Clemson University
TigerPrints

Publications | Physics and Astronomy

8-2011

A Missense Mutation in CLIC2 Associated with Intellectual Disability is Predicted by in Silico Modeling to Affect Protein Stability and Dynamics

Shawn Witham
Clemson University

Kyoko Takano
J.C. Self Research Institute of Human Genetics

Charles Schwartz
Clemson University

Emil Alexov
Clemson University, ealexov@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/physastro_pubs

 Part of the Biological and Chemical Physics Commons

Recommended Citation
Please use publisher's recommended citation.

This Article is brought to you for free and open access by the Physics and Astronomy at TigerPrints. It has been accepted for inclusion in Publications by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

NIH Public Access
Author Manuscript
Proteins. Author manuscript; available in PMC 2012 August 1.

Published in final edited form as:
Proteins. 2011 August ; 79(8): 2444–2454. doi:10.1002/prot.23065.

A Missense Mutation in *CLIC2* Associated with Intellectual Disability is Predicted by *In Silico* Modeling to Affect Protein Stability and Dynamics

Shawn Witham[1,$], **Kyoko Takano**[2,$], **Charles Schwartz**[2,3,*], and **Emil Alexov**[1,*]

[1] Computational Biophysics and Bioinformatics, Department of Physics, Clemson University, Clemson, SC 29634

[2] J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC 29646

[3] Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29642

Abstract

Large-scale next generation resequencing of X chromosome genes identified a missense mutation in the *CLIC2* gene on Xq28 in a male with X-linked intellectual disability (XLID) and not found in healthy individuals. At the same time, numerous nsSNPs (nonsynonomous SNP) have been reported in the *CLIC2* gene in healthy individuals indicating that the CLIC2 protein can tolerate amino acid substitutions and be fully functional. To test the possibility that p.H101Q is a disease-causing mutation, we performed *in silico* simulations to calculate the effects of the p.H101Q mutation on CLIC2 stability, dynamics and ionization states while comparing the effects obtained for presumably harmless nsSNPs. It was found that p.H101Q, in contrast with other nsSNPs, (a) lessens the flexibility of the joint loop which is important for the normal function of CLIC2, (b) makes the overall 3D structure of CLIC2 more stable and thus reduces the possibility of the large conformational change expected to occur when CLIC2 moves from a soluble to membrane form and (c) removes the positively charged residue, H101, which may be important for the membrane association of CLIC2. The results of *in silico* modeling, in conjunction with the polymorphism analysis, suggest that p.H101Q may be a disease-causing mutation, the first one suggested in the CLIC family.

Keywords

CLIC2; missense mutations; mental disorder; energy calculations; pKa calculations; electrostatics; molecular dynamics simulations

INTRODUCTION

Intellectual disability (ID) is defined as a condition of significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social and practical adaptive skills and is diagnosed before the age of 18 years [1]. Its prevalence is estimated to be 1–3% of the general population [2]. ID has multiple causes and is clinically diverse with varying degrees of intellectual functioning with or without additional features.

[*]corresponding authors: Charles Schwartz, ceschwartz@ggc.org, tel: (864) 941-8100, webpage: http://www.ggc.org/faculty/schwartz.html. Emil Alexov, ealexov@clemson.edu, tel: (864) 656-5307, webpage: www.ces.clemson.edu/compbio.
[$]contributed equally to the paper

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

X-linked ID (XLID) results from defects in X chromosome genes and has been estimated to account for 10–15% of males with ID [3]. XLID is classified as syndromic XLID or non-syndromic XLID depending on the presence or absence of dysmorphic, neurological, muscular, and metabolic features respectively. More than 90 genes associated with XLID have been identified thus far. Among them, 38 genes are considered to be non-syndromic XLID genes, with each gene contributing to no more than 0.2–0.5% of all ID [2]. Large-scale resequencing of X chromosome genes was conducted in 200 males with XLID to discover rare disease causing sequence variants [4]. Following this study, resequencing using Solexa next generation technology was conducted in 15 probands. This approach detected a missense alteration (c.303C>G, p.H101Q) in the *CLIC2* gene in Xq28 in one proband (Schwartz et al. in preparation). The family had 2 affected males with profound ID, seizures, thumb position abnormality, large ears and large testes and a female with ID. This change was observed in the affected brothers and their mother.

Chloride intracellular channel 2, CLIC2, is the least studied CLIC protein in its class. Very little is known about the functional properties of CLIC2, but *in vitro* experiments [5] and the high sequence similarity to other chloride channel proteins suggest that it may function as either a chloride channel or as a chloride channel regulator [6]. There are some *in vivo* experiments favoring the hypothesis that the CLIC2 protein is indeed a chloride channel [7]. At the same time, it is known that the CLIC2 protein exists in a soluble form as well [8], and this is the reason why CLIC2 belongs to the Janus protein family [5]. Since no obvious transmembrane motif is present in the CLIC2 structure, the transition from a soluble to membrane form of CLIC2 should involve a dramatic structural change [7]. Thus the conformational flexibility should be an intrinsic and very a important characteristic of the CLIC2 protein.

At the present time, only one interacting partner of CLIC2 is known. It has been shown that CLIC2 interacts and regulates the ryanodine receptors 1–2 (RyR1,2) [9]. RyRs are important players in Ca^{+2} channel activity, and CLIC2 may be involved in the regulation of Ca^{+2} signaling [7]. Following this, it has been shown that the presence of CLIC2 ultimately induces conformational changes of the RyR1 complex [7]. From these studies, it is hypothesized that the CLIC2 protein is a common regulator of various RyR complexes, which could give rise to a further understanding of the functional roles of CLIC2. Furthermore, it has been shown that there are high concentrations of CLIC2 in the muscle tissues of the skeleto- and cardiac system, along with a presence in the lung, spleen, and kidney [7]. Ultimately, it is reasonable to conclude that the CLIC2 protein serves a variety of functions, some of them yet to be discovered.

CLIC2 is a member of the CLIC protein family comprised of six members, CLIC1–6. Structurally, CLIC proteins are very similar to each other. There are several structural features present in all CLIC proteins while in soluble form (Fig. 1). Following the original structural report [10], the three dimensional (3D) structure is comprised of an N-domain (a.a. 1–94), which is connected to the C-terminal (a.a. 107–245) domain through a loop, termed the “joint” loop (a.a. 95–106). The C-terminal domain itself, has an extended loop, termed the “foot” loop (a.a. 152–180), which is composed of acidic groups. Despite the high structural and sequence similarity among the members of the CLIC family, noticeable differences are seen in the length and conformation of the “foot” loop. It has been suggested that CLIC2 may use its flexible “foot” loop to interact with various complexes [5,8]. Most of the known non-synonymous single nucleotide polymorphisms (nsSNPs) (p.P160A and p.D161H/Y) are located within this region. The other distinctive loop, the “joint” loop, is well conserved in both structural and sequence space. This conservation in the joint loop could indicate the importance of the loop for the function and structure of CLIC2. For example, a small change in the “joint” loop conformation will cause a large change in the N-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

domain position. The potential disease-causing missense mutation, p. H101Q, resides within the joint loop as does a known nsSNP, p.S109C. This study was undertaken to determine if any differences exist between their effects on the CLIC2 biophysical properties.

In silico modeling allows for predicting the molecular effects of missense mutations and comparing the effects induced by disease-causing and harmless mutations. Since very little is known about the CLIC2 function(s) and interacting partners, an *in silico* investigation must be focused on physical characteristics of the CLIC2 protein, namely the stability and dynamics. The available X-ray structures [5,8] provide an excellent source of information and their superimposition indicates that the foot loop has the ability to adopt multiple structures/conformations in different crystal forms, making it somewhat flexible. Also, the flexible foot loop shows low sequence conservation. All of these aspects indicate that mutations residing within this loop are less likely to disrupt the function of CLIC2. At the same time, if the CLIC2 protein functions as a chloride channel, the insertion into the membrane should involve large conformational changes, which presumably affects the N- and C-domains mutual orientation. Such an event will depend on the joint loop flexibility and therefore any mutation that affects its flexibility is expected to have a significant effect on the function of the CLIC2 protein.

In silico techniques were conducted to assess how the newly identified mutation (p.H101Q) and known nsSNPs affect the CLIC2 protein stability and dynamics. The *in silico* techniques involved pKa calculations and analysis, an evaluation of folding free energy changes, and an analysis of the changes of flexibility (molecular dynamics simulation) caused by the mutations. Plausible roles of various domains were also explored. The missense mutation, H101Q, not only causes a noticeable difference in the physical motion of the various domains, but also causes a significant difference in the folding free energy. This mutation is predicted to cause both a decrease in physical motion and a decrease in the flexibility of the CLIC2 protein. As for the presumably harmless nsSNPs, it is shown that each alteration site causes an increase in flexibility, or in other words, a decrease in the stability.

METHODS

I. Experimental

Patients and mutation screening—DNA from a proband in K8015 was analyzed using Solexa DNA sequencing technology (Illumina). The c.303C>G alteration was confirmed by direct sequencing and enzyme digestion following polymerase chain reaction (PCR). The polymorphism study for this change was performed using enzyme digestion. The exon 4 of the *CLIC2* gene was amplified from the proband and control persons. PCR products were digested with AluI and visualized by agarose gel electrophoresis. DNA and protein mutation numbering was based on NM_001289 and NP_001280, respectively.

II. *In silico* methods

Protein structures—From the work of Cromer *et al* [5], the soluble CLIC2 structure has been determined from two crystal forms, crystallized at different space groups and different pH. The structure with PDB ID 2R4V was obtained at pH=8.0, while the other structure with PDB ID 2R5G at pH=7.5. A third alternative structure was reported by Wei *et al* (PDB ID: 2PER) [8], which was crystallized at pH=6.9. The three aforementioned structures were used in the *in silico* portion of this study. A comparison of these three structures gives a global RMSD of 1.17 Å. The highest contributors to the global RMSD are the flexible "foot" loop (a.a. 152–180) in the C-terminal and several short loops (a.a. 1–14 and a.a. 53–71) located within the N-terminal domain (Note that 2PER structure has a long expression tag at the N-terminal domain, which was removed from the *in silico* modeling). The ligands,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

glutathione and 5-mercapto-2-nitro-benzoic acid, were also removed from the structures prior to *in silico* modeling since they were shown to be an artifact of the crystallization procedure [5] and not to be functionally or structurally important.

Each 3D structure was subjected to the profix program [11] to add missing heavy atoms or missing residues. The mutations were then inserted into each fixed CLIC2 structure using the program SCAP [12] within the JACKAL package (URL=http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal_General_Description). Default parameters were used with an Amber force field and the heavy atoms only option.

Non-synonymous single nucleoside polymorphisms (nsSNPs)—The known nsSNPs within CLIC2 were obtained from the NCBI SNP database (http://www.ncbi.nlm.nih.gov/snp/). We focused only on the non-synonymous SNPs, since these mutations change the amino acid sequence and their effect can be directly compared with the p.H101Q mutation. Since nsSNPs are taken from the general population, presumably from healthy individuals, it can be thought that they are harmless, especially if they are found to occur with a high frequency. The analysis revealed five nsSNPs resulting in the following protein substitutions: p.S109C (Ser is selected to be the wild type residue based on the available X-ray structures), p.P160A, p.D161H/Y, and p.R201C. This study included all substitutions, except for the 201 position. The p.R201C frequency is very low, since it was found in only one individual and thus can not be considered to be a polymorphism (unpublished data, The Wellcome Trust Sanger Institute).

pKa calculations: A point mutation can dramatically change the hydrogen bond network and even alter the ionization states of neighboring amino acids. To investigate such possibilities, the Multi Conformation Continuum Electrostatics (MCCE), version 2.2 (http://www.sci.ccny.cuny.edu/~mcce/index.php) [13–15] was used to perform pKa calculations on the WT monomers and the corresponding mutants. The MCCE calculates the ionization states of titratable groups and optimizes the hydrogen bond network. Default parameters of MCCE were used with internal dielectric constant 8[15]. The corresponding pKa shifts were calculated as:

$$\Delta pKa_i(mut) = pKa_i(WT) - pKa_i(mutation) \quad (1)$$

where $pKa_i(WT)$ is the *pKa* of the i-th amino acid calculated with the WT structure while $pKa_i(mutation)$ is the *pKa* of the same amino acid calculated using the mutant structure.

Folding free energy calculations: Following the procedure described in our earlier work [16], the energies of the structures of the corresponding proteins were minimized using the "minimize.x" module from the TINKER package [17] applying the Limited Memory BFGS Quasi-Newton Optimization algorithm and a convergence criterion of 0.01 gradient per atom. The solvent was modeled with the Still Generalized Born model [18] and the protein internal dielectric contact was 1.0. The charged states were assigned according to the pKa predictions, namely all titratable groups ionized, except Cys's. For each structure, three separate runs were performed using three different force fields, Amber98 [19], Charmm27 [20] and OPLS [21], and the results were averaged. Global RMSDs of the minimized wild type structures for each force field are 1.34 Å for 2PER, 1.32 Å for 2R4V, and 1.06 Å for 2R5G. This was done to generate results that are less sensitive to a particular force field and to provide an estimation of the error of the computational protocol. After a successful minimization was obtained, a structural segment of a length of seven amino acids, with the mutation site at the center, was extracted from the minimized structure. The minimized

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

structure of the corresponding protein and the seven residue segment were then subjected to the "analyze.x" module of TINKER to obtain the total potential energy.

The folding free energy calculations follow the approach described in Ref. [16]. The potential energy of each monomer *G(folded)* is the potential energy obtained with TINKER using the corresponding energy minimized structures (PDB IDs 2PER, 2R4V and 2R5G). The unfolded state was modeled as made of two components: (a) a component which is affected by the mutations and it is represented as a seven residue segment centered at the corresponding mutation site and (b) a component which is unaffected by the corresponding single mutation and is the same for the wild type and the corresponding mutant (see Ref [16] for details). The potential energies corresponding to these two components of the unfolded state are G_0(unfolded) and G_7(unfolded), correspondingly. The folding free energy is then described as:

$$\Delta G(folding)=G(folded)-G(unfolded)=G(folded)-G_0(unfolded)-G_7(unfolded) \quad (2)$$

where *G(folded)* is the total potential energy of the folded state and *G(unfolded)* is the energy of the unfolded state.

The effect of a mutation is calculated as:

$$\begin{aligned}\Delta\Delta G(folding\_mut)&=\Delta G(folding:WT)\\&\quad-\Delta G(folding:mutation)\\&=G(folded:WT)\\&\quad-G_7(unfolded:WT)\\&\quad-G(folded:mutation)\\&\quad+G_7(unfolded:mutation)\end{aligned} \quad (3)$$

Since G_0(unfolded) is mutation independent, it cancels out in eq. (3). Such an approach avoids the problems with modeling unfolded states since it focuses on the energy difference between WT and mutant structures [16].

An additional correction to the energy calculation was made for the cases for which it is predicted that the folding will cause proton uptake/release. For these cases, an additional term ($\Delta G^{pH}_{folding}$) is added to the energies in eq. (3) as [22,23]:

$$\Delta G^{pH}_{folding}=1.37\int_{pH_1}^{pH_2}(\Delta q)dpH \quad (4)$$

where the energy is in kcal/mol, Δq is the proton uptake/release in the interval pH_1 to pH_2, and pH_2 is taken to be 7.0, and pH_1 is taken to be the pH at which the group of consideration is fully ionized in an unfolded state. For example, a His residue having pKa = 7.0 within a protein is 50% ionized at pH = 7.0 (q = +0.5e). The same His residue in an unfolded state is assumed to have an unperturbed pKa = 6.1, and thus will be fully ionized at pH = 5.1 (pH_1 = 5.1) according to the Henderson-Hasselbalch equation [24].

Molecular dynamics (MD) simulations: The main goal of *in silico* modeling of the CLIC2 structure is to not only determine the overall effects on the folding free energy of different known mutations, but to also reveal what the effects are on the structural flexibility. For

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

such a purpose we carried out molecular dynamics simulations on the wild type (WT) and mutant proteins as described below.

The WT and mutant PDB files were first subjected to the "minimize.x" module of TINKER [17]. This was done to minimize the effect of plausible structural imperfections which could give rise to very high fluctuations (overheating) of the initial structures. Then, the minimized structures were subjected to the "dynamic.x" module of the TINKER package [17]. The necessary parameters that were used for this simulation are that of the number of simulation steps (10^6), the time interval of simulation steps (1 femtosecond), the data saving/output interval (1 picosecond), and temperature (298 K). The amber98 TINKER parameter file was used for this simulation with the Still GB model for the water phase. The length of MD simulations was not fixed, because there is no information about the time scale of CLIC2 conformational dynamics. However, our main goal is to reveal the plausible differences between H101Q mutant, the wild type CLIC2 and other CLIC2 mutants. Thus, the results from the MD simulation of H101Q mutant were used to determine the length of MD simulations for all other cases. It was done using both the RMSD and the displacement of the center of mass (ΔCOM) as a function of simulation time (see below) and the runs were considered sufficiently long once the RMSD and ΔCOM for H101Q reached saturation (simulation time < 0.5 *n*s). The saturation can be seen in Fig. 2, where ΔCOM for H101Q mutant reaches 1.6Å after about 100 – 200*p*s and further simulations do not alter this value, i.e. ΔCOM remains at 1.6Å. The same observation can be made for RMSD of H101Q mutant: at t=100 – 200*p*s, it reaches RMSD=2.2Å and stays at this level for the rest of the simulation time.

The resulting snapshot structures were compared to the original starting structures using the Multiple Alignment with Translations and Twists (MATT) program [25]. The CLIC2 structure is comprised of an N-terminal domain, linked to the C-terminal domain through so called "joint loop" and a so called "foot loop". To avoid possible artifacts arising from superimposing the structure-snapshots, which domains may move relative to each other, the superimpositions were done using the main body of the CLIC2 protein, the C-terminal domain (a.a. 107–245), excluding the flexible foot loop. The MATT program has the capability of taking a full PDB structure and superimposing it with a pseudo-structure (in our case, the stable parts of the C-terminal). The resulting superimposed structures were then available for the determination of various quantities.

The plausible disease-causing mutation, p.H101Q, and all other known nsSNPs, are located either in the "joint" or "foot" loop with the exception of the nsSNP p.S109C, which is located at the beginning of the first C-terminal helix coming from the "joint" loop. Because of this, we focused our investigation on the properties of these structural entities, while focusing our calculations on the mutation-site/location correspondence. Thus, the superimposed structures were used to calculate either the root mean square deviation (RMSD) or the displacement of the center of the mass (ΔCOM) as a function of the simulation time. While such plots are provided in the supplementary materials, they do not allow for an objective assessment of the results since the simulation time is not the absolute time. Instead we introduce the *timed average* ($\overline{X}$) of either RMSD or ΔCOM as:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{5}$$

where "*X*" stands for either RMSD or ΔCOM and "*n*" is the snapshot number. Such a quantity has an important property to tend asymptotically to a particular value and thus to be

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

used to assess if the quantity of interest was properly estimated in the particular simulation time window.

In the cases of the p.H101Q and p.C109S mutation, the RMSD and ΔCOM were calculated for the N-terminal domain (a.a. 1–94). The beginning of the N-terminal domain (a.a. 1–14) is extremely flexible as well as the loop (a.a. 53–71) so these structural segments were excluded from both the RMSD and ΔCOM calculations to reduce noise. In the cases of p. P160A and p. D161H/Y mutants, the RMSD and ΔCOM were calculated for the foot loop (a.a.152–180) since they are located within this structural segment.

RESULTS and DISCUSSION

I. Experimental/clinical results

Solexa sequencing of X-chromosome genes and polymorphism study—A missense alteration in the *CLIC2* gene p.H101Q (c.303C>G) was found in a male with XLID by Solexa sequencing. This amino acid is located in the joint region and is a highly conserved residue from human to zebrafish. Sequencing analysis of the family of the proband showed that affected brothers and their mother also had this change while two sisters of the proband did not (data not shown). Analysis of 1059 normal X chromosomes (685 males, 187 females) did not detect this alteration.

Clinical reports—This family consists of 3 siblings. Two affected males have profound ID, seizures, large testis, abnormal positioning of the thumb and large ears with an over folded helix. The proband (II-1) walked independently and said his first words at the age of two years. Developmental regression was noted after he had the measles at two years of age. At the age of 61, his head circumference was 56 cm. He had prognathism with a prominent nasal bridge, colimella extending below the ala nasi and a flat upper lip. He had normal knee reflexes and no spasticity. He had a high stepping, wide spaced gait. His brother (II-3) had macrocephaly (63.2cm, >97th percentile) without hydrocephalus. At the age of 48, he had coarse facial features, a depressed nasal bridge, broad bulbous nose, and a large and protruding tongue with an everted lower lip. He had some joint contructures and was wheelchair bound. Both patients were negative for *FMR1* and *L1CAM* testing. A sister (II-4) has ID, but other information was not available. There was no family history of ID.

II. Results of *in silico* modeling

pKa calculations—The pKa calculations on the WT structures showed no amino acids with very unusual pKa's (see Table 1S). All His residues, including H101, were calculated to have elevated pKa's suggesting that they are ionized at pH=7.0 and below. This has important implications for our energy analysis, since the folding of the CLIC2 is then expected to cause proton uptake at neutral pH (pH=7.0). Assuming that in an unfolded state the pKa of His is close to the standard pKa = 6.1, the Histidines in an unfolded state should be neutral at pH=7.0. Thus, the folding will cause proton uptake from the water phase, which in turn will destabilize the CLIC2 protein according to the theory of ion binding [22,23] (see eq. 4).

Several titratable groups were predicted to have pKa shifts away from standard pKa values using the 2PER structure, but were found to have standard pKa's using the other two structures (2R4V and 2R5G), and thus their averaged pKa's were almost unperturbed (Table 1S). Thus, according to our pKa analysis, the CLIC2 does not have amino acids with significantly perturbed ionization states that could serve as catalytic residues or be involved in some kind of proton transfer reaction, although it should be noted that all Cys residues,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

not involved in disulfide bonds, are calculated to have very high pKa's indicating that they may hold a proton.

Among the alterations that are investigated in this work, three of them (p.H101Q, p.D161H and p.D161Y) involve a titratable group, while the rest of them engage non-titratable amino acids. However, even the mutation does not introduce or remove a charge; the substitution can still affect the neighboring pKas by perturbing the dielectric boundary, altering proton network or affecting side chain packing. Because of this, we carried out pKa calculations for all alterations and compared the predicted pKas with those calculated with the WT protein. The results are summarized below for each alteration separately:

a. p.H101Q. This mutation removes a positive charge from the WT CLIC2 protein, but the substitution to Gln from His was found not to affect any other titratable amino acid. The reason is that H101 in the WT structure predominantly interacts with the backbone oxygens of S103 and P104; while the interaction with E91 is relatively weak. Thus, the removal of H101 does not affect E91 since it is hydrogen bonded to S103. In addition, the H101 is exposed to the water phase and thus the long range electrostatic interactions are effectively screened and do not affect distant ionizable groups. The finding that the p.H101Q mutation does not cause ionization changes, combined with the predicted proton uptake for the WT protein, has implications for the energy calculations of the effect of the mutation since the mutant is not destabilizing by the means of proton uptake.

b. p.S109C. This alteration does not involve a typical titratable group, since Cys residues usually have high pKa's and hold a proton (or are involved in sulfate bridges). In the WT structure, S109 is buried and makes a hydrogen bond with Y149. In the altered structure, the C109 is predicted to be protonated and be involved in the same hydrogen bond interactions as S109. Thus, this alteration does not affect the WT hydrogen bond network.

c. p.P160A. Both residues are non-titratable, non-polar residues, and although the alteration site is located within a cluster of acidic groups, no significant pKa change upon mutation is predicted. This result was maintained using all three CLIC2 structures, regardless of the different conformations that the foot loop takes in these structures. The main reason is that site 160 is totally exposed to the water phase and does not interfere with any pair wise interaction.

d. p.D161H. The WT residue is predicted to be negatively charged at neutral pH (Table 1S). It is located in a cluster of acidic groups within the foot domain. However, all acidic groups in the foot domain are predicted to have standard and even lower than standard pKa, because they are fully exposed to the water. The alteration introduces histidine, which is calculated to have an elevated pKa of about 7.0 and thus to be partially ionized at neutral pH. Thus, this mutation significantly alters the net charge of the foot domain, making it less negative. However, it is unclear what the physiological importance may be. In terms of proton uptake, the altered protein will be destabilized by the proton uptake coming from H161 at neutral pH (see eq. (4)). However, the destabilization of this loop may have no significant effect since the foot loop is intrinsically unstable.

e. p.D161Y. This alteration alters the net charge of the foot domain by one electron unit, since D161 is fully ionized in the WT protein. However, it does not affect proton uptake and has no effect on the pKa of any other ionizable group in the CLIC2 protein.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Folding free energy calculations—Below we summarize the results of the folding free energy calculations. The calculations were done assuming that acidic groups (Asp and Glu) and basic groups (Arg, Lys and His) are fully ionized as predicted by the pKa calculations. Once the results were obtained, further corrections were, if applicable, made to add the effect of predicted proton uptake/release (see eq. (4)). The results are presented for each mutation separately. In the supplementary material, we also show the effects of mutations on the solvation energy component (taken from the analyze module of TINKER) of the folding energy (Table 2S).

a. p. H101Q. Position 101, which resides in the joint loop, is exposed to the water phase, has very little desolvation penalty to pay and is involved in weak interactions with neighboring backbone oxygens. The change from His to Gln takes away a weak positive charge. However, the Gln residue has the capability of creating a hydrogen bond with surrounding residues, consequently stabilizing the joint loop almost in the same manner as the WT residue. This is the reason why the predicted energy change is almost zero. However, by taking into account the proton uptake/release, which makes the WT structure less stable because of the proton uptake caused by the WT His101, the missense mutation p.H101Q ultimately results in the folding free energy to be a positive number. This indicates that the mutation makes the CLIC2 structure more stable than the WT (Table 1). The same conclusion is made in terms of the solvation energy component (Table 2S, supplementary materials), where the mutant solvation energy is much more favorable than of the WT. This increase in stability may cause lasting effects with the motion of the N-terminal, since the movement of the N-terminal is directly effected by the flexibility of the joint loop.

b. p. S109C. The stability of the WT CLIC2 changes little as a result of the p.S109C alteration. The alteration introduces a Cys residue, but as shown above, the Cys is predicted to be neutral, and thus to be a proton donor. A Ser residue and a protonated Cys residue poses very similar biochemical properties: they are both good hydrogen donors and have similar side chain dimensions. Thus, the substitution has a modest effect on the folding free energy making the mutant less stable than the WT (Table 1). The effect on the solvation energy component is also much smaller as compared to H101Q mutant (Table 2S, supplementary material). No additional proton uptake/release is associated with this mutation.

c. p. P160A. Site 160 is totally exposed to the surrounding water phase and the WT residue Pro is not involved in any specific interactions. Replacement with Ala does not involve any van der Waals (vdW) clashes, does not introduce new interactions and does not produce any conformational changes. As a result, the numerical protocol predicts that there will be no change on the folding free energy due to p. P160A mutation. The effect on the solvation energy component is also much smaller as compared to H101Q mutant (Table 2S, supplementary material).

d. p. D161H. The foot loop carries a large negative charge due to the large number of acidic groups, one of them being Asp161. However, all titratable groups within the foot loop are totally exposed to the water phase and thus their mutual interactions are highly damped. This is demonstrated by the pKa calculations (see Table 1S, supplementary materials), which show that almost all pKa's within the foot domain are close to standard, unperturbed pKas. The alteration introduces an oppositely charged residue, a positively charged His, but has a modest effect on the stability, including the solvation energy component (Table 2S, supplementary material). This is because the His residue is also totally exposed to the water phase and the effect of its positive charge is screened and damped. In addition, the His residue is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

predicted by the pKa calculations to be fully ionized at neutral pH, and therefore will introduce proton uptake in the mutant, further destabilizing it with respect to the WT structure (Table 1).

e. p. D161Y. This is another alteration occurring at the same site, site 161, which is Asp in the WT CLIC2. As was said above, this WT residue does not contribute to the stability of CLIC2, since it is not involved in specific interactions and is exposed to the water phase (no desolvation penalty). The mutation introduces a Tyr residue, which is a polar, does not carry net charge, and has a relatively larger side chain than the WT Asp. As a result, the Tyr mutation destabilizes the CLIC2 structure by more than 2 kcal/mol (Table 1). The effect on the solvation energy component is also much smaller as compared to H101Q mutant (Table 2S, supplementary material).

Molecular dynamics simulations—Within the molecular dynamics simulation results, the WT structure showed a time average RMSD of 3.2 Å and ΔCOM of 2.4 Å for the N-terminal, and an RMSD of 3.1 Å and ΔCOM of 1.4 Å for the foot loop. These two separate domain calculations are used for a baseline comparison with the mutation results, which reside in the N/C-terminal "joint" loop and C-terminal "foot" loop. Since there is no previous knowledge of how much each domain moves (or should move) throughout the given simulation interval, the averaged WT values of both RMSD and COM are used as an acceptable range. All results shown are an average of all three CLIC2 structures (2PER, 2R4V, 2R5G). Below we present the results separately for the mutations at the "joint" and "foot" loops.

a. Alterations occurring at the "joint" loop (p. H101Q and p.S109C). The effect of the alterations on the conformational dynamics of CLIC2 is assessed by monitoring the change of the center of the mass (ΔCOM) of the N-terminal domain or the corresponding RMSD. This was done for the WT, H101Q and S109C structures (Figs. 2a,b and Figs. 1Sa,b in supplementary materials). Both ΔCOM and RMSD rapidly increase within the first 20–50 *p*s of the simulations and there is no significant difference in the behavior of the WT and the altered CLIC2 proteins during this initial period. As the simulation time increased, we observed several fluctuations (Fig. 2) and after 50 *p*s up to 150 *p*s simulation time, a significant difference was noticed for the WT and S109C CLIC2 proteins as compared with the H101Q CLIC2 protein (Fig. 2a). Both the time average RMSD and ΔCOM for the H101Q protein saturate and became almost constant after 100 – 200 *p*s of simulation time. This indicates that perhaps we have reached the simulation time that captures all important motions of the N-terminal domain within the H101Q mutant. These fluctuations, both ΔCOM and RMSD, are much smaller in magnitude as compared to the same quantities for the WT and S109C mutant (Fig. 2). The time averages for WT and S109C do not reach saturation in the time interval of the simulations (up to 0.5 ns). Instead, their magnitudes kept increasing with simulation time indicating that longer simulations are needed to capture all important motions in the WT and S109C mutant. Numerical values for the RMSD and ΔCOM of H101Q are 2.4 Å and 1.7 Å, respectively. The same quantities for S109C mutant are 2.9 Å and 2.1 Å.

The time average ΔCOM shown in Fig. 2 clearly indicates that the mobility of the N-terminal domain is much more restricted in the H101Q mutant as compared to both the WT and the S109C mutant. The same is valid for the internal flexibility, as measured through the time average RMSD, which is much smaller in the H101Q mutant than in the WT and S109C mutant. Thus, the MD simulations suggest that the H101Q mutation rigidifies the "joint" loop, which in turn restricts the mobility

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of the N-terminal domain and reduces its internal structural flexibility. In contrast, the nsSNP, S109C, leaves the conformational properties almost the same as they were in the WT CLIC2.

b. Alterations occurring at the "foot" loop (p. P160A, p. D161H and p. D161Y). The corresponding time average ΔCOM and RMSD are shown in Fig. 3a,b,c,d (the absolute ΔCOM and RMSD are shown in supplementary materials, Figs. 2a,b). At the end of the simulation time the corresponding foot loop ΔCOMs are: 1.7 Å, 2.0 Å and 1.4 Å for p.P160 Å, p.D161H and p.D161Y, respectively. In terms of time average RMSD the results are 3.1 Å, 3.4 Å and 2.8 Å, respectively. Some of these numbers are lower, others are higher than the corresponding quantity calculated for the WT, however, no significant difference can be pinpointed. In addition, as mentioned above, these fluctuations occur in the "foot" loop, which is intrinsically flexible and takes different conformations in different X-ray structures. Variations of its internal flexibility, either making it slightly more or less flexible are not expected to alter its WT properties. The same COM and RMSD values were determined for the N-terminal region for each mutational site residing in the foot loop. These values follow the same pattern of showing a slight increase in flexibility (decrease in stability) as stated previously. These values are shown in Figures 3c,d.

Multiple sequence alignment and residue conservation—The sequence of CLIC2 is highly conserved among species, but the sequences of the CLIC family are not. Multiple sequence alignment of the six members (see for example Refs. [5,8]) indicates that none of the amino acid positions investigated in this work is conserved. However, position 161 has a Tyr residue in other CLIC proteins, which may indicate that it can tolerate a D→Y substitution. The presence of 3 different residues at position 161 also supports the notion that position 161 can hold almost any type of amino acid. In the case of position 160, in CLIC5,6 the wild type residue is Ala, which favors the suggestion that the P160A mutation in CLIC2 is a harmless variation. Position 109 is very interesting since CLIC2 exists in almost two equivalent fractions having either Ser or Cys at position 109 and thus supports our prediction that the S109C mutation is harmless. In contrast, position 101 in all CLIC proteins is a positively charged residue (except for CLIC3), which indicates the importance of a positive charge at this position and indirectly supports our suggestion that the H101Q mutation is a disease-causing mutation.

Critical assessment of the computational approaches used—The computational part of the work is aimed to provide support of the clinically found missense mutation in CLIC2 gene in XLID patients and to explain the molecular effect causing the disease by comparing to the effect induced by harmless mutations. The in silico analysis and the corresponding conclusions were based on investigating the effects of mutations on stability, flexibility and hydrogen bond network of the CLIC2 protein. Such an approach, which delivers a consensus prediction, has an advantage with respect to a single component analysis. For example, if our predictions were based on the energy calculations alone, we will not be able to distinguish between H101Q and other missense mutations, because of the large standard error (see Table 1). However, our computational modeling suggests differences between the effects caused by H101Q mutation and all other (harmless) mutations on the level of folding energy, local flexibility and pKa values. In particular, the predicted change of the local flexibility (rigidifying the joint loop in case of H101Q mutation) does not necessarily mean changes in the stability (folding energy). Instead, such a change of the flexibility is expected to affect the CLIC2 plausible function as a chloride channel by reducing its ability to undergo large conformational change in order to enter the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

membrane. Combining different effects to deliver the predictions provides extra confidence that H101Q, perhaps, is a disease-causing mutation.

An important component of our analysis is pKa calculations and this was demonstrated on two levels: (a) predicting proton uptake/release due to CLIC2 fording and thus adding to the folding energy calculations the corresponding correction and (b) assigning the correct protonation states of titratable residues in the CLIC2 protein. A typical example is S109C mutation. The solution pKa value of Cys is about 9.1 and at neutral pH it is not ionized in solution, i.e. it is a polar residue. Our pKa calculations predicted that Cys109 is protonated in the CLIC2 protein as well and thus is a polar but uncharged residue, with properties similar to the properties of Ser, i.e. proton donor. This helps understand why S109C mutation is found in the general population of healthy individuals, since the substitution is a very conservative substitution.

Our MD simulations are relatively short (0.5*n*s). It can be argued that such a short simulation time may not be enough to reveal the changes of the conformational flexibility caused by the mutations. However, the primary focus of our study is the effect of the H101Q mutation, and the Gln101 mutant was used to decide how long the MD simulations should be carried out. Fig. 2a,b shows that both the RMSD and ΔCOM of Gln101 mutant reach a saturation at very early stages of the simulations (t=100 – 200*p*s). Despite this, we extended the simulations to 0.5*n*s simply to make sure that the saturation is stable. All other mutations showed larger RMSD and ΔCOM compared with Gln101 and most of them have not converged in the simulation time (t=0.5*n*s). However, it is very unlikely that longer simulations will bring their RMSD and ΔCOM lower and this is the reason why longer simulations were not attempted.

CONCLUSION

The study reports a clinical finding that the CLIC2 gene in XLID patients differs from that in healthy individuals resulting in a missense mutation H101Q in the corresponding CLIC2 protein. The computational part of the work provided support that H101Q may indeed be disease-causing mutation based on the comparison of the calculated effects on stability, dynamics and hydrogen bond network with respect to those predicted for harmless mutations (human genetic variations). It was predicted by *in silico* analysis that H101Q mutation, in contrast to the all other harmless mutations, will increase CLIC2 protein stability, will reduce the flexibility of the N-domain and will remove a positively charged residue (His 101). Since flexibility is expected to be crucial for the transformation of CLIC2 protein from soluble to membrane form and since the attachment to the membrane is presumed to be facilitated by positively charged residues, these findings were considered as indicators for plausible pathological effect of the H101Q missense mutation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by awards from NLM/NIH, grant numbers 1R03LM009748 and 1R03LM009748-S1. We thank Brian Dominy for the useful comments and suggestion regarding molecular dynamics simulations and evaluating the results. The support of Clemson University CCIT group, especially Corey Ferrier, is greatly appreciated. Dedicated to the memory of Ethan Francis Schwartz (1996–1998).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

References

1. Luckasson, R.; Borthwick-Duffy, S.; Buntinx, W. Mental retardation: Definition, Classification, and Systems of Supports. Washington: American Association on Mental Retardation; 2002.
2. Gecz J, Shoubridge C, Corbett M. The genetic landscape of intellectual disability arising from chromosome X. Trends Genet. 2009; 25(7):308–316. [PubMed: 19556021]
3. Ropers HH, Hamel BC. X-linked mental retardation. Nat Rev Genet. 2005; 6(1):46–57. [PubMed: 15630421]
4. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Teague J, Butler A, Jenkinson A, Jia M, Richardson D, Shepherd R, Wooster R, Tejada MI, Martinez F, Carvill G, Goliath R, de Brouwer AP, van Bokhoven H, Van Esch H, Chelly J, Raynaud M, Ropers HH, Abidi FE, Srivastava AK, Cox J, Luo Y, Mallya U, Moon J, Parnau J, Mohammed S, Tolmie JL, Shoubridge C, Corbett M, Gardner A, Haan E, Rujirabanjerd S, Shaw M, Vandeleur L, Fullston T, Easton DF, Boyle J, Partington M, Hackett A, Field M, Skinner C, Stevenson RE, Bobrow M, Turner G, Schwartz CE, Gecz J, Raymond FL, Futreal PA, Stratton MR. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet. 2009; 41(5):535–543. [PubMed: 19377476]
5. Cromer BA, Gorman MA, Hansen G, Adams JJ, Coggan M, Littler DR, Brown LJ, Mazzanti M, Breit SN, Curmi PM, Dulhunty AF, Board PG, Parker MW. Structure of the Janus protein human CLIC2. J Mol Biol. 2007; 374(3):719–731. [PubMed: 17945253]
6. Board PG, Coggan M, Watson S, Gage PW, Dulhunty AF. CLIC-2 modulates cardiac ryanodine receptor Ca2+ release channels. Int J Biochem Cell Biol. 2004; 36(8):1599–1612. [PubMed: 15147738]
7. Meng X, Wang G, Viero C, Wang Q, Mi W, Su XD, Wagenknecht T, Williams AJ, Liu Z, Yin CC. CLIC2-RyR1 interaction and structural characterization by cryo-electron microscopy. J Mol Biol. 2009; 387(2):320–334. [PubMed: 19356589]
8. Mi W, Liang YH, Li L, Su XD. The crystal structure of human chloride intracellular channel protein 2: a disulfide bond with functional implications. Proteins. 2008; 71(1):509–513. [PubMed: 18186468]
9. Dulhunty AF, Pouliquin P, Coggan M, Gage PW, Board PG. A recently identified member of the glutathione transferase structural family modifies cardiac RyR2 substate activity, coupled gating and activation by Ca2+ and ATP. Biochem J. 2005; 390(Pt 1):333–343. [PubMed: 15916532]
10. Wu H, Min J, Zeng H, McCloskey DE, Ikeguchi Y, Loppnau P, Michael AJ, Pegg AE, Plotnikov AN. Crystal structure of human spermine synthase: implications of substrate binding and catalytic mechanism. J Biol Chem. 2008; 283(23):16135–16146. [PubMed: 18367445]
11. Petrey D, Xiang Z, Tang C, Xie L, Gimpelev M, Mitros T, Soto C, Goldsmith-Fischman S, Kernytsky A, Schlessinger A, Koh I, Alexov E, Honig B. Uning Multiple Structure Alignments, Fast Model Builing, and Energetic Analysis in Fold Recognition and Homology Modeling. Proteins. 2003; 53:430–435. [PubMed: 14579332]
12. Xiang Z, Honig B. Extending the Accuracy Limits of Prediction for Side-chain Conformations. J Mol Biol. 2001; 311:421–430. [PubMed: 11478870]
13. Alexov E, Gunner M. Calculated Protein and Proton Motions Coupled to Electron Transfer: Electron Transfer from QA- to QB in Bacterial Photosynthetic Reaction Centers. Biochemistry. 1999; 38:8253–8270. [PubMed: 10387071]
14. Georgescu R, Alexov E, Gunner M. Combining Conformational Flexibility and Continuum Electrostatics for Calculating Residue pKa's in Proteins. Biophys J. 2002; 83:1731–1748. [PubMed: 12324397]
15. Song Y, Mao J, Gunner MR. MCCE2: Improving Protein pKa Calculations with Extensive Side Chain Rotamer Sampling. Comp Chem. 2009; 30(14):2231–2247.
16. Zhang Z, Teng S, Wang L, Schwartz CE, Alexov E. Computational analysis of missense mutations causing Snyder-Robinson syndrome. Hum Mutat. 2010; 31(9):1043–1049. [PubMed: 20556796]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

17. Ponder, JW. TINKER-software tools for molecular design. 3.7. St. Luis: Washington University; 1999.

18. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. J Am Chem Soc. 1990; 112:6127–6129.

19. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem. 2005; 26(16):1668–1688. [PubMed: 16200636]

20. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem. 2009; 30(10):1545–1614. [PubMed: 19444816]

21. Jorgensen WL, Tirado-Rives J. The OPLS potential function for proteins. J Am Chem Soc. 1988; 110:1657–1666.

22. Schellman JA. Marcromolecular Binding. Biopolymers. 1975; 14:999–1018.

23. Yang A-S, Honig B. On the pH dependence of protein stability. J Mol Biol. 1993; 231:459–474. [PubMed: 8510157]

24. Hasselbalch KA. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. Biochemische Zeitschrift. 1917; 78:112–144.

25. Menke M, Berger B, Cowen L. Matt: local flexibility aids protein multiple structure alignment. PLoS Comput Biol. 2008; 4(1):e10. [PubMed: 18193941]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

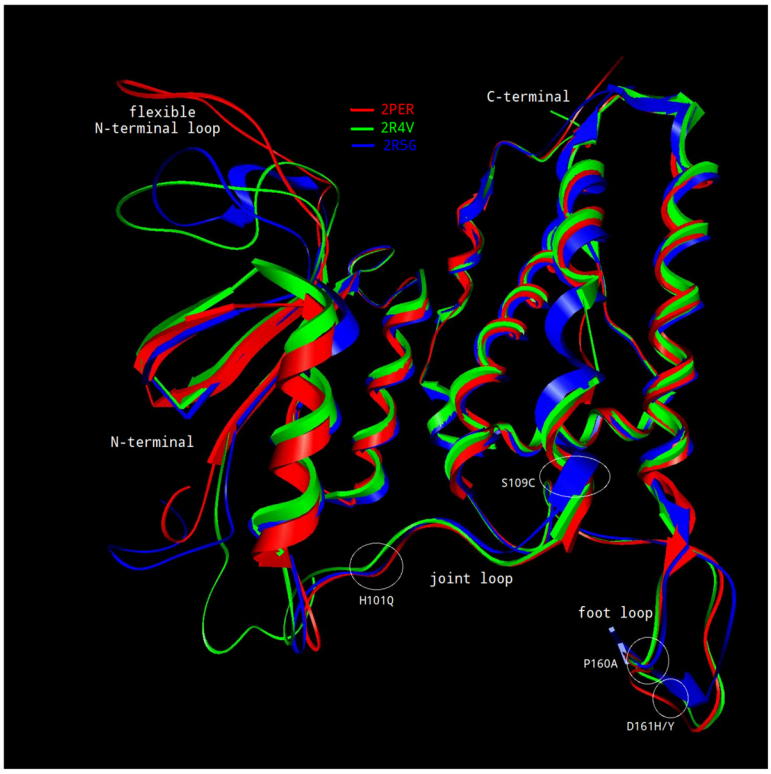

Fig. 1.
Overall 3D structure of CLIC2 protein revealed by structural superimposition of three available structures (2PER green, 2R4V red and 2R5G blue) shown in ribbons presentation. The distinctive structural features are marked and properly labeled.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

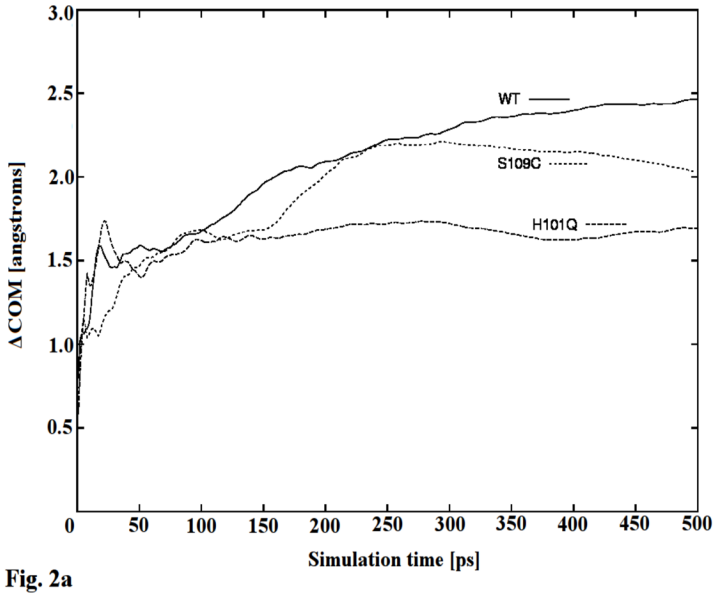

Fig. 2a

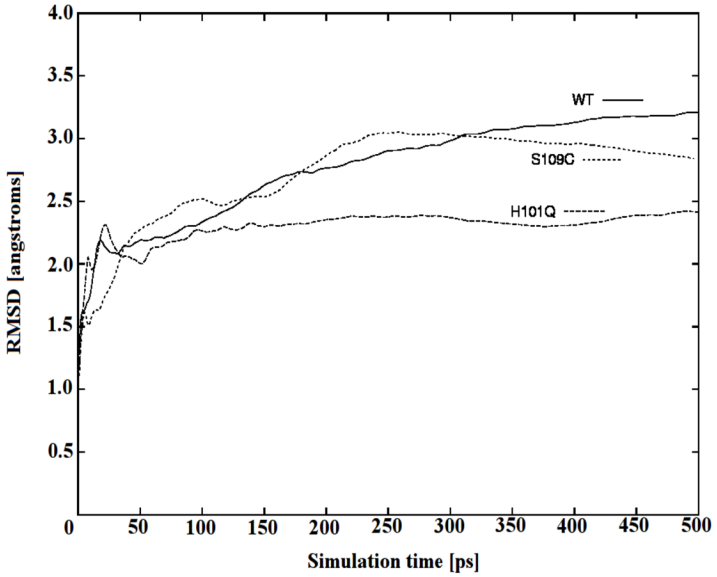

Fig. 2b

Fig. 2.

The results of MD simulations taken over the N-domain of CLIC2. The results were time-averaged and further averaged over the results obtained with each of the available three (WT) structures and the corresponding mutants (p.H101Q and p.S109C).

(a) the center of the mass (ΔCOM) of N-domain

(b) the RMSD of N-domain

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

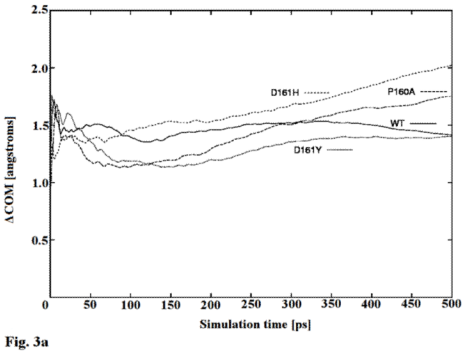

Fig. 3a

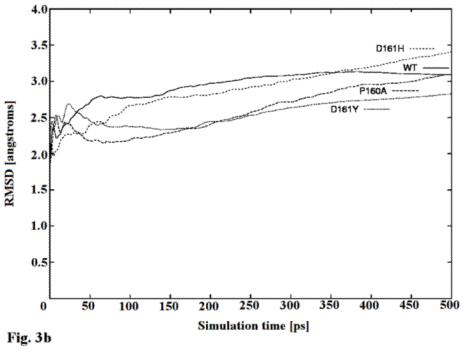

Fig. 3b

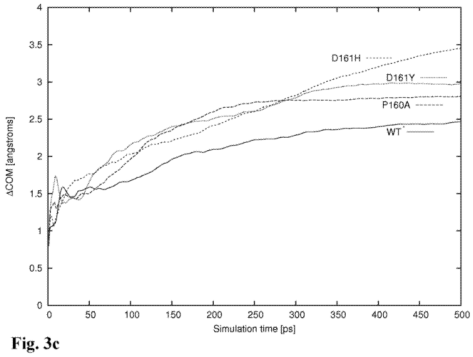

Fig. 3c

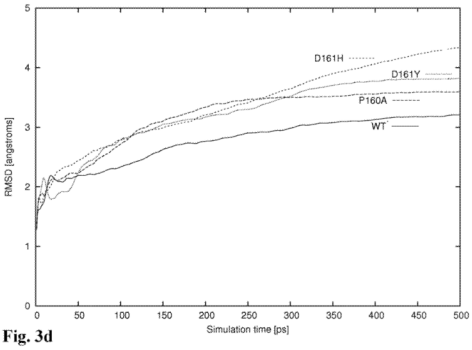

Fig. 3d

Fig. 3.

The results of MD simulations taken over the foot of CLIC2. The results were time-averaged and further averaged over the results obtained with each of the available three (WT) structures and the corresponding mutants (p.P160A, p.D161H/Y).

(a) the center of the mass (ΔCOM) of foot loop

(b) the RMSD of foot loop

(c) the center of the mass (ΔCOM) of the N-domain

(d) the RMSD of the N-domain

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Folding free energy results in kcal/mol. The effects of the proton/uptake release will add 1.37 to alterations that remove a His and subtract 1.37 with alterations that add a His.

Mutation	2PER (kcal/mol)	2R4V (kcal/mol)	2R5G (kcal/mol)	Average (kcal\mol)	Including the effect of proton uptake/release (kcal/mol)	Standard Error (kcal/mol)
p.H101Q	−0.34	−1.3	0.67	−0.32	1.05	1.95
p.S109C	−0.22	−2.04	−1.07	−1.11	−1.11	1.52
p.P160A	1.10	0.29	−1.39	0.00	0.00	1.46
p.D161H	4.66	−0.45	−4.74	−0.18	−1.55	2.47
p.D161Y	0.47	−2.69	−4.83	−2.35	−2.35	2.15