form (Lai, Slaughter, Peoples, & Greenberg, 1999). Surface polarity analysis and ^{43}Ca NMR experiments also suggest that weaker binding sites exist and the dissociation constant of calciums for the zymogen FXIII-A_2 is 0.51 mM. However, based on NMR experiments, in the presence of thrombin the Kd increases significantly (5.9 mM) (Ambrus et al., 2001). Due to claims regarding the limitations of NMR techniques, hydrogen/deuterium exchange experiments were conducted in order to examine the percentage of deuteraion of four different FXIII-A_2 systems (zymogen, thrombin-cleaved, thrombin cleaved + 1 mM Ca^{2+}, and the zymogen activated by 50 mM Ca^{2+}) using MALDI-TOF mass spectrometry, which resulted in the identification of several regions with elevated deuteration (Turner & Maurer, 2002). The *in vitro* activated zymogen FXIII-A^o structure has been published by Stieler et al. in the presence of a covalent inhibitor and using high ionic strength. Within this study they have found that β-barrel 1 and β-barrel 2 undergoes remarkably large movements and exposes its active site (Stieler et al., 2013). More recently a detailed structure/ function study has been conducted in which the generated pathway between the active and inactive conformations were studied and the interaction between the A and B subunits was also modelled (Gupta et al., 2016). More detailed information can be found about the structure and function of FXIII in numerous other relevant publications (Komáromi et al., 2011; Lorand, 2001; László Muszbek, Bereczky, Bagoly, Komáromi, & Katona, 2011).

Tissue translgutaminase (TG2 or Gh protein), another member of the TG group of enzymes, is present in different cellular compartments including the plasma membrane, cytosol and the nucleus, although only as a monomer. It is common for both TG2 and FXIII-A that activation only occurs in the presence of Ca^{2+}, however structural evidence is available in the absence of Ca^{2+} and presence of GDP (Liu, Cerione, & Clardy, 2002), GTP (Jang et al., 2014) or ATP (Han et al., 2010). The same lack of observed Ca^{2+} has been reported with the active conformation as well (Pinkas et al., 2007). TG2 possesses GTPase activity and the binding of guanosine-phosphates regulates the transamidation activity in a negative way. Interestingly, based on equilibrium dialysis experiments, TG2 can bind up to six Ca^{2+} which can decrease with the binding of GTP (Bergamini, 1988). Modest sequence homology can be observed between the eight known human TGs (Stieler et al.,

2013), but they share a highly conserved secondary and tertiary structure. Using the known structure of FXIII-A_2 (V. C. Yee et al., 1994), as a base template for homology modelling, 50 ps unconstrained MD simulations were carried out in the absence of explicit solvent molecules both with and without Ca^{2+} (3 pcs.). In this, it was suggested that the short simulation time and low stoichiometric ratio were the main causes for the lack of large scale movements, although both simulations predicted that the added calciums would increase the gyration radius. When digesting the enzyme in limited proteolysis experiments, it was found that only the loop region between the catalytic and the first β-barrel domain has been cleaved (Casadio et al., 1999). Results of ^{43}Ca NMR investigations suggested the presence of some low affinity sites, more interestingly the Kd was somewhat similar to the thrombin activated FXIII-A_2 value (6.0 mM versus 5.9 mM), hence these two enzymes probably follow a similar activation mechanism (Ambrus et al., 2001). In a detailed *in vitro* experiment five calcium binding sites were identified following systematic mutations of certain residues within the putative binding sites (Király et al., 2009). These sites were established based on computer modelling and structural analogy with FXIII-A and epidermal translgutaminase (TG3). It should be noted that the 1KV3 structure (Liu et al., 2002) contains a glycine residue in the position of 224 while its native form contains valine in the very same position (https://www.uniprot.org/uniprot/P21980). As of now, there are no known structures available of the native closed TG2 except 4PYG (Jang et al., 2014), which has the native Val224, but contains three other point mutations, and the reasons behind these were not discussed in the original article. Kanchan et al. pointed out, that there are several differences between the calcium affinity, GTPase activity and other measures of these different TG2 forms (Kanchan et al., 2013). The open conformation of TG2 was resolved by Pinkas et al. via incubating the enzyme in 150 mM NaCl, 10 mM Ca^{2+} and in the presence of an irreversible, modified-pentapeptide inhibitor (Pinkas et al., 2007). A recent review concerning the protein-protein interactions of the TG2 can be found in the literature (Kanchan, Fuxreiter, & Fésüs, 2015).

In summary, coagulation factor XIII and TG2 represent well studied systems, however, several details of their calcium induced activation process still remains unclear. In this present work a series of microsecond long all-atom MD simulation were conducted on both the zymogen and on the computationally cleaved FXIII-A_2', both in the presence and absence of Ca^{2+} cofactors and also using activation peptide free proteins. Individual and combined effects of Ca^{2+} and GDP were also considered in the study of TG2, then the outcomes of these MD simulations were compared to existing *in vitro* experimental results. The effect of the Gly224Val mutation on calcium binding affinity and overall protein dynamics has also been investigated within this study.

Materials and methods

The simulation systems

In order to examine the effect of Ca^{2+} on the dynamics of FXIII-A_2, the following simulation sets were assembled. The first simulation set (i) was used to study the zymogen enzyme via the calcium bound form of FXIII-A_2 (i/a) (PDB ID: 1GGU (Fox et al., 1999)) and in addition to the two bound calcium ions, the total Ca^{2+} concentration was set to 14 mM. A second zymogen model was created from the 1F13 PDB structure (i/b) (Weiss, Metzner, & Hilgenfeld, 1998) and a total of 50 mM Ca^{2+} was added into the simulation cell. To clarify the important structural features of the thrombin treated FXIII-A_2', the peptide bond between Arg37 and Gly38 was cleaved using computational methods and from these, four different model systems were constructed. The ii/a and ii/b model systems were analogues of i/a and i/b, respectively. The ii/c model was similar to ii/b, but contained only 14 mM Ca^{2+} (practically 20 calcium ions), while the final model of this set of four contained only the neutralized protein with a total of 150 mM NaCl (ii/d) and was used as a reference. We have suspected that after the proteolysis of the APs, the next step in the activation process should be the relocation or at least partial dissociation of the APs in the presence of calcium. The simulations in the absence of APs (FXIII-A_2') were carried out under different conditions; with no calciums present at all (iii/a), with 14 mM Ca^{2+} (iii/b) and with 1000 mM Ca^{2+} (iii/c), in order to examine the protein dynamics under extreme ionic strength. The reason why we chose ten times higher calcium concentration than *in vivo* (14 mM vs. 1.4 mM) was partially due to computational time limitations. To address this, we have increased the number of interacting calcium ions in order to reveal all possible binding sites within time frame, including previously hypothesized weak ones. Further reinforcement for using 14 and 50 mM Ca^{2+} concentration came from the results of MALDI MS experiments where these concentrations were sufficient to detect changes in the percentage of deuteration (Turner & Maurer, 2002).

In the fourth simulation set individual and combined effects of calciums and guanosin di- or triphosphates were studied on TG2 systems. In the reference simulation the bound GDP was removed and enzyme solvation was achieved in the presence of 150 mM NaCl (iv/a), and also containing 8 mM Ca^{2+} (iv/b), which corresponds to 10 calcium ions. We have also repeated these two simulations in the presence of GDP (iv/c,d) as well. Due to the suspected importance of Gly224, the final model of this set (iv/e) was based on GTP bound TG2 in the presence of 6 mM Ca^{2+} (practically 8 calcium ions).

Table 1. Details of FXIII-A_2 and TG2 simulation sets.

All simulations of iv/a-d were based on the 1KV3 (Liu et al., 2002) crystallographic structure, however iv/e was based on the 4PYG X-ray structure (Jang et al., 2014). The final simulation set (v) contained the open conformation of the FXIII-A^O monomer (PDB ID: 4KTY) with three bound calcium ions (v/a) (Stieler et al., 2013) and the calcium free open conformation of TG2 (v/b) (Pinkas et al., 2007). Both of these model systems contained 14 mM Ca^{2+} concentration and were considered to be interesting cases, however they did not couple closely to any part of this present study. Throughout the work we have numbered the residues according to existing literature (*id est* the first methionine was omitted in the case of FXIII-A but not in the case of TG2) and the recommended abbreviations of Muszbek et al. were used for factor XIII (L. Muszbek, Ariëns, & Ichinose, 2007), where it was possible.

Molecular dynamics simulations

The protein models were solvated using explicit TIP3P water molecules in octahedral boxes with the closest distance between the box and the protein being 12 Å. In the next step all systems were neutralized and the NaCl concentration was set to 150 mM/dm^3. After a short energy minimization step, a 2 ns long simulated annealing equilibration was carried out where the minimized systems were heated up to 310 K and the protein heavy atoms were kept restrained by a force constant of 1000 $kJmol^{-1}nm^{-2}$. Following the equilibration procedure, model system dependent, one or two microsecond long (Table 1) isobaric-isothermal (NpT) production runs were performed using periodic boundary conditions with the aid of virtual sites that allowed a 4 fs step size for integrating the Newtonian laws of motions. All bonds were constrained using the LINCS algorithm (Hess, Bekker, Berendsen, & Fraaije, 1997). In simulations where guanine di- or triphosphate was present, a 2 fs step size was applied and these simulations lasted 1 µs. A cut-off of 10.0 Å was used for the Lennard-Jones and short-range electrostatic interactions and a force-switch was applied to smoothly switch forces between 7.0 Å and 10.0 Å. The long-range electrostatics interactions were calculated by Particle Mesh Ewald summation. For the coupling of thermal bath the velocity-rescaling method was used (Bussi, Donadio, & Parrinello, 2007), while pressure was regulated with the isotropic Parrinello-Rahman method (Parrinello & Rahman, 1981). The Amber99SB-ILDN-NMR (Hornak et al., 2006; Li & Brüschweiler, 2010; Lindorff-Larsen et al., 2010) force field (Hornak et al., 2006; Li & Brüschweiler, 2010; Lindorff-Larsen et al., 2010) was used throughout the simulations since it was successfully used in the study of calcium and magnesium binding features of small antifungal proteins (Fizil et al., 2018). In the systems where

TG2 was complexed with GDP or GTP the AMBER GAFF parameters (Meagher, Redman, & Carlson, 2003) were used for the substrates. A total of 21 μs all-atom MD simulations were carried out in this study to be able to examine the details of the early events of calcium induced activation of human TGs.

In two cases (ii/c and iv/d) an additional 50 ns NpT MD simulation was carried out with the Amber ff14SB force field (Maier et al., 2015) and using multi-site Ca^{2+} models (Saxena & Sept, 2013) in order to simulate more precisely the already established binding sites. The starting structures were extracted from the very end of the corresponding simulations and all bound calcium ions were kept as is. If any virtual interaction sites were present, they were removed. The XZ (central) atom of the multi-site Ca^{2+} model was aligned to all of the bound calciums. The reconstructed systems were solvated in an octahedral box filled with TIP3P water molecules and the final NaCl concentration was set to 150 $mmol/dm^3$. A 10.0 Å cut-off was used over the initial energy minimization step, the 1.7 ns long (3-stage) equilibration protocol and the 50 ns long production runs.

The MD simulations were performed using GROMACS 5.1.4 (Abraham et al., 2015; Páll, Abraham, Kutzner, Hess, & Lindahl, 2015) and with the pmemd software of the Amber16 package (Case et al., 2005; D.A. Case et al., 2016; Salomon-Ferrer, Case, & Walker, 2012). The trajectories were analysed with the software tools of GROMACS and the cpptraj program (Roe & Cheatham, 2013). The dynamic cross-correlation matrices (DCCM) (McCammon, 1984) were calculated over 12.500 frames with the aid of the Bio3D v2.3 R package (Grant, Rodrigues, ElSawy, McCammon, & Caves, 2006; R Core Team, 2013). R was used to perform all matrix operations as well. For visualization of the protein structures either UCSF Chimera 1.11.2 software (Pettersen et al., 2004) or Visual Molecular Dynamics (VMD) 1.9.1 (Humphrey, Dalke, & Schulten, 1996) was used. All unresolved segments of the protein structures were reconstructed with MODELLER 9.10 and only those models were used which had the best DOPE score (Šali & Blundell, 1993) via the graphical unit interface of UCSF Chimera (Yang et al., 2012). It should be noted that the conformation of the missing loop region of 1GGU (between Thr508 and Ser516) was found to have an overall importance, therefore, it was remodelled as it can be found in the 1F13 structure with the loop of the A chain of 1F13 placed to the B chain of 1GGU and vice versa. All of the data that can be seen in this present work was plotted by various in-house written Python scripts using matplotlib (Droettboom et al., 2018) and numpy (van der Walt, Colbert, & Varoquaux, 2011). The protonation states of titratable residues were predicted using the H++ webserver (Anandakrishnan, Aguilar, & Onufriev, 2012; Gordon et al., 2005; Myers, Grothaus, Narayanan, & Onufriev, 2006) at pH=7.4, however, it should be noted that all of the aspartate and glutamate side-chains of the TG2 systems were predicted to be deprotonated (i.e. negatively charged).

Analysis of molecular dynamics simulations

Frames of the dynamic trajectories were saved every 80 ps and contained the corresponding protein plus Ca^{2+} ions and any GDP or GTP, if any were present. For the analysis of root mean square deviation (RMSD), root mean square fluctuations (RMSF) and radius of gyration (r_{gyr}), the heavy atoms within the protein were used. In order to ensure that the N-terminal activation peptide does not interfere our conclusions, the RMSD and r_{gyr} values were calculated for residues between Gly38 and Met731 in the case of FXIII-A_2 simulations. For dimers, the per-residue averaged values of RMSFs were calculated per individual chain and then averaged. For the calculation of DCCM matrices the coordinates of Cα atoms were used which yielded symmetrical N x N matrices. Since the interpretation of correlation matrices can be quite difficult in such a large proteins, we have tried to focus only on the difference DCCM matrices (($c_{ij}(2)$-$c_{ij}(1)$, where $c_{ij}(1)$ was taken as a reference)), however all reference matrices are also included in the SI. The Ca^{2+} binding properties were based on the calculation of distance between the Cγ atom of aspartate or the Cδ atom of glutamate and any Ca^{2+}. The large amplitude low frequency motions were extracted from the Cα trajectory for principal component analysis and the first 10 eigenvectors were calculated via the ProDy software (Bakan et al., 2014; Bakan, Meireles, & Bahar, 2011). For helping the interpretation of the domain motions of monomers relative to each other, the centre of masses of main structural domains were used to calculate a pseudo torsion angle and the standard deviations were also indicated in each case.

Results and discussion

Results of molecular dynamics simulations

Simulations of FXIII-A_2

Due to the complex nature of the activation mechanism of FXIII-A_2, we followed a step-by-step approach to clarify the key events which may contribute to an enzymatically active protein conformation and can also be related to the binding of calcium. This meant starting from zymogen structures (i/a,b) through proteolitically activated ones (ii/a-d) and finally simulating the AP free proteins (iii/a-c). Since it was the very first instance where FXIII-A_2 was studied by the means of long MD simulations, we have primarily assessed the general dynamic properties (RMSD, r_{gyr}, RMSF) of these runs.

The root mean square deviations (Fig. 1) of zymogen and proteolitically activated FXIII-A_2 models suggest that the presence of calcium ions in general increases the heavy atom RMSDs in every case, compared to the reference simulation (ii/d). More importantly, on the timescale studied, the high level of Ca^{2+} concentration (50 mM) does not appear to influence the dynamics significantly, although it should be noted that these simulations converge much slower towards the ~2.5 Å final RMSD value, than any of the other ones. Therefore, we cannot exclude the importance of high calcium concentration on a much longer timescale since the increased fluctuations of APs are clearly demonstrated (see later).

The reason why dynamics of simulation sets *i* and *ii* are quite similar, lies in the presence of the APs. Independently from the fact whether the APs are cleaved or not, it seems that their main role is to keep the monomer subunits together. Since the fluctuations of the first 37 residues are significantly smaller in the absence of calcium ions, it is highly probable that calciums play an essential role in releasing the APs. Two arginine residues (Arg11, Arg12) play a critical role in holding the APs in their place, as it was found in earlier studies (László Muszbek et al., 2011). These residues are located in a relatively deep cavity and several acidic residues (Asp343, Asp345, Asp367 and Glu401) hold them in place via salt-bridge contacts. We assume that full dissociations of the APs are rare events on the MD timescale, if they are possible at all, but in the ii/c simulation we can observe a slight relocation of an arginine contained within one AP, however it does not dissociate completely. The added calcium ions increase the fluctuations of N-terminal peptides significantly (Fig. 2), contributing directly to the release of APs, however other additional requirements cannot be excluded at this stage. The 2 μs long simulations without APs (iii/a-c) show that the RMSDs increase in every scenario even in the absence of calcium ions. The detailed effect of extremely high calcium concentration will be discussed in a different section.

In the context of gyration radius (Fig. S2), all of the studied zymogen and cleaved models represent almost identical final values, with the average radius of gyration of these simulations being 36.90 ± 0.1 Å. The RMSD and r_{gyr} were also calculated for each monomer chain and can be found in the SI (Figure S1, S3).

Notable regions with high local RMSF are shown on (Figure S4-S6) and generally speaking these regions are loop segments or located in close proximity to the terminals of the AP. Nevertheless we would like to highlight the possible importance of four regions which may have distinguished significance. These regions can be found between the residues of Val274-Asp280, Met350-Trp370, Ile440-Ile460 and Pro505-Arg515. Other than the first segment, all of the others are situated in close proximity to each other. The loop regions between Val274-Asp280 show moderate fluctuations except for the third simulation set, where the RMSF values are significantly higher. The second region is also in contact with sixteen residues of the other subunit N-terminal

and a helix can be found at the N-terminal end of the last region. This helix contains the residues of the main Ca^{2+}-binding site, hence an allosteric pathway can be drawn from the main binding site to the AP. Since the residues between Pro505-Arg515 are usually poorly or non-resolved at all in crystallographic structures, even in the 1F13 structure the subunits show different conformations in these regions, one can obviously associate a regulatory function to these segments. The Met350-Trp370, Ile440-Ile460 regions are organized in three adjacent anti-parallel β-strands and one disordered loop. Despite contradictions in the details of the proposed activation mechanisms (Komáromi et al., 2011; Stieler et al., 2013) the above mentioned segments should be equally important in the movement of β-barrel 1. Simulation iii/c corroborate this theory, since the local fluctuations are higher than in simulation i and ii and also a special monomer movement can be observed within this system as well.

Our findings here point to several valid conclusions; (1) the presence of calcium ions generally increases the RMSD values compared to the reference simulation; (2) in the absence of APs slightly higher RMSD can be observed, however extremely high Ca^{2+} concentration causes a significant increase in RMSD values; (3) the fluctuation of the N-terminal part of the APs were greatly influenced by the presence of calciums.

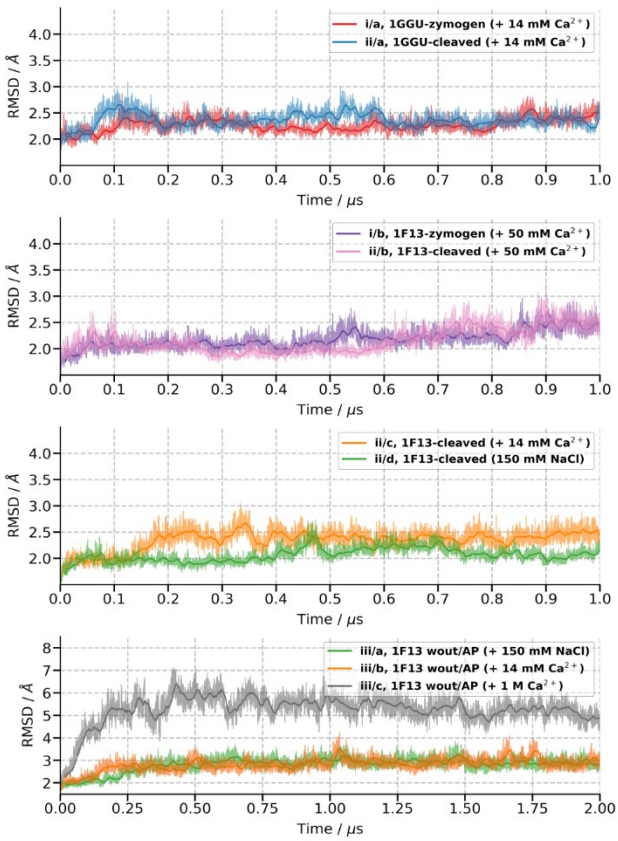

Figure 1. Heavy atom root mean square deviation during FXIII-A_2 simulations. (The activation peptide was omitted in the calculation of RMSD in order to have comparable results.)

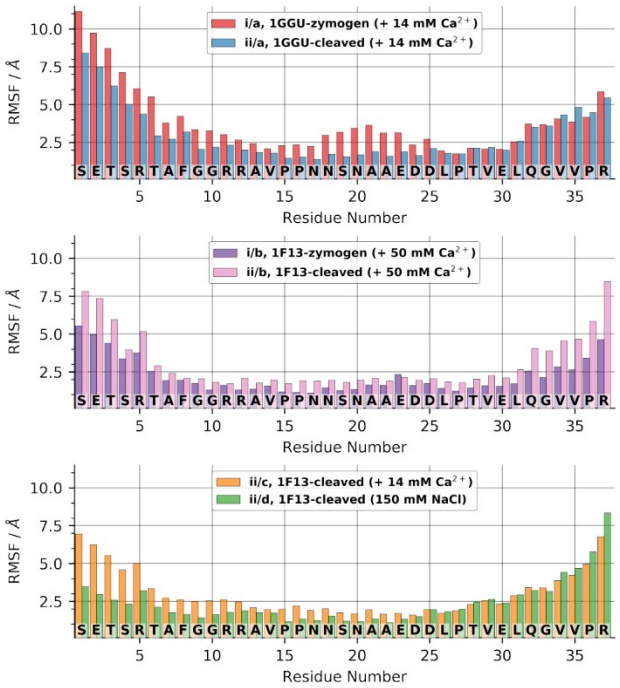

Figure 2. Per-residue averaged root mean square fluctuation of the activation peptide heavy atoms during FXIII-A_2 simulations.

Simulations of TG2

Following a similar detailed approach for TG2 as in the case of FXIII-A_2, not only the individual effects of Ca^{2+} ions and guanosine-phosphates were studied, but also their combined effects via time-dependent properties.

Here we cannot observe notable differences in RMSD and r_{gyr} values between iv/a and iv/b simulations (both without GDP), however in the presence of GDP we can see clear differences in RMSDs which follow the iv/a ~ iv/b > iv/d > iv/c trend (Fig. 3(A)). These results can be elucidated by the GDP-bound forms being *in vivo* inactives and the smaller RMSDs can also be the sign of the conservative behaviour of the inactive form. It should be noted here that the added calcium ions obviously increase the RMSD, indicating their significance in creating an eventually active (open) TGase conformation. The r_{gyr} values in this case are much higher suggesting that the parallel occurrence of the Ca^{2+} ions and the GDP leads to different dynamics. The simulation using the 4PYG structure as a starting point (iv/e) also supports the hypothesis that has been stated in the case of iv/d. It is worth pointing out that after 650 ns the RMSD values of iv/e show an elevation, probably since the initial structure contains three point mutations (Gln186Glu, Thr533Asn, Val655Leu) which influence the dynamics in condensed phase. The r_{gyr} values do not indicate significant changes, i.e. the protein remains stable and there are no signs of unfolding (Fig. 3(B)).

Due to the highly conserved secondary and tertiary structure of the inactive (closed) FXIII-A_2 and TG2 we can observe similarities in RMSF values (Fig. S7). The earlier highlighted regions of FXIII-A_2 correspond to the following residues in the TG2 structure: Ile313-Asn333, Val401-Val422 and Asn460-Thr471 (according to the TG2 residue numbering). Indeed, these regions seem to share many common features with the FXIII-A, thus we focus on the fine details between the TG2 simulations by looking at the RMSF differences between the corresponding sets.

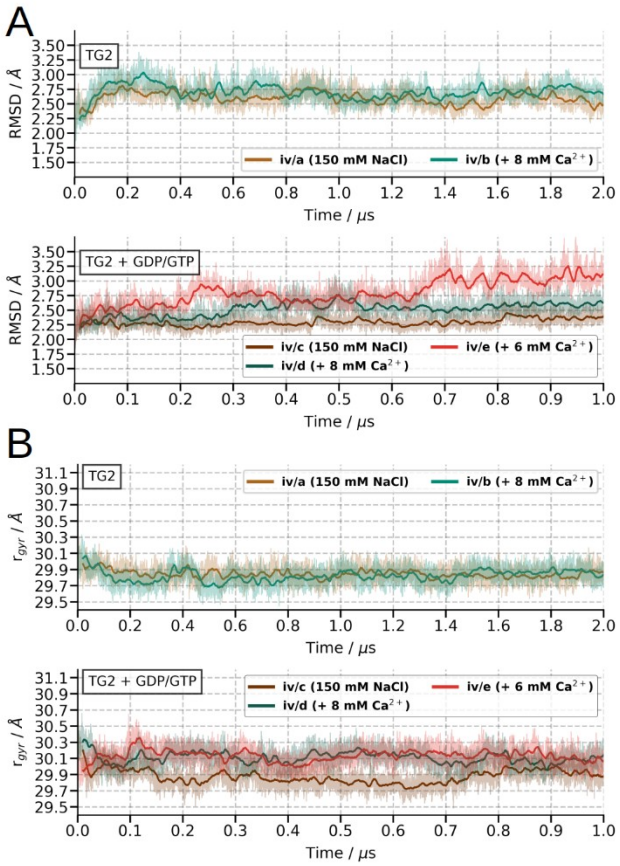

Figure 3. (A) Heavy atom root mean square deviation and (B) heavy atom radius of gyration during TG2 simulations.

The graphical representation of the RMSF differences is shown in Figure S8. As it can be seen, the largest differences affect the β-sandwich domain, however these discrepancies can be attributed to regions with no known importance, such as Leu12-Lys30, Gly64-Gly72, Leu79-Asp87. The region between Arg240-Tyr245 shows high fluctuation in the presence of Ca^{2+} (Fig. S8(A)), although importance of this segment is not fully understood. More importantly, the active site Cys277 is concealed by this loop and a binding site can also be found near Asp232 and Asp233 (Asp270 and Asp271 in FXIII-A_2). Since this loop most likely adapts an ordered conformation in the active TG2 and FXIII-A_2 as well, it is possible that it can control the opening of the active site

during activation not only in TG2, but also in FXIII-A_2. The loop between Thr343-Glu352 also shows high fluctuation, but only in the absence of calcium and GDP (Fig. S8, panel A, D). In the presence of calcium ions this loop is fixed by a bound calcium, which seems like an acceptable interpretation of the observed high fluctuation (iv/a).

Concerning the loop region between Asn460-Thr471, we can provide the following plausible explanation. Despite the residues of the main binding site in FXIII-A_2 (Asp438, Gly457, Glu485, Glu490) being almost perfectly conserved in TG2 (Asp400, Ser419, Glu447, Glu452), we can only assume, based on the above mentioned identities, that this site would actually bind any calcium. Since the β-strand in FXIII-A (Thr449-Ile460) is partially disordered in TG2 (Ser411-Val422), this assumption seems highly plausible indeed. On the other hand, the space available in TG2 is much wider than in FXIII-A_2. Another possibility could be the formation of a binding site under the helix which pulls it downward, hence strengthening the Asn460-Thr471 loop, just like bending a note on a guitar string. The fluctuations only increased when there were calciums added to the simulation box, meaning that the C-terminal end of this loop is connected to the β-barrel 1, hence on a much longer timescale it may initiate more significant rearrangements, such as pulling down the first β-barrel and turning both barrels upside down. This mechanism would be completely in line with the “swiss army knife” theory for the activation of TGs (Komáromi et al., 2011).

Correlated motions

Due to the complex and large structure of these studied systems, it was deemed necessary to inspect correlated motions within, in order to identify differences which may be in connection with the presence of calciums and large scale domain movements during the activation process.

By taking the differences between the zymogen (i) and the cleaved (ii) simulation sets of FXIII-A_2, the main common feature that clearly appears is the strong correlation of the β-sandwich and β-barrel 1 domains within the subunits. This positive correlation indicates the movement of the domains in the same direction, thus these motions are practically the same as represented by the principal component analysis in the next section. Taking the comparison of i/a and ii/a and also the i/b and ii/b simulations several important features can be observed (Fig. S11). As mentioned above, the region between the residues of Met350-Trp370 and Ile440-Ile460 are assumed to play an important role regardless of the activation mechanism, meaning that these regions should be involved in both suggested processes. Supporting this hypothesis, these segments show increased correlation with the β-barrel 1 and also with the very same segments of the other subunit in the cleaved structures, while in the zymogen structures they seem much less correlated (Fig. S11). Another interesting observation is that these regions also correlate significantly with the residues

between Gly210-Trp225. This region is located next to the C-terminal end of the β-barrel 1 and the N-terminal of β-barrel 2 and also shows somewhat weaker correlation to these in the zymogen models. The effects of bound calcium ions become evident when comparing ii/c and ii/d, where their correlation to the β-sandwich domains can mainly be attributed to the presence of these bound ions (Fig. S12(A)). A comparison of simulations starting from different PDB structures carried out in the presence of 14 mM Ca^{2+} can be seen in Figure S11(B). Last but not least, all of the notable correlations are present in the third simulation set as well and the extreme ionic strength causes even higher correlation between the β-sandwich domains and strong anti-correlation between the β-sandwich and the β-barrel domains (Fig. S13). These findings can predict the observed monomer rotation and are in good agreement with the results of the PCAs.

Despite the well-preserved tertiary structures of FXIII-A_2 and TG2, it is important to note that in the latter the β-barrel 1 is correlated with the β-sandwich and the β-barrel 2 seems to correlate rather with the core domain (Fig. S14), making it a unique feature of TG2. The second notable difference can be found in the β-strand regions between the residues of Ile313-Asn333 and Val401-Val422 (correspond to the Met350-Trp370 and Ile440-Ile460 in FXIII-A_2), where these segments correlate with β-barrel 1 and also with the core domain, while in the case of FXIII-A_2 we cannot observe this latter correlation.

Concerning the simulation of open transglutaminases (v/a,b) the cross-correlation matrices suggest a clear difference. Namely, in v/b (TG2, 2Q3Z) the β-barrel 2 is correlated with the β-sandwich domain and is strongly anti-correlated with the core and β-barrel 1 domains (Fig. S15, panel B). In the v/a (FXIII-A^o, 4KTY) we can observe exactly the opposite phenomena (Fig. S15, panel A). To our current knowledge, we emphasize that either these enzymes follow two completely different activation mechanisms or one of the open conformations was kinetically trapped in a transient state during the large scale motions.

Principal component analysis and the interdomain salt-bridge network

The large amplitude motions are usually difficult to visually detect even over longer MD simulations, therefore we have decided to examine these possible motions in another way. The calculated pseudo torsion angles of the dynamic centre of masses for corresponding domains can be seen in Figure 4. Most of these dihedrals are distributed closely around -40° with one exception. In the case of iii/c this value is approximately -70°, what means a further 30 degrees of rotation.

Interestingly, the principal component analyses of iii/a-c systems shows that the motion along the first eigenvector is in perfect agreement with our findings in the case of iii/c and with the domain correlations discussed before. The high Ca^{2+} concentration induced monomer rotation from

the MD trajectory also can be observed in the PCA analysis as the first eigenvector in each simulation (Fig. 5(A)). It is important to point out that in the presence of the AP, these eigenvectors are hardly present at all or are significantly smaller, which provides further evidence on the protective roles of APs. Moreover, these findings support the experimental results which suggest that a slow progressive activation occurs and also that added calciums increase the speed of activation for the zymogen FXIII-A_2. Since these motions are not visible in the presence of APs, it can be assumed that the relocation of the N-terminal part of the AP is necessary for the activation, which is in agreement with the enhanced fluctuations of these regions. The rotation takes place as a contrary motion of the subunits involved, where the axis perpendicular to the plane of the paper goes through an interdomain salt-bridge network (Fig. 5(B)). This salt-bridge network has been never discussed previously, and a graphical representation can be seen in Figure S14 with the characteristic distances also depicted. The Asp404, Asp427 and the Arg260, Arg408 residues of both subunits play a role in the formation of this salt-bridge network and despite the significant movement these contacts remain stable, excluding minor differences in the case of iii/c.

Moreover, it was found earlier that the mutation of Arg260 causes serious deficiencies, therefore the effect of this mutation was also investigated by computational means and then compared to the energies of optimized structures (Ichinose et al., 1998; Peyvandi et al., 2004). Our results suggest that after 1 μs simulation time the subunits do not suffer significant changes, despite the obvious fact that Cys260 was unable to establish any native contacts (Fig. S10). It is also possible that this mutations causes structural changes within the A subunit, hence after folding the native dimer could not be formed. Last but not least, an interesting difference can be observed between the sequences of FXIII-A and TG2. It has been identified that Arg408 is not conserved in TG2, but Cys370 can be found in this exact position, which can lead to the conclusion that the Arg260Cys can cause serious malfunction and the subunits are unable to assemble. Based on this analogy, the Cys370 could then be the reason why TG2 preserves its monomer structure *in vivo*. Structurally, Cys370 and Cys371 residues are important and are known as a disulphide switch which helps to stabilize the active (open) conformation of TG2.

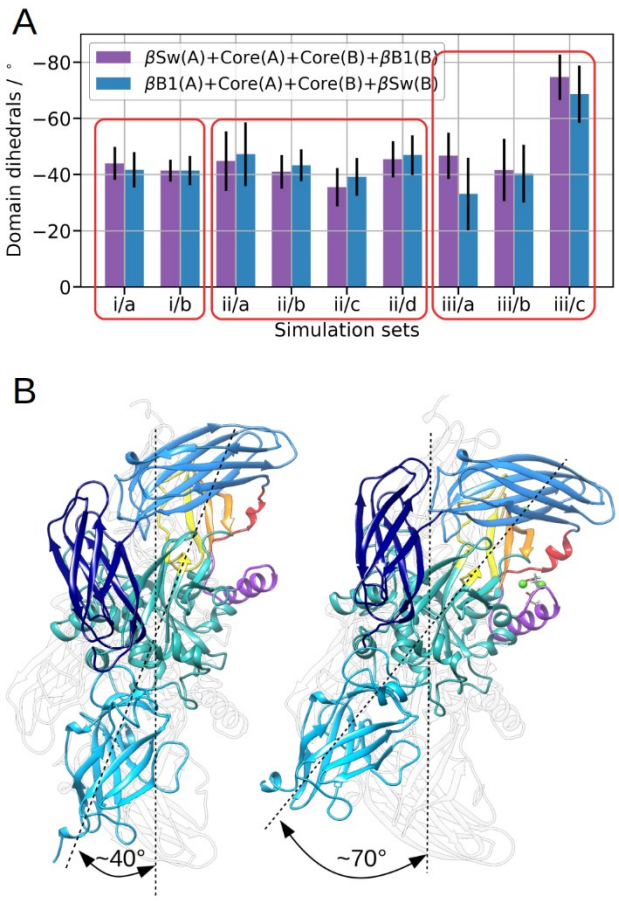

Figure 4. (A), the calculated pseudo torsion angles in the FXIII-A_2 simulations. The black vertical lines on the top of bars show the standard deviation. (B), result of the iii/c simulation (FXIII-A_2), the first frame of the trajectory (left side) compared to the last frame (right side).

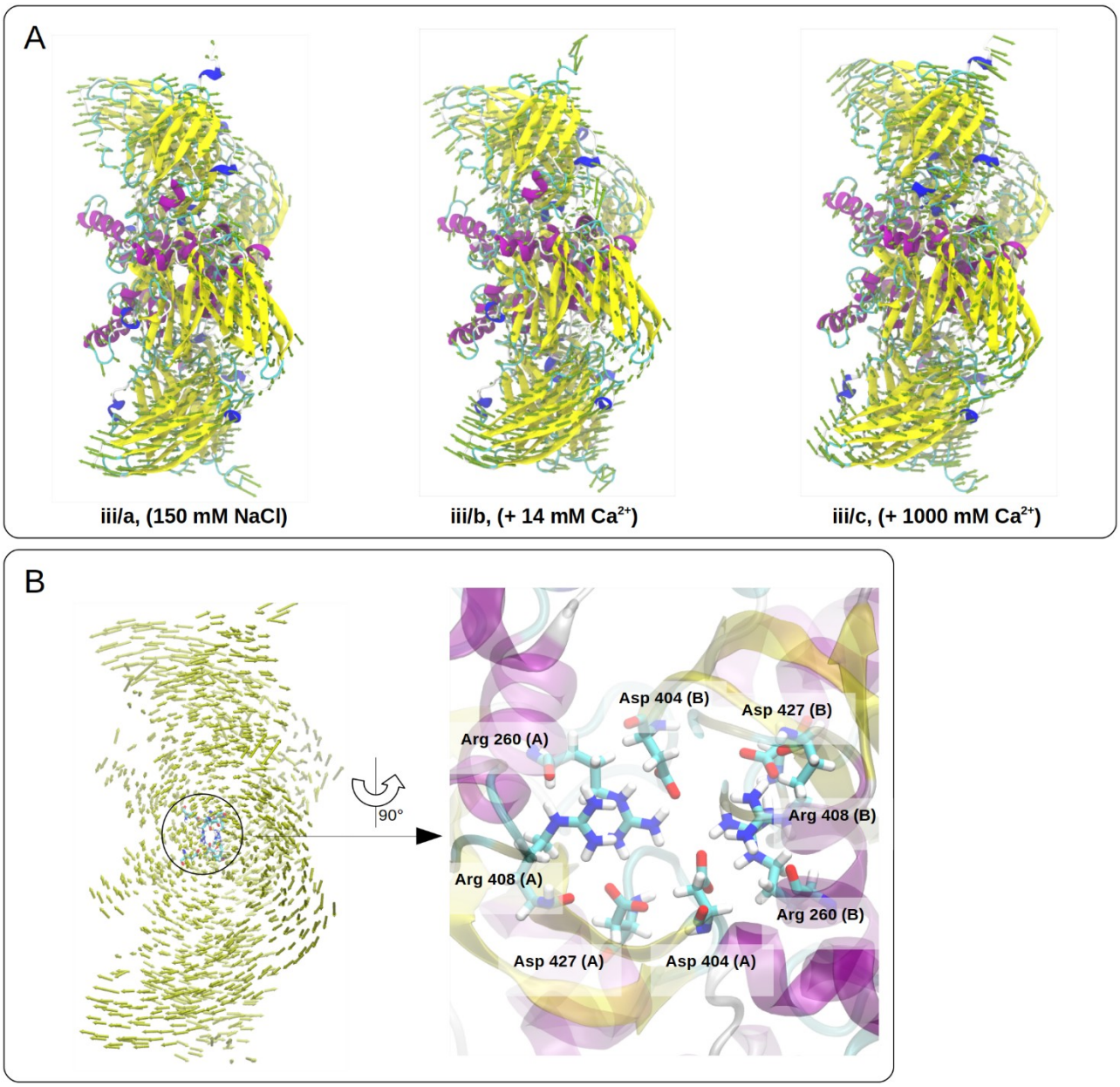

Figure 5. (A), the first principal component analysis eigenvectors of the C-alpha trajectories. (B), the central, inter-subunit salt-bridge network and its location. (The colouring of the secondary structure of FXIII-A_2 correspond to the structural elements (purple=α-helix, yellow=β-sheet, blue=3_{10} helix), the residues of central salt-bridge network was shown with sticks.)

Calcium binding

FXIII-A_2

It was known that FXIII-A_2 can bind one calcium per A subunit and several structures can be found in the Protein Data Bank that contain di- or trivalent cations bound to FXIII-A_2, eg. 1GGU,

1GGY, 1QRK, 1EVU (Fox et al., 1999). Based on these studies, factor XIII-A_2 can bind only one Ca^{2+} or Sr^{2+} per subunit in the very same binding site (Asp438, Gly457, Glu485, Glu490) and can accommodate up to eight Yb^{3+} in three binding sites (4 Yb^{3+} - chain A/B Asp270, Asp271, Glu272; 2 Yb^{3+} - chain B: Asp438, Glu485, Glu490; 1 Yb^{3+} - chain A: Asp438, Glu485, Glu490; 1 Yb^{3+} - chain A: Asp574, Glu585). It is worth noting that the 4KTY X-ray structure of the open FXIII-A^O also contains three calcium ions. One of them is located in the main binding site (Asp438, Gly457, Glu485, Glu490), the second bound by Asp270 and Asp271, and the third bound calcium was accommodated by the side chains of Asp343, Asp345, Asp351 and Asp367. The last one is buried in the closed (inactive) FXIII-A_2 by the APs, which could explain why it was unoccupied in the case of Yb^{3+} ions. On the other hand, these three occupied sites seem to be well conserved features, since they are also present in the TG3 structure (PDB ID: 1NUD). As the radius of the previously applied di- or trivalent cations are very close to each other, one can assume that beside the main site, the interfacial site of the homodimer (Asp270, Asp271 of both subunits) can also bind Ca^{2+} not only Yb^{3+}, and other sites can also exist.

For the analysis of Ca^{2+} binding properties of FXIII-A_2, our examination was driven by three main goals; (1) critically evaluate occupancy at the main binding sites of A subunits during the simulations; (2) validate the existence of weak binding sites proposed earlier and also if they exist, shed light on their locations; (3) investigate the impact of extreme high Ca^{2+} concentration used during *in vitro* activation.

Based on our simple MD simulations several possible binding sites have been identified, with most of them being formed by either spatially close or consecutive negatively charged side-chains. We have applied a selection criteria in order to reduce the number of possible binding sites and to remove temporary or weak binding modes. This meant that only those carboxylate carbon containing residues were collected which had a calcium ion within 4.5 Å and the ion also stayed within this threshold in more than 5 % of the total simulation time. The full numerical analysis of calcium binding can be found in the SI as an ASCII log file. Here we discuss primarily the results of ii/a (FXIII-A_2', 1GGU + 14 mM Ca^{2+}) and ii/c (FXIII-A_2', 1F13 + 14 mM Ca^{2+}) simulations and preferably underline those sites which are most likely to be conserved in FXIII-A, TG2, TG3 enzymes (Fig. 6). The detailed binding-event analysis can be found in the SI for each simulation systems (Fig. S16-S20).

Both ii/a and ii/c systems are equally important for understanding calcium binding. It is important to note, despite the initial structure of ii/c not containing any bound calciums, that after a few tenths ns of simulation time one of its main sites becomes occupied and the Ca^{2+} ion preserves its position throughout the whole simulation, while the other monomer does not bind any calciums at all in that place.

In a preliminary simulation, what was conducted under the very same conditions as in ii/c, both main sites were filled. In this particular case, the major site of chain A has been filled after 500 ns and after 900 ns both sites contained bound calciums (Fig. S19.). The protonation states of titratable residues were not predicted in this case, which ultimately lead to the binding of multiple calciums at the interfacial site (Asp270, Asp271 and Glu272 of both subunits).

The main sites of ii/a appear to bind the calciums quite tightly, even though the Pro505-Arg515 loops were remodelled based on how they were found in the 1F13 structure. Thus, we can assume that the loop conformation does not influence the calcium binding at all and, as it can be seen in Fig. S24, there is no clear connection between the secondary structures of these loops and the calcium binding events. Comparing the geometries of main sites in the ii/a and ii/c simulations, we can see an almost identical arrangement in their cluster representations (Fig. 7). Throughout the simulations the remodelled loop regions preserved their initial structure other than some minor changes (Fig. S24), as one of them became almost completely disordered while the other one still contained a one turn helix element. Based on this information it is unclear whether the loop regions have any other role over the *in vivo* activation or they are just in a fine equilibrium between the disordered and the one-turn helix arrangements.

In the i/b and ii/b simulations those major sites were occupied, all of which belongs to chain B, while in the simulation system of the AP free protein model (iii/b) the main site of chain A was found to bind one calcium ion. The extreme high calcium load (iii/c) caused the binding of at least one calcium ion per binding site in both A subunits. Thus, a straight conclusion cannot be drawn concerning these seemingly random preferences based on these relatively long equilibrium MD simulations.

In order to study this “blind spot” in great detail, 17 simulations were performed with different initial seeds, starting from the last frame of ii/c by keeping the bound calciums and resolvating the system in the presence of 100 mM Ca^{2+}, each one lasting 25 ns. In only one out of the 17 simulations the second site has become filled after 11 ns of simulation time (Fig. S23). We assume that this latent preference comes from the stochastic nature of the simulations and the random ion placement.

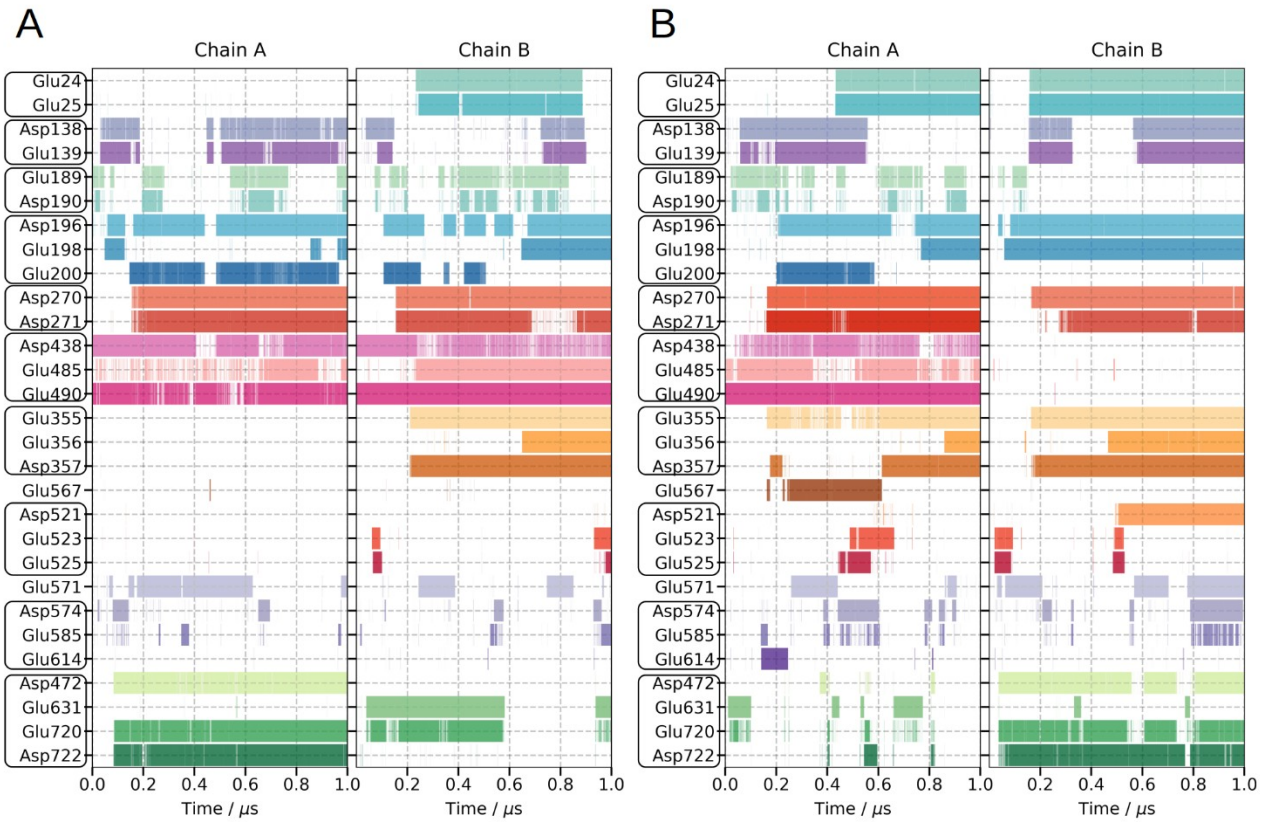

Figure 6. Calculated calcium binding events during the (A) ii/a (1GGU, cleaved APs, + 14 mM Ca^{2+}) and (B) ii/c simulations (1F13, cleaved APs, + 14 mM Ca^{2+}). (The coloured thin vertical bars represent the binding events as a function of simulation time in every row. The frequent binding events form continuous horizontal stripes which can be considered as a stronger binding at the corresponding residue. Different colours were applied in order to help the interpretation and comparison of results.)

The second binding site which became occupied quite frequently in our simulations, formed by Asp270 and Asp271 of the monomers and mentioned earlier as an interfacial site, is known to be able to accommodate Yb^{3+} ions while these aspartates also bind the Ca^{2+} ions in the 4KTY (non-proteolitically activated) structure (Stieler et al., 2013) of factor XIII-A and in the transglutaminase 3 (TG3) (PDB ID: 1NUD) (Ahvazi, Boeshans, Idler, Baxa, & Steinert, 2003). Our results suggest that in this site at least one bound calcium ion can always be found. The prediction of protonation states suggests that the sidechain of Glu272 is possibly protonated and, dependent on its protonation state, is able to bind up to two calcium ions. Despite our significant efforts, the role of this site is still not fully understood, although these loop regions are quite flexible and therefore can easily adapt a conformation that can be seen in the 4KTY or 1NUD structures. It is important to note that this region is situated very close to the central salt-bridge network, so we cannot exclude its regulatory role over the whole activation process. Also, these sites are directly connected with the

loops between Ser278-Gly294 of the A subunits, which serve as the boundary between the active site and the partially neighbouring β-barrel-2.

The third calcium binding site, which is also present in 4KTY and in 1NUD, can be found under the activation peptide and is formed by the side-chains of Asp343, Asp345, Asp351 and Asp367. This site was not easily accessible while the AP was in its place and the Asp343, Asp367 residues were in direct contact with the guanidium group of the side-chains of Arg11 and Arg12, therefore binding was not possible at this site. However, in the iii/b simulation a calcium can be found near the Asp343 and Asp345 of the A chain, with the dimer structure remaining intact and the other two residues being further away, so additional changes would be required for binding to occur at this site. As mentioned above, the extreme high calcium concentration (iii/c) induces notable changes in the overall position of the A subunits compared to each other. It seems that in both main chains the site is formed after 250 ns of simulation and additionally almost all of the possible anionic surface sites became occupied (Fig. S18(B)). Sites with reduced affinity (Glu355, Glu356, Asp357; Asp521, Glu523, Glu525; Asp574, Glu585, Glu614; Asp472, Glu631, Glu720, Asp722) may also be important in the non-proteolytic activation. We suggest that the reason why extreme high ionic strength is able to cause large scale monomer rotation is probably due to the breakdown of the electrostatic interactions between the A subunits. On the other hand, it remains unclear what causes the separation of the β-barrels from the catalytic core domain. In fact, the site formed by Asp472, Glu631, Glu720 and Asp722 represents an interdomain contact between the β-barrel 2 and the core domain.

An important question could be whether the simplified molecular mechanics parameters of divalent cations are able to predict the geometries of binding sites appropriately? To try and answer this question, we have decided to continue the ii/c simulation for a further 50 ns NpT MD by replacing the bound calciums to multi-site calcium models in order to validate the geometries of the already established binding sites. Our results suggest that all sites preserve their potency perfectly (Fig. S21) and the final frame of the simulation is included in the SI as a pdb file, which also contains the water molecules coordinated to the bound calcium ions.

It is worth mentioning that at least five binding sites were identified in the case of v/a (4KTY, 14 mM Ca^{2+}) (Fig. S20) and each one binds its calcium ion continuously, while in the zymogen or cleaved systems some of these sites seem to have somewhat weaker potency (eg. Asp138, Glu139 or Glu355, Glu356, Asp357).

Our work supports the presence of two new binding sites with distinguished importance beside the main Ca^{2+} binding site of the A subunits. The first one is formed by the very same Asp270 and Asp271 residues of the A subunits and hence can be realized as an inter-subunit site which was known earlier as a place suitable for binding Yb^{3+} ions. The second site has a more

complex structure and will be relevant only in the absence of APs. This site can be established by the contribution of Asp343, Asp345, Asp351 and Asp367 residues of the A subunits after the dissociation of APs. Moreover, this site can be formed only in the presence of extremely high Ca^{2+} concentration where rotation of the A subunits occurs. This site can also be found in the active FXIII-A^{O} (v/a) which enhances its relevance even further.

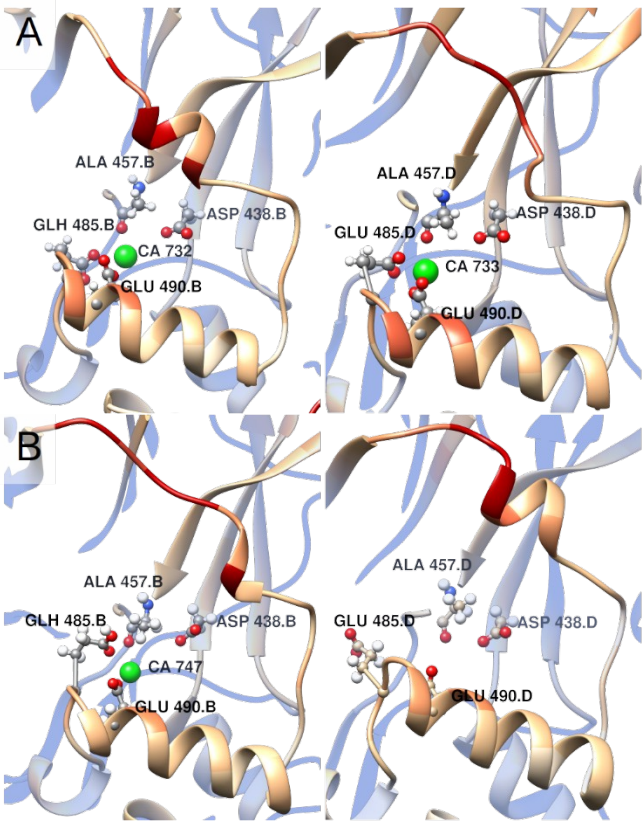

Figure 7. Main binding sites of FXIII-A_2 in the cluster representatives after clustering the dynamic trajectories. The results of simulation ii/a (A) and the results of simulation ii/d (B). (The secondary structure elements of FXIII-A_2 was represented as “cartoons”, the important residues of main sites were shown as “ball and stick” models and the bound calciums were depicted with green spheres.)

TG2

Based on site-directed mutagenesis studies, five out of six binding sites of TG2 were identified. Despite the significant efforts of this work, the static snapshot used as starting point, provided by X-ray crystallography, was lacking the true dynamic behaviour of proteins *in vivo*, nevertheless, MD simulations provide a good insight to study these time-resolved atomic-level events. Our detailed work serves multiple purposes; (1) since such detailed mutagenesis studies are not available for FXIII-A_2, it was important to compare the results of our microsecond long MD

simulations to *in vitro* experiments and TG2 embodies a perfect target for such a purpose; (2) we wished to review the relevance of previously identified binding sites and also intended to identify the missing sixth site of TG2. Instead of plotting binding events alone, we have concentrated on the representation of distances between the carboxylate carbon atom of acidic residues and the calcium ions. In the following we use the site labelling suggested by Király et al. (Király et al., 2009).

The calcium binding features of GDP bound TG2 (iv/d) is depicted in Figure 8. and the location of the binding sites on the protein can be seen in Figure 9, using the same panel labelling as in Figure 8.

Compared to *in vitro* results and static modelling, an important difference is that at least two sites exist on the β-sandwich domain (Fig. 8(A, B, C, D, E)). In the case of the GTP bound form of TG2 (iv/e) these sites appear much less significant (Fig. S26(A, B, C, D, E)). In agreement with previous experiments, the S1 site was also found in all of the simulations except in iv/b (GDP-free). It should be noted that this is formed via the contribution of the Asp232 and Asp233 residues (Fig. 8(G), Fig. S26(E), Fig. S27(F)). Concerning the acidic residues of the S2A site, we can see that the sidechain of Glu396 orientates to the inner side of the core domain and therefore is buried within. Besides Glu396, there is another negatively charged residue located at this site, Asp400. In the case of iv/a-d the distance between the Cγ of Asp400 and the Nζ of Lys464 is 5.55 ± 1.25 Å on average, which does not deviate considerably during the production runs. Comparing this value to those which can be calculated in the case of ii/a-d (FXIII-A_2'), the corresponding residues are significantly further apart (6.25 ± 1.30 Å), thus the shorter distance can be attributed to the salt-bridge blocking of the Asp400 sidechain. Instead of the S2A and S2B sites a new site has been identified with a high rate of occupancy. By the residues of Glu381, Glu451 and Glu454 a new site can be formed (Fig. 8(J), Fig. S25(I), Fig. S27(E)) and the occupancy of this site is almost continuous in every case. A second new site has also been formed by the residues of Glu319, Asp408 and Asp409 (Fig. 8(I), Fig.S25(H), Fig. S27(H)). It is worth noting that in the open conformation the Asp581 takes part in the formation of this site. The Glu467, Glu469, Glu470 residues of the 460s loop have weak binding affinity and the simulation with multi-site calcium models also confirm that the Ca^{2+} remains close to Glu467 and Glu470 (Fig. 8(L), Fig. S25(K), Fig. S26(G), Fig. S29). Two additional binding sites, namely the S3A and S3B, were proposed with one acidic residue in each. It has been found that both amino acids bind the same calcium ion and, in the minority of the simulation time, the Glu363 also contributes to the binding (Fig. 8(H), Fig. S26(F), Fig. S27(G)). The S4 site is composed from the consecutive five acidic residues between Asp151 and Glu158.

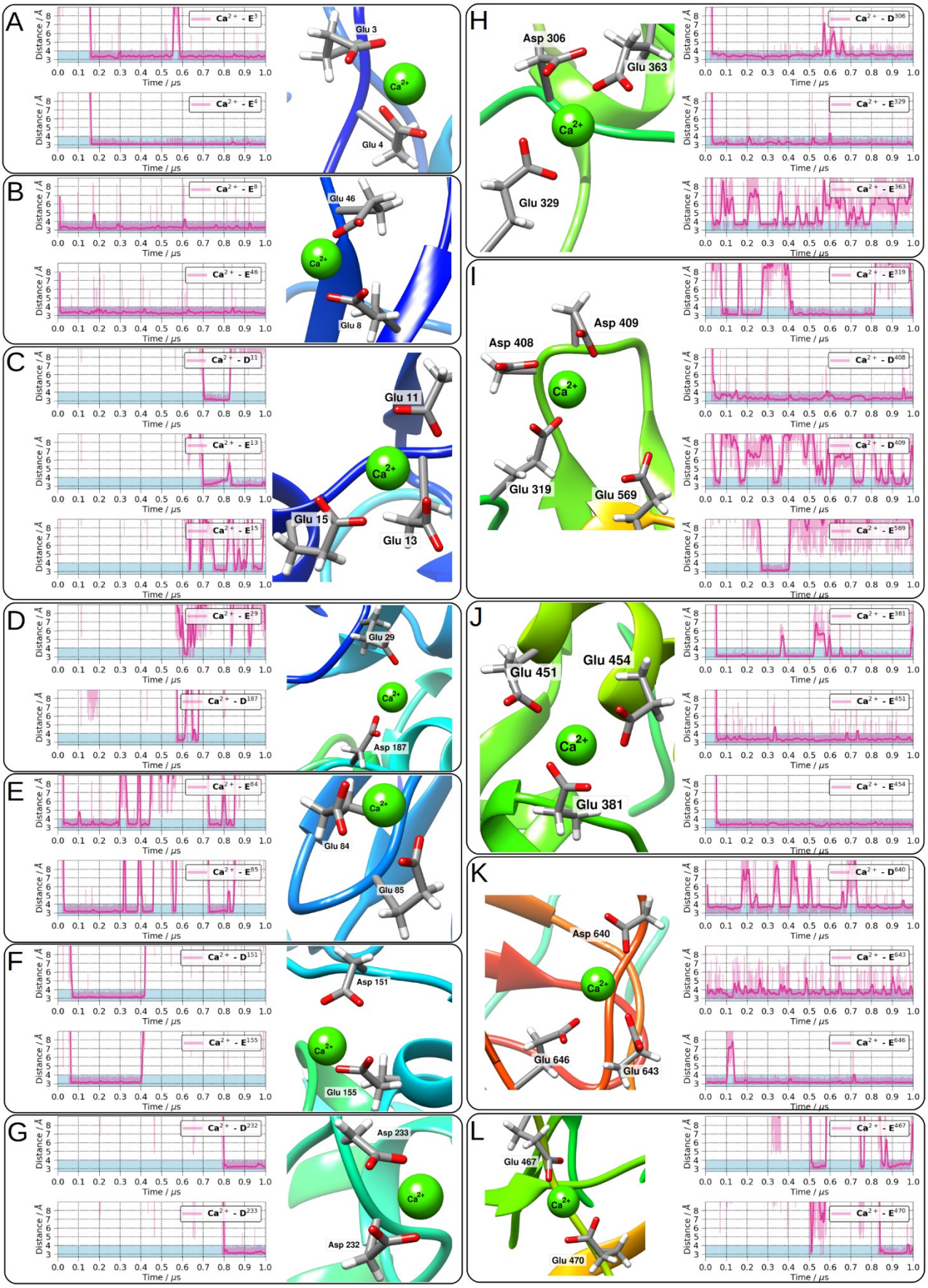

Figure 8. Calcium binding of TG2 during the iv/d simulation. (The distance between any calcium ion and the corresponding carboxylate carbon atom was calculated. The protein segments were

represented with coloured “cartoons” and the corresponding negatively charged side-chains were shown with “sticks”. The bound calciums can be seen as green spheres.)

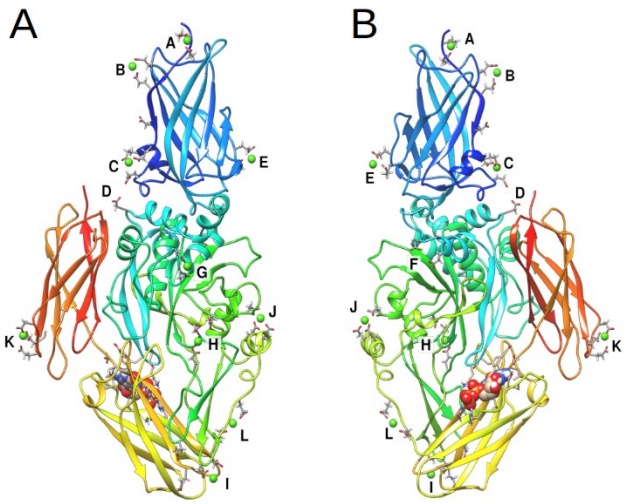

Figure 9. The location of Ca^{2+}-binding sites in TG2. ((A), front view, (B), back view. The site labelling used is the same as Fig. 8. The colouring of cartoons is the same as it was in the case of Figure 8.)

Our simulations are in good agreement with experimental results, meaning that Asp151 and Glu155 seem to be an appropriate diad for binding calcium ions (Fig. 8(F), Fig. S26(C), Fig. S25(F), Fig. S27(D)). On top of the already known sites a new one has also been identified, formed by Asp640, Glu643 and Glu646 in every case (Fig. 8(K), Fig. S26(J), Fig. S25(J), Fig. S27(K)). This site was almost continuously occupied, independently from the simulation systems, hence we reckon that it is the sixth site of TG2.

Similarly to the FXIII-A_2 simulation, iv/d was continued for a further 50 ns using multi-site calcium models and the majority of the ion binding contacts remained unchanged (Fig. S29). The final frame of this simulation can be found in the SI.

Regarding the calcium binding properties of TG2, based on our recent results, the following conclusions can be made on the calcium binding sites of TG2 compared to the experimental results by Király et al. (Király et al., 2009). The S1, S3A/S3B and S4 sites are found to be suitable for binding calciums in our simulations, ie. they support the experimental findings and *vice versa*, however opposing results have been found regarding the S2A/S2B and S5 sites. The S2A/S2B sites correspond to the main site of FXIII-A where we have observed the binding of Ca^{2+} during the 1 or 2 μs long simulations, but no binding event has happened in the case of TG2. In respect to the S5 site we have also found that this site has no significance in the early stage of activation. We cannot rule out that in a later phase of the activation process the significance of these sites will increase,

but starting from crystallographic structures even during these relatively long MD simulations we were unable to observe binding events at these places. Instead of the S2A/S2B and S5 sites we suggest the presence two other sites near the residues of Glu381, Glu451, Glu454 and by the Glu319, Asp408, Asp409. Finally, a strong binding site was identified formed by the contribution of Asp640, Glu643 and Glu646 residues. Due to its outstanding binding affinity we consider it as the previously unknown, sixth calcium binding site (S6) of TG2.

The results of the calcium binding properties study of TG2 are not only clarify the existence of its Ca^{2+} binding sites, but also to indirectly support our findings and observations which were made in the case of FXIII-A_2.

Conclusion

A series of microsecond long all-atom MD simulation were carried out in order to clarify the important details which may initiate the activation of the studied TGs and are primarily linked to the binding of calcium ions to blood coagulation factor XIII-A_2 and TG2.

In the case of the dimer A subunit of blood coagulation factor XIII we have found that the calcium concentration increases the flexibility of the N-terminals of the APs and three calcium binding sites were identified with possible significance in the activation mechanism. The main binding site of FXIII-A_2 was known and based on our work it was found that binding events can be observed in at least one of the main sites on the studied 1 or 2 μs long timescale. In those models which were based on 1GGU (Ca^{2+} bound crystallographic structure), the calcium ions preserved their position throughout all of the simulations, thus these sites can be considered strong binding sites. Suggestions have been made for the possible significance of the binding site near Asp270 and Asp271 of both subunits during the activation procedure. A third binding site was identified, which has possible importance only in the absence of APs and the rotation of A subunits are needed to form this site (Asp343, Asp345, Asp351 and Asp367). It should be noted that this site was known in the case of TG3 already, and simulation with the open FXIII-A also confirmed the existence of this site.

It was also pointed out that in the absence of APs the high Ca^{2+} concentration (1000 mM) was able to cause large scale monomer movements, probably by breaking down the intermolecular electrostatic interactions. The calculated PCA eigenvectors predict this rotation even in the absence of any calcium ions and this finding is supported directly by *in vivo* experiments in the case of zymogen structures, if we assume that the high ionic strength causes immediate dissociation of APs.

The calcium binding ability of TG2, a structurally very similar protein although with other distinct functions, was also investigated using *in silico* modelling methods, based on the profound

work of Király and co-workers. Our current work provides further evidence on most of these proposed sites, however in a few cases results from our simulation studies seem to point to different conclusions. According to the labelling of the binding sites of Király et al., the S1, S3A/S3B and S4 sites are in good agreement with the results of previously conducted *in vitro* studies, however instead of S2A/S2B we would like to propose another site which can be formed by the contribution of Glu381, Glu451 and Glu454 in this early phase of activation. Based on our work, we could not confirm whether the S5 site really is appropriate or not for binding a calcium ion. Among other possible and earlier unknown binding sites, another site was also identified which can be formed by the Asp640, Glu643 and Glu646 residues and found to be frequently populated by a calcium ion during the MD simulations. Thus this has been identified as the sixth (S6), previously unknown binding site of TG2.

Supplementary material

Acknowledgements

The authors truly appreciate the computational resources made available on the Debrecen2 (Leo) GPU cluster of the Governmental Information-Technology Development Agency. We are thankful to Prof. László Muszbek for his helpful comments. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

Funding

This work was supported by the Hungarian Scientific Research Fund (OTKA K-106294). AF is also thankful for the PhD scholarship to the Richter Gedeon Talentum Foundation. The work at Swetox (DM) was supported by Knut and Alice Wallenberg Foundation [2013.0253] and Swedish Research Council FORMAS [2016–02031].

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015).

GROMACS: High performance molecular simulations through multi-level parallelism from

laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U., & Steinert, P. M. (2003). Roles of Calcium Ions in the Activation and Activity of the Transglutaminase 3 Enzyme. *Journal of Biological Chemistry*, *278*(26), 23834–23841. https://doi.org/10.1074/jbc.M301162200

Ambrus, A., Bányai, I., Weiss, M. S., Hilgenfeld, R., Keresztessy, Z., Muszbek, L., & Fésüs, L. (2001). Calcium Binding of Transglutaminases: A ^{43}Ca NMR Study Combined with Surface Polarity Analysis. *Journal of Biomolecular Structure and Dynamics*, *19*(1), 59–74. https://doi.org/10.1080/07391102.2001.10506720

Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, *40*(W1), W537–W541. https://doi.org/10.1093/nar/gks375

Bakan, A., Dutta, A., Mao, W., Liu, Y., Chennubhotla, C., Lezon, T. R., & Bahar, I. (2014). Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics*, *30*(18), 2681–2683. https://doi.org/10.1093/bioinformatics/btu336

Bakan, A., Meireles, L. M., & Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, *27*(11), 1575–1577. https://doi.org/10.1093/bioinformatics/btr168

Bergamini, C. M. (1988). GTP modulates calcium binding and cation-induced conformational changes in erythrocyte transglutaminase. *FEBS Letters*, *239*(2), 255–258. https://doi.org/10.1016/0014-5793(88)80928-1

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, *126*(1), 014101. https://doi.org/10.1063/1.2408420

Casadio, R., Polverini, E., Mariani, P., Spinozzi, F., Carsughi, F., Fontana, A., … Bergamini, C. M. (1999). The structural basis for the regulation of tissue transglutaminase by calcium ions. *European Journal of Biochemistry*, *262*(3), 672–679.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., … Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688. https://doi.org/10.1002/jcc.20290

Credo, R. B., Curtis, C. G., & Lorand, L. (1978). Ca2+-related regulatory function of fibrinogen. *Proceedings of the National Academy of Sciences*, *75*(9), 4234–4237. https://doi.org/10.1073/pnas.75.9.4234

D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, … P.A. Kollman. (2016). *AMBER 2016*. University of California, San Francisco. Retrieved from http://ambermd.org

Droettboom, M., Caswell, T. A., Hunter, J., Firing, E., Nielsen, J. H., Lee, A., … Würtz, P. (2018 17). Matplotlib/Matplotlib V2.2.2. Zenodo. https://doi.org/10.5281/zenodo.1202077

Fizil, Á., Sonderegger, C., Czajlik, A., Fekete, A., Komáromi, I., Hajdu, D., … Batta, G. (2018). Calcium binding of the antifungal protein PAF: Structure, dynamics and function aspects by NMR and MD simulations. *PLOS ONE*, *13*(10), e0204825. https://doi.org/10.1371/journal.pone.0204825

Fox, B. A., Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E., & Teller, D. C. (1999). Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. *The Journal of Biological Chemistry*, *274*(8), 4917–4923.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., & Onufriev, A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, *33*(Web Server), W368–W371. https://doi.org/10.1093/nar/gki464

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, *22*(21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461

Gupta, S., Biswas, A., Akhter, M. S., Krettler, C., Reinhart, C., Dodt, J., … Oldenburg, J. (2016). Revisiting the mechanism of coagulation factor XIII activation and regulation from a structure/functional perspective. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep30105

Han, B.-G., Cho, J.-W., Cho, Y. D., Jeong, K.-C., Kim, S.-Y., & Lee, B. I. (2010). Crystal structure of human transglutaminase 2 in complex with adenosine triphosphate. *International Journal of Biological Macromolecules*, *47*(2), 190–195. https://doi.org/10.1016/j.ijbiomac.2010.04.023

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 712–725. https://doi.org/10.1002/prot.21123

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Jang, T.-H., Lee, D.-S., Choi, K., Jeong, E. M., Kim, I.-G., Kim, Y. W., … Park, H. H. (2014). Crystal Structure of Transglutaminase 2 with GTP Complex and Amino Acid Sequence Evidence of Evolution of GTP Binding Site. *PLoS ONE*, *9*(9), e107005. https://doi.org/10.1371/journal.pone.0107005

Kanchan, K., Ergülen, E., Király, R., Simon-Vecsei, Z., Fuxreiter, M., & Fésüs, L. (2013). Identification of a specific one amino acid change in recombinant human transglutaminase 2 that regulates its activity and calcium sensitivity. *Biochemical Journal*, *455*(3), 261–272. https://doi.org/10.1042/BJ20130696

Kanchan, K., Fuxreiter, M., & Fésüs, L. (2015). Physiological, pathological, and structural implications of non-enzymatic protein–protein interactions of the multifunctional human transglutaminase 2. *Cellular and Molecular Life Sciences*, *72*(16), 3009–3035. https://doi.org/10.1007/s00018-015-1909-z

Király, R., Csősz, É., Kurtán, T., Antus, S., Szigeti, K., Simon-Vecsei, Z., … Fésüs, L. (2009). Functional significance of five noncanonical Ca2+-binding sites of human transglutaminase 2 characterized by site-directed mutagenesis: Ca2+-binding sites of TG2. *FEBS Journal*, *276*(23), 7083–7096. https://doi.org/10.1111/j.1742-4658.2009.07420.x

Komáromi, I., Bagoly, Z., & Muszbek, L. (2011). Factor XIII: novel structural and functional aspects: FXIII: structure and function. *Journal of Thrombosis and Haemostasis*, *9*(1), 9–20. https://doi.org/10.1111/j.1538-7836.2010.04070.x

Lai, T. S., Slaughter, T. F., Peoples, K. A., & Greenberg, C. S. (1999). Site-directed mutagenesis of the calcium-binding site of blood coagulation factor XIIIa. *The Journal of Biological Chemistry*, *274*(35), 24953–24958.

Li, D.-W., & Brüschweiler, R. (2010). NMR-Based Protein Potentials. *Angewandte Chemie International Edition*, *49*(38), 6778–6780. https://doi.org/10.1002/anie.201001898

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, NA-NA. https://doi.org/10.1002/prot.22711

Liu, S., Cerione, R. A., & Clardy, J. (2002). Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proceedings of the National Academy of Sciences*, *99*(5), 2743–2747. https://doi.org/10.1073/pnas.042454899

Lorand, L. (2001). Factor XIII: structure, activation, and interactions with fibrinogen and fibrin. *Annals of the New York Academy of Sciences*, *936*, 291–311.

Lorand, L., & Konishi, K. (1964). ACTIVATION OF THE FIBRIN STABILIZING FACTOR OF PLASMA BY THROMBIN. *Archives of Biochemistry and Biophysics*, *105*, 58–67.

McCammon, J. A. (1984). Protein dynamics. *Reports on Progress in Physics*, *47*(1), 1–46. https://doi.org/10.1088/0034-4885/47/1/001

Meagher, K. L., Redman, L. T., & Carlson, H. A. (2003). Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*, *24*(9), 1016–1025. https://doi.org/10.1002/jcc.10262

Muszbek, L., Ariëns, R. A., Ichinose, A., & ON BEHALF OF THE ISTH SSC SUBCOMMITTEE ON FACTOR XIII. (2007). Factor XIII: recommended terms and abbreviations: Factor XIII. *Journal of Thrombosis and Haemostasis*, *5*(1), 181–183. https://doi.org/10.1111/j.1538-7836.2006.02182.x

Muszbek, László, Bereczky, Z., Bagoly, Z., Komáromi, I., & Katona, É. (2011). Factor XIII: A Coagulation Factor With Multiple Plasmatic and Cellular Functions. *Physiological Reviews*, *91*(3), 931–972. https://doi.org/10.1152/physrev.00016.2010

Myers, J., Grothaus, G., Narayanan, S., & Onufriev, A. (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Structure, Function, and Bioinformatics*, *63*(4), 928–938. https://doi.org/10.1002/prot.20922

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In S. Markidis & E. Laure (Eds.), *Solving Software Challenges for Exascale* (Vol. 8759, pp. 3–27). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-15976-8_1

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182–7190. https://doi.org/10.1063/1.328693

Pedersen, L. C., Yee, V. C., Bishop, P. D., Trong, I. L., Teller, D. C., & Stenkamp, R. E. (1994). Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. *Protein Science*, *3*(7), 1131–1135. https://doi.org/10.1002/pro.5560030720

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pinkas, D. M., Strop, P., Brunger, A. T., & Khosla, C. (2007). Transglutaminase 2 Undergoes a Large Conformational Change upon Activation. *PLoS Biology*, *5*(12), e327. https://doi.org/10.1371/journal.pbio.0050327

Polgár, J., Hidasi, V., & Muszbek, L. (1990). Non-proteolytic activation of cellular protransglutaminase (placenta macrophage factor XIII). *The Biochemical Journal*, *267*(2), 557–560.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, *9*(7), 3084–3095. https://doi.org/10.1021/ct400341p

Šali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, *234*(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626

Salomon-Ferrer, R., Case, D. A., & Walker, R. C. (2012). An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *3*(2), 198–210. https://doi.org/10.1002/wcms.1121

Saxena, A., & Sept, D. (2013). Multisite Ion Models That Improve Coordination and Free Energy Calculations in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, *9*(8), 3538–3542. https://doi.org/10.1021/ct400177g

Stieler, M., Weber, J., Hils, M., Kolb, P., Heine, A., Büchold, C., … Klebe, G. (2013). Structure of Active Coagulation Factor XIII Triggered by Calcium Binding: Basis for the Design of Next-Generation Anticoagulants. *Angewandte Chemie International Edition*, *52*(45), 11930–11934. https://doi.org/10.1002/anie.201305133

Takagi, T., & Doolittle, R. F. (1974). Amino acid sequence studies on factor XIII and the peptide released during its activation by thrombin. *Biochemistry*, *13*(4), 750–756. https://doi.org/10.1021/bi00701a018

Turner, B. T., & Maurer, M. C. (2002). Evaluating the Roles of Thrombin and Calcium in the Activation of Coagulation Factor XIII Using H/D Exchange and MALDI-TOF MS †. *Biochemistry*, *41*(25), 7947–7954. https://doi.org/10.1021/bi025630n

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Weiss, M. S., Metzner, H. J., & Hilgenfeld, R. (1998). Two non-proline cis peptide bonds may be important for factor XIII function. *FEBS Letters*, *423*(3), 291–296.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., … Ferrin, T. E. (2012). UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *Journal of Structural Biology*, *179*(3), 269–278. https://doi.org/10.1016/j.jsb.2011.09.006

Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E., & Teller, D. C. (1994). Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII.

Proceedings of the National Academy of Sciences of the United States of America, *91*(15), 7296–7300.

Yee, Vivien C., Pedersen, L. C., Bishop, P. D., Stenkamp, R. E., & Teller, D. C. (1995). Structural evidence that the activation peptide is not released upon thrombin cleavage of factor XIII. *Thrombosis Research*, *78*(5), 389–397. https://doi.org/10.1016/0049-3848(95)00072-Y