Molecular Dynamic Simulation of Neurexin1α Mutations Associated with Mental Disorder

Ashraf Hendam[1] · Ahmed Farouk Al-Sadek[1] · Hesham Ahmed Hefny[2]

Abstract

Neurexin1 gene is essential for formulating synaptic cell adhesion to establish synapses. In a previous work, 38 SNPs in Neurexin1 recoded in mental disorder patients have been collected. Five computational prediction tools have been used to predict the effect of SNPs on protein function and stability. Only four SNPs in Neurexin1α have deleterious prediction results from at least four tools. The current work aims to use molecular dynamic simulation (MD) to study the effects of the four mutations on Neurexin1α both on the whole protein as well as identifying affected domains by mutations. A protein model that consists of five domains out of six domains in the real protein was used; missing residues were added, and model was tested for quality. The MD experiment has last for 1.5 μs where four parameters have been used for studying the whole protein in addition to three more parameters for the domain analysis. The whole protein study has shown that two mutations E427I for Autism and R525C for non-syndromic intellectual disability (NSID) have distinctive behavior across the four used parameters. Domain study has confirmed the previous results where the five domains of R525C have acted differently from wild type (WT), while E427I has acted differently for four domains from wild type. The other two mutations D104H and G379E have three domains that only acted differently from wild type. The fourth domain of all mutations has an obvious distinctive behavior from wild type. Further study of E427I and R525C mutations can lead to better understanding of autism and NSID.

Introduction

Neurexin1 is a member of the neurexins family, which contains 3 members, and together they represent 0.1% of the human genome where Neurexin1 has the smallest length with 112,039 bp (Rowen et al. 2002). The sequence of this group is producing more than 1000 gene isoforms due to alternative splicing technique (Ullrich et al. 1995). A common feature between these genes is the production of two forms of the same gene which are the shorter β form and the longer α

form. Neurexin1 is the first member in this group where α form is consisting of 23 exons numbered 1–23, while β form has the last 6 exons of α form which are 18–23. Neurexin1α form has 5 splice sites SS1-SS5, while β form has the last two splice sites SS4 and SS5 (Rowen et al. 2002; Reissner et al. 2013). The promoter of the β form is found within the intron just by the end of α form (Missler et al. 1998a).

Neurexin1α protein consists of six 1–6 laminin neurexin sex (LNS) hormone binding domains, three 1–3 epidermal growth factor-like (EGF), signal peptide (SP), and transmembrane region (TM). The structure of the protein has three repeats; each repeat consists of two LNSs flanked by one EGF (Fig. 1) (Shen et al. 2008; Missler et al. 1998a).

Non-synonymous single nucleotide polymorphisms (nsSNP) of Neurexin1α and β forms have been reported in many studies especially those concerned with mental disorders. nsSNPs found in autistics have been reported in Persico and Napolioni (2013), Tordjman et al. (2017), and Tromp et al. (2021) and high concentration of nsSNP in SP region of Neurexin1β (Feng et al. 2006). Also Neurexin1 nsSNPs have been reported as a possible risk gene for schizophrenia

✉ Ashraf Hendam
ashraf.hendm@gmail.com

Ahmed Farouk Al-Sadek
afsadek@gmail.com

Hesham Ahmed Hefny
hehefny@ieee.org

[1] Agricultural Research Center, Giza, Egypt

[2] Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt
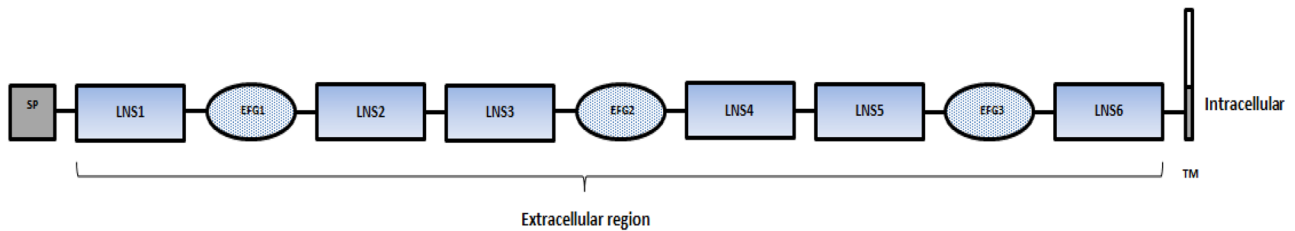
Fig. 1 Structure of Neurexin1α protein

(SCZ) (Rujescu et al. 2008; Gauthier et al. 2011; Tromp et al. 2021). Recently a study has investigated the relation between nsSNP in Neurexin1 as a possible risk for suicide death (William et al. 2021).

Molecular dynamics (MD) simulation is a technique that uses computer capabilities of high calculations to build a model for simulating molecules and atoms behavior under different conditions. MD has been used to study different molecular systems covering a large area of interest (Ryckbosch et al. 2017; Perilla and Schulten 2017; Hunter et al. 2018; Nadvorny et al. 2018).

MD has been widely used in identifying the effect of mutations on protein behavior and its association with diseases. The study of the two mutations in STN1 gene which are R135T and D157Y and their association with Coat Plus (CP) syndrome has been confirmed by running 100-ns all-atom MD (Amir et al. 2019). Shank3 protein binds to SAPAP protein, and this complex plays an important role in the formation of excitatory synapse and its plasticity. Mutations in both Shank3 and SAPAP are associated with some neurological diseases. The study of the mutation's effects on the binding of the two proteins was performed by MD experiment, which leads to better understanding of the basis on the molecular level of their binding (Piao et al. 2019). Another MD experiment was performed to study the effect of mutations in CA2 enzyme which lead to CA2-inherited enzyme deficiency disorder causing mental retardation and growth problems, among many other symptoms (Shaik et al. 2020). The appearance of Coronavirus pandemic (COVID-19) has led to the use of MD to help in the rapid discovery of drugs for this disease (Padhi et al. 2021). It represents an important pillar in a chain of steps, including molecular docking (Hosseini and Amanlou 2020; Chikhale et al. 2020; Kumar et al. 2020). ACE2 enzyme is binding to RBD domain to start the infection of SARS-CoV-2. Five mutations in RBD domain have studied using MD for 150 ns in order to be able to develop SARS-CoV-2 drugs (Dehury et al. 2021). Further, one mutation in RBD domain which is N501Y has been studied by MD for understanding molecular causes of increased affinity to hACE2 enzyme associated with SARS-CoV-2 (Luan et al. 2021).

The current work aims to use MD to study the effects of Neurexin1α mutations on whole protein in order to compare the results with the previously used computational tools and also to determine which of the five domains are affected by the mutations since domains contain binding sites of partner proteins and significant changes in their behavior. Changes can be an indicator of a possible domain malfunction and could be a cause of symptoms of psychiatric and neurological diseases such as autism and schizophrenia.

Methodology

Data Set

In a previous work (Hendam et al. 2020), a dataset consists of 38 nsSNPs in Neurexin1α and β forms that have been collected from different studies. Two groups of computational tools for predicting the effects of nsSNPs on protein have been applied. The first group includes PolyPhen-2 (Adzhubei et al. 2010), SIFT (Kumar et al. 2009), and PROVEAN (Choi and Chan 2015) for predicting the effects on protein function, while second group includes MUpro (Cheng et al. 2005) and I-Mutant 2.0 (Capriotti et al. 2005) for predicting the effects on protein stability.

Among the 38 nsSNPs used in the study, a full agreement of deleterious results has been achieved in two nsSNPs which are Glu715Ile and Arg813Cys as shown in Table 1. Two additional nsSNPs which are Asp392His and Gly667Glu have a full agreement of deleterious results between tools predicting the effects on protein function in addition to deleterious results of MUpro from the tools predicting the effects on protein stability as shown in Table 1. All the four nsSNPs have been found in Neurexin1α with NCBI gene reference transcript id NM_004801.4 with a size of 9368 pb, and the protein reference transcript id was NP_004792.1 with a length of 1477AA.

The four nsSNPs in Neurexin1α have been reported in two different studies and will be subject to more computational study in the current work to explore more functional impacts on the protein. Information about SNPs, such as location, SNP ID, found observation in sequenced persons, deleterious prediction results of tools, and nsSNPs source of the four mutations, are shown in Table 1.

Table 1 SNPs information, deleterious prediction results, and SNP source

SNP nucleotide position	SNP protein position	SNP ID	Observation	Functional tools			Stability tools		SNP source
				PROVEAN	SIFT	PolyPhen-2	MUpro	I-Mutant2.0	
c.1174G > C	p.Asp392His	NR	SCZ or control	√	√	√	√	X	Gauthier et al. (2011)
c.2000G > A	p.Gly667Glu	NR	Autism	√	√	√	√	X	Gauthier et al. (2011)
c.2143G > A	p.Glu715Ile	NR	Autism	√	√	√	√	√	Yan et al. (2008)
c.2437C > T	p.Arg813Cys	rs201150987	NSID	√	√	√	√	√	Gauthier et al. (2011)

NSID non-syndromic intellectual disability, *SCZ* schizophrenia, *NR* not registered

Method

As mentioned earlier, the four deleterious SNPs existed in the Neurexin1α. RCSB database contains PDB file for Neurexin1α which is 3poy.pdb and contains 1005 residues. Ten residues were missing from 3poy.pdb which are residues 292, 293, 551, 552, and from 810 to 815. To model the protein, including the missing residues SWISS-MODEL (Waterhouse et al. 2018) has been used, and the result was a model containing 1009 residues.

The resulted model contains five LNS domains starting from LNS2 to LNS6 in addition to two EGFs according to Swissprot Accession Q9ULB1 of Neurexin1α (Fig. 2).

In order to map positions of the four SNPs in the model, a sequence alignment was conducted between NP_004792.1 sequence and the model sequence.

The new locations of the SNPs according to the model file and SNP abbreviation are summarized in Table 2.

The model was validated using protein structure validation tool Prochek (Laskowski et al. 1993). Validation results have shown that residues in most favored regions are 85.6%, residues in additional allowed regions are 11.6%, residues in generously allowed regions is 2.0%, and only 0.8% of the residues were in the disallowed regions as shown in Fig. 3A, while the secondary structure of the model is shown in Fig. 3B.
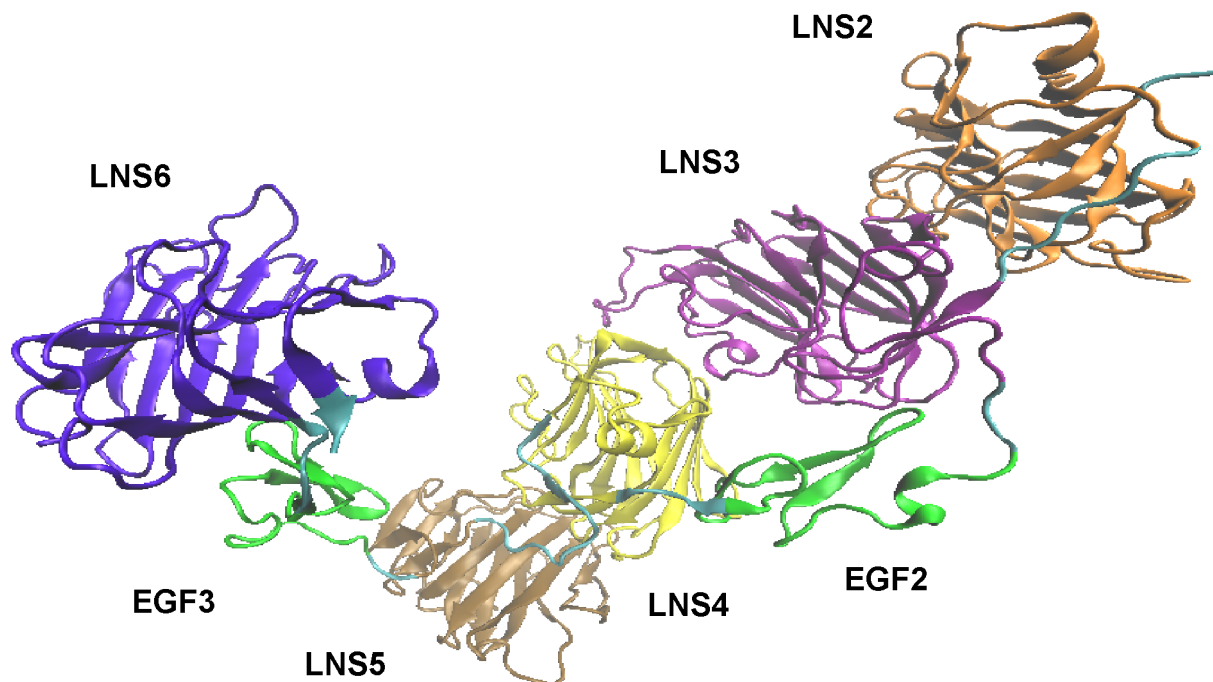

Fig. 2 LNSs and EFGs of Neurexin1α model

Table 2 Mapping SNPs to model sequence

Protein SNP position	Model SNP position	SNP abbreviation	Observation
Asp392His	Asp104His	D104H	SCZ or control
Gly667Glu	Gly379Glu	G379E	Autism
Glu715Ile	Glu427Ile	E427I	Autism
Arg813Cys	Arg525Cys	R525C	NSID

The mutated versions of the protein have been generated using SWISS PDB Viewer (Guex and Peitsch 1997). Figure 4 shows wild type (WT) protein and mutations.

The molecular dynamics simulation was conducted using Gromacs (Van Der Spoel et al. 2005; Pronk et al. 2013; Abraham et al. 2015) version 5.1.2. The file containing topology parameters of proteins was created by the Gromacs program with OPLS-AA/L force field. WT and four mutations were sunken in a cubic box of water molecules using extended simple point charge model (SPCE) water molecule model. Ions of Na + and Cl − have replaced water molecules in order to gain neutral system. The energy minimization was carried out to make sure that there are no clashes in protein geometry with the steepest descent minimization algorithm. To keep a steady number of pressure at 1 bar and temperature at 300 K, periodic boundary condition (PBC) has been implemented. The production of MD for WT and mutations was conducted for 1.5 μs.

The analysis of the MD results was based on extracting parameters from the MD trajectories. Four parameters will be common when conducting whole protein comparisons and domain comparisons which are root mean square deviation (RMSD), radius of gyration (Rg), number of hydrogen bonds (Hbond), and solvent accessible surface area (SASA), and three parameters will be used only in domain comparisons which are root mean square fluctuation (RMSF), secondary structure prediction using DSSP (Dictionary of Secondary Structure for Proteins), and principle component analysis (PCA).

The extraction of the parameters has been done by GROMACS commands. A consistent color scheme was used in the parameters where WT is black, D104H is magenta, G379E is green, E427I is red, and R525C is blue.

Results

Whole Protein Comparison

The analysis of the whole proteins was conducted using four extracted parameters which are RMSD, Rg, Hbond, and SASA.

RMSD

The individual comparison of the RMSD of the whole proteins (Fig. 5) has shown that RMSD values of WT have equilibrated around 250 ns and started the production phase to the end of the simulation where the RMSD values convergenced around 1.35 nm, while G379E has equilibrated faster than WT near 35 ns, and the production phase continues from this point to the end of the simulation where the RMSD values convergenced around 1.32 nm slightly lower than WT. D104H has a long equilibration time around 800 ns and the production phase from this point to the end of the simulation and RMSD values convergenced also around 1.4 nm slightly higher than WT and G379E. R525C tend to have higher RMSD values compared to WT with equilibration near 700 ns, and the convergence value was around 1.75 nm, while E427I is having a lower RMSD values than WT and has equilibrated close to 1200 ns, and its convergence value has been around 1.2 nm.

Rg

Rg is used to measure compactness of the protein. If the Rg values of a protein during MD are high, this implies that the protein is having a less compactness, while low Rg values are indicating high compactness of the protein. The Rg values of the whole proteins (Fig. 6) have shown that Rg values for WT have stabilized around 3.87 nm and D104H have stabilized near 3.90 nm, which is slightly higher than WT, while G379E is showing higher compactness than WT and D104H where it has stabilized around 3.77 nm. R525C is having higher Rg values indicating an evident of lower compactness than WT and stabilized after 700 ns around 4.17 nm, while E427I is having higher Rg values until 1150 ns; it converts its direction to be lower than WT and stabilized around 3.7 nm.

Hbond

The number of internal hydrogen bonds (Hbond) formed between protein residues is an indicator of protein stability. The Hbond of whole proteins, including WT, are shown in Fig. 7. WT is achieving the highest number of Hbond towards the end of the simulation compared to the
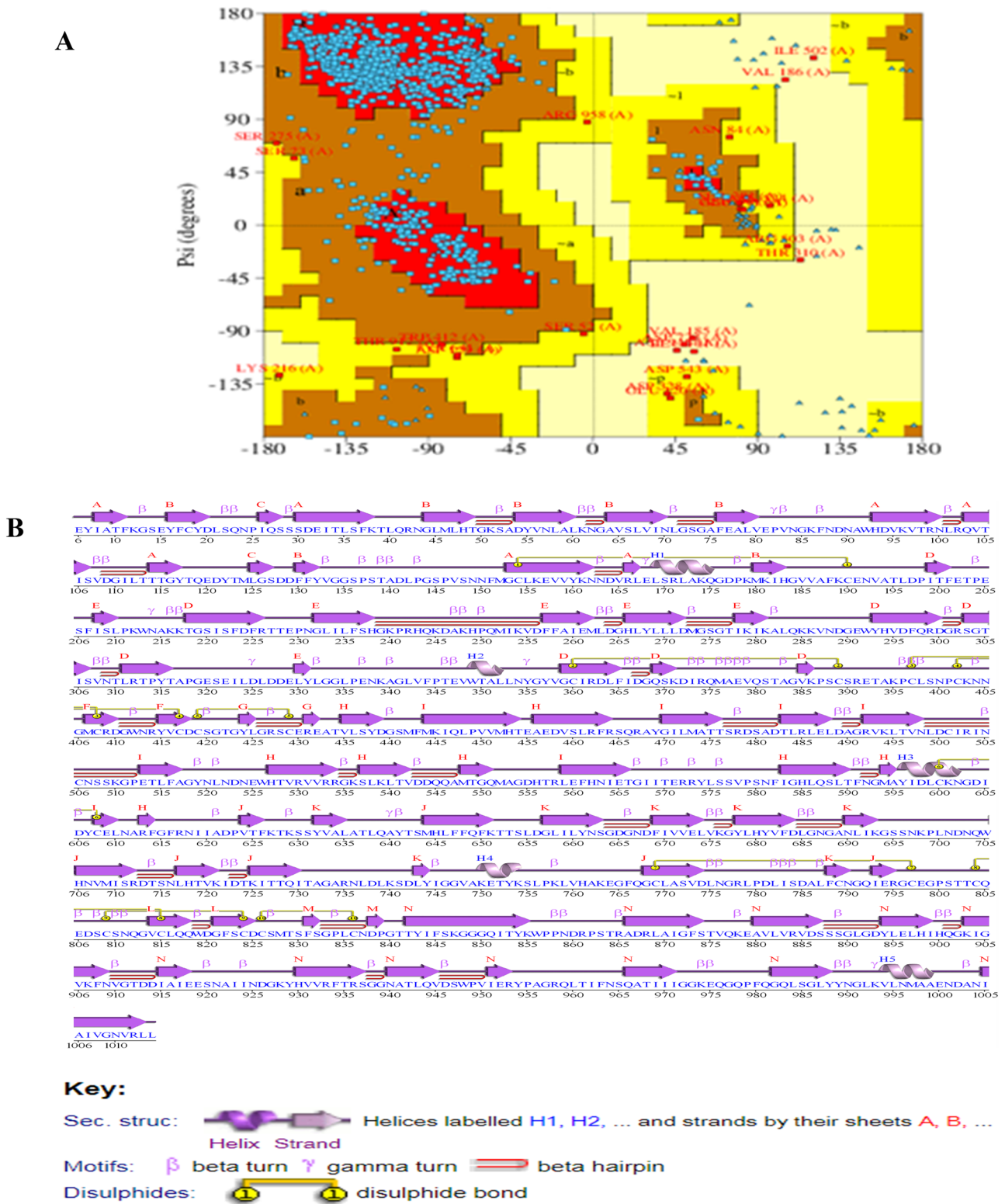
A

B

Key:

Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...

Helix Strand

Motifs: β beta turn γ gamma turn beta hairpin

Disulphides: disulphide bond

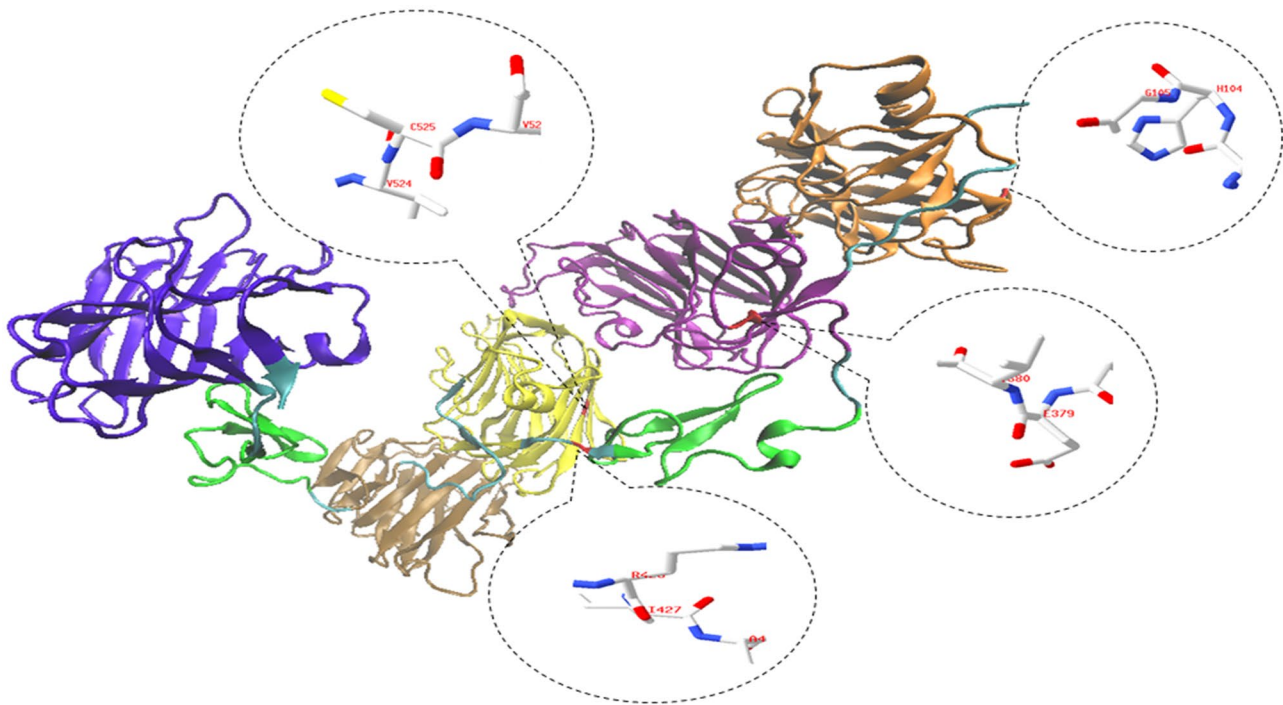**Fig. 3** **A** Prochek results. **B** Secondary structure of protein

Fig. 4 3D structure of the WT and mutations

four mutations, and the number of Hbond for stabilization is 740 bonds. The closest mutated protein to the WT is G379E towards the end of simulation where the highest achieved Hbond number is 775 bonds at 621.6 ns, and it has stabilized near 730 bonds, while D104H is achieving lower Hbond than both WT and G379E where the highest achieved Hbond number is 768 bonds at 854.79 ns. Both E427I and R525C are sharing the lowest achieved Hbond where both have stabilized around 700 bonds.

SASA

SASA values are an indicator of protein size where smaller SASA indicates a smaller size of the protein. WT has achieved the lowest SASA values most of the time of simulation compared to the four mutated proteins where the SASA values have stabilized near 430nm2 (Fig. 8). The nearest SASA values of mutated proteins to WT is D104H where the lowest achieved SASA value is 415.64

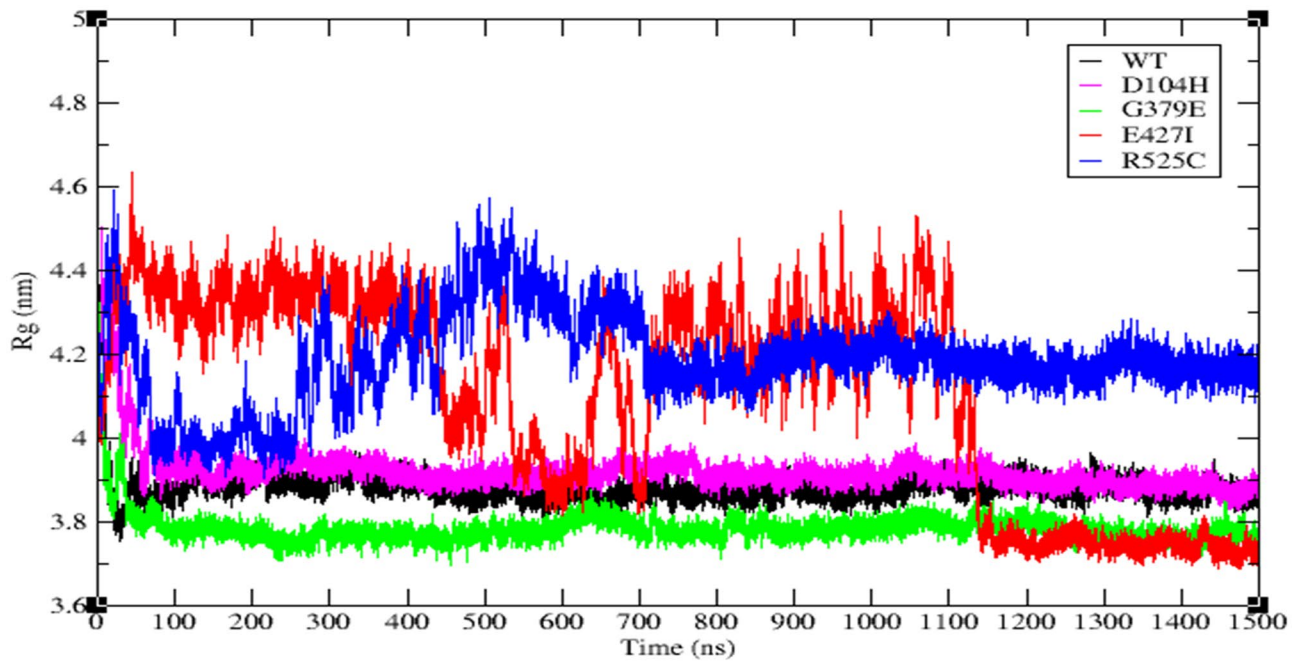**Fig. 5** Comparison of backbone RMSD for WT and four mutated proteins
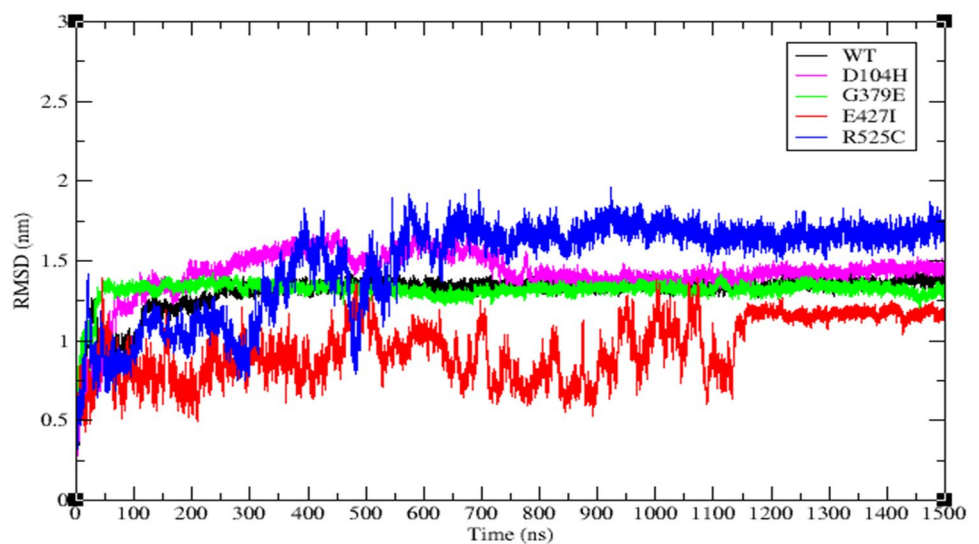
Fig. 6 Rg analysis of WT and 4 mutated whole proteins for 1.5 μs at 300 K

nm2 at 584.99 ns, and it has stabilized around 435 nm2. G379E and E427I have closer SASA values from 1150 ns and towards the end of the simulation where both have stabilized around 440nm2, while R525C values have stabilized near 435nm2.

Domain Comparison

Seven parameters have been used in order to compare between the behavior of the domains of the mutated proteins and WT domains.
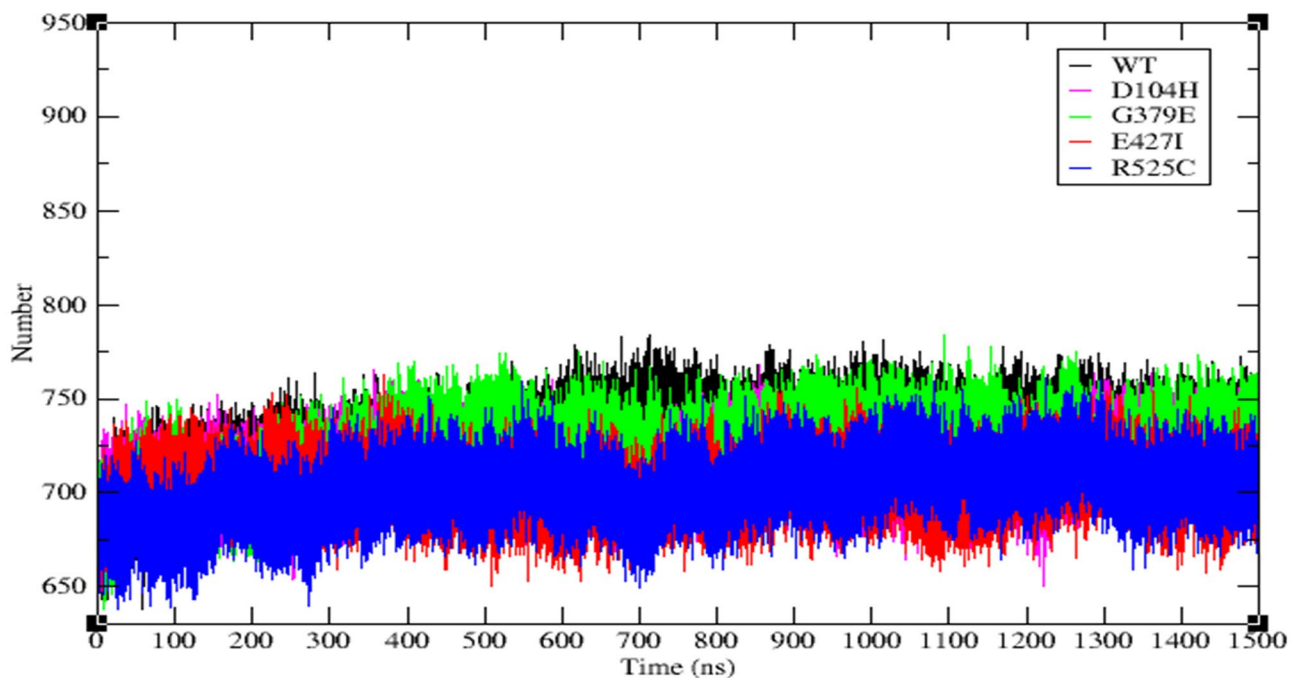

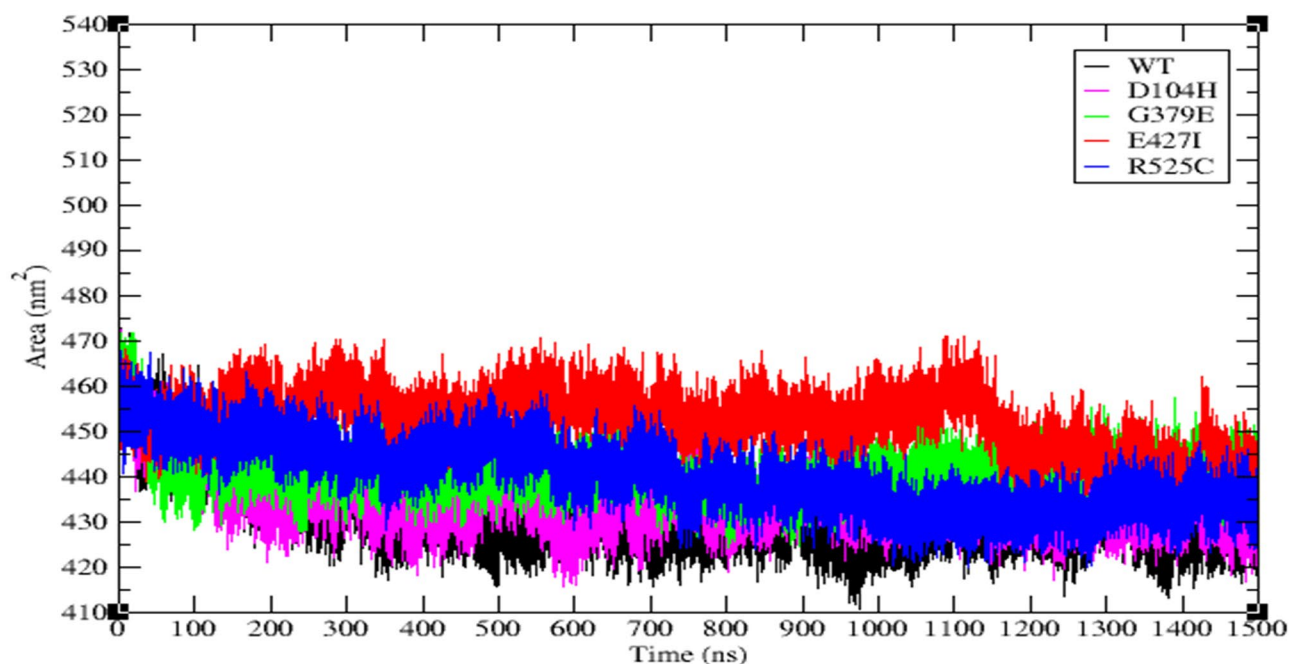**Fig. 7** Hbond of WT and 4 whole mutated proteins

Fig. 8 SASA of WT and 4 whole mutated proteins

RMSD

RMSD values of the five domains of WT and four mutated proteins have been shown in Fig. 9. The behavior of the 5 domains of WT is different from each other where LNS4 has acted uniquely. It equilibrated around 200 ns and starts the production phase after 200 ns to the end of the simulation where the RMSD values have convergenced less than 3.5 Å. LNS3 and LNS5 both have a shorter period to be equilibrated around 100 ns and continue with the production phase to the end of the simulation, and RMSD values have convergenced less than 2 Å, while LNS2 and LNS6 also equilibrated around 100 ns and RMSD values convergenced less than 1.5 Å.

LNS2 of D104H and G379E are acting very near to LNS2 of WT, and RMSD values have also convergenced, lower than WT and less than 1.5 Å. On the other hand, LNS2 of E427I and R525C are acting higher than LNS2 of WT and the other two proteins, but RMSD values of E427I have convergenced around 2 Å, while RMSD of LNS2 for R525C have convergenced higher than 1.5 Å.

RMSD of LNS3 for WT has convergenced around 1.75 Å which is the lowest achieved value compared to mutated proteins. R525C and E427I both have convergenced around 2 Å, and R525C has the nearest value to WT. The highest values have been achieved by G379E and D104H where both have convergenced around 2.25 Å.

LNS4 of WT has acted differently from all mutations where its RMSD values have convergenced not only earlier

than all mutations, but also higher with a value around 3.5 Å. D104H, G379E, and E427I have convergenced around 2.75 Å near the end of the simulation despite their fluctuation across the simulation time. R525C values have fluctuated and convergenced from 600 to 1400 ns, but near the end of the simulation, it drops to lower level.

The comparison of LNS5 has shown that WT, R525C, and D104H have convergenced around 1.75 Å, while G379E and E427I have convergenced in a higher level around 2 Å.

The RMSD values for LNS6 of WT and E427I have convergenced around 1.25 Å, while mutations have a slight difference compared to WT. D104H has acted similar to WT, but near 1400 ns, it turn to be higher than WT. G379E and R525C have convergenced close to 1.5 Å which is higher than WT.

RMSF

RMSF is used to study the structure and dynamics of the protein through measuring the flexibility of the protein residues. The comparisons between RMSF values for the domains of the four mutated proteins and WT have shown higher and lower values as shown in Fig. 10.

LNS2 of D104H has higher values for residues 80–90, 110–115, and 135–150 while lower values for residues 20–30. Higher residue's values for G379E are 80–90, 135–145, and 160–170, while lower values are 15–30 and 115–130. E427I has higher values for residues 70–90, 105–115, and 135–150 which are the highest achieved
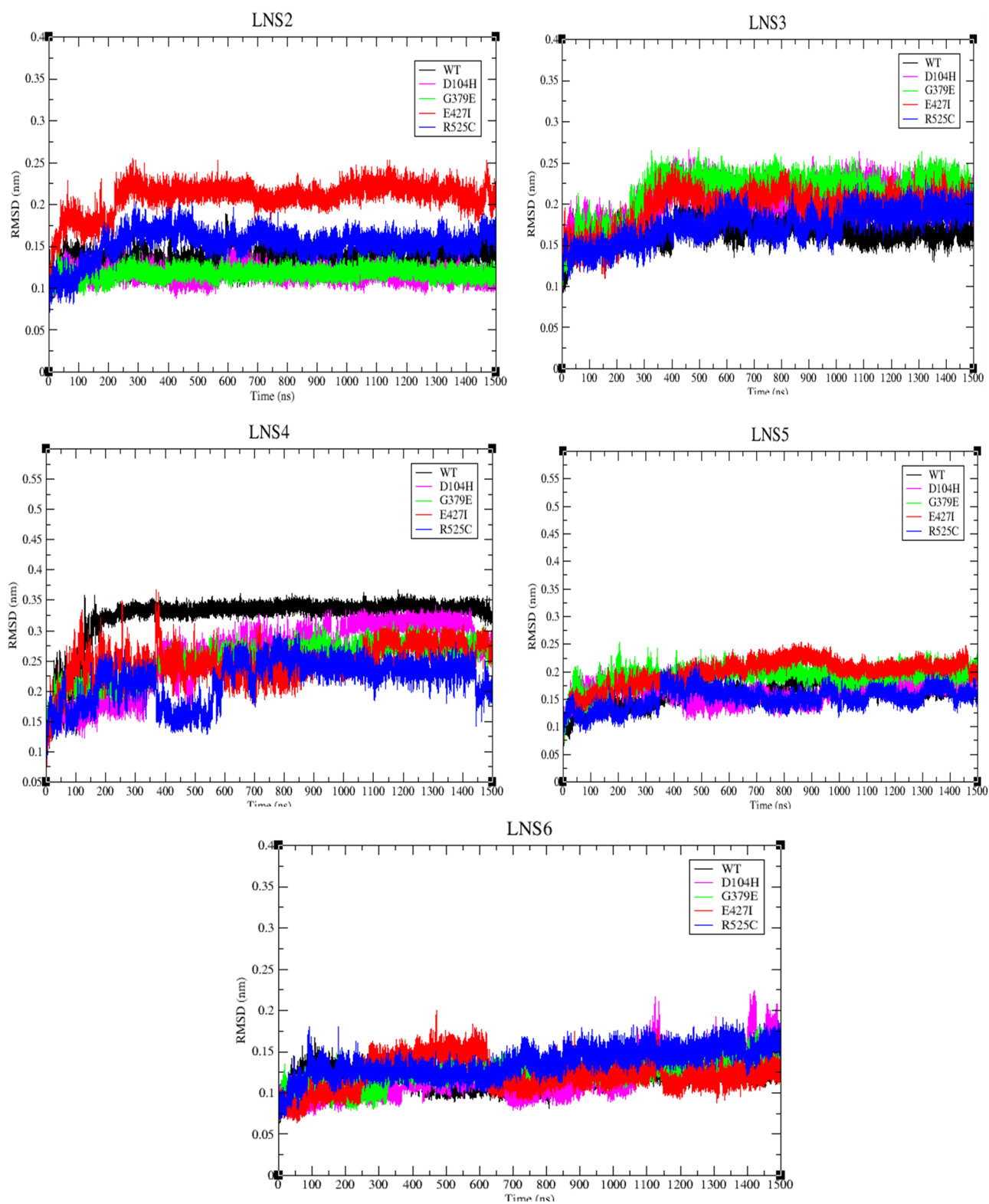
Fig. 9 Comparison of backbone RMSD for domains of WT and 4 mutations
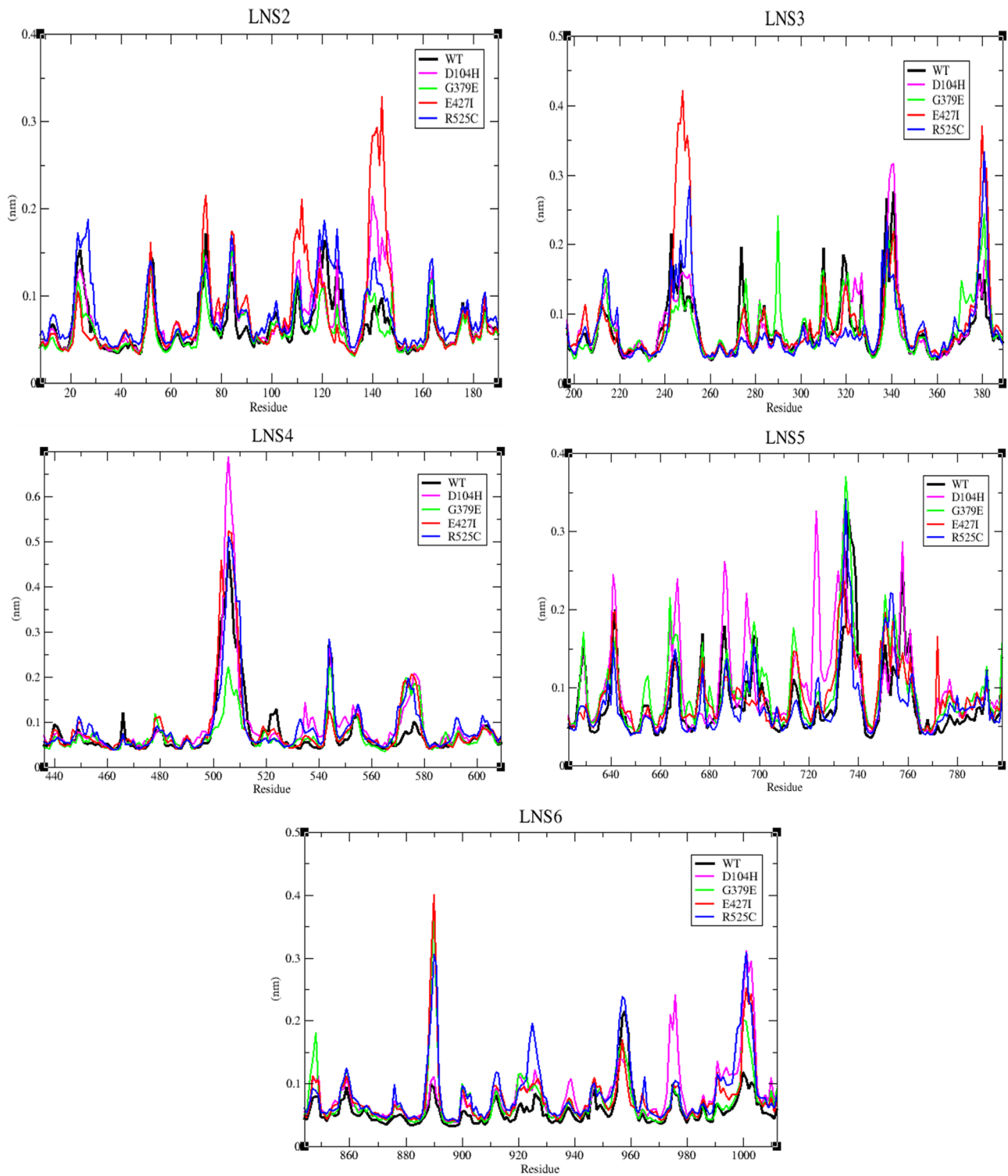
Fig. 10 RMSF of Cα atoms for domains of WT and four mutations

values between all mutated proteins, while lower values are for residues 20–30 and 116–130. R525C has higher values for residues 20–30, 80–90, 115–130, and 135–145 with almost no lower values.

RMSF values of LNS3 for D104H are higher values for residues 202–207, 210–220, 245–255, 322–327, and 335–342, while lower values are for residues 240–245, 272–277, 307–312, and 312–320. G379E has higher values

for residues 210–215, 288–292, and 370–385, while lower values are residues 202–207, 237–249, 272–275, 318–321, and 335–342. E427I has achieved higher values for residues 202–207 and the highest values among all mutated proteins for residues 245–255 and 377–382, while lower values are for residues 272–275, 315–321, and 335–342. Most of R525C residues have lower values than WT except for higher values for 212–215, 245–255, and 377–382.

Most of the RMSF residue values for LNS4 of D104H have higher values than WT. Residues 505–510, 535–543, and 567–580 are the highest among all mutated proteins, while lower values are for residues 465–468 and 518–525. G379E residues are acting close to WT residues except for residues 463–468, 500–515, 520–526, and 568–580 which are lower than WT. E427I is also closer to WT values except for residues 543–547 which are lower than WT, and 568–580 are higher than WT. R525C residues are higher than WT especially residues 568–580, and lower values are for residues 435–445, 464–468, and 518–522.

D104H has a higher RMSF values for LNS5 except for residues 675–680, 735–740, and 770–798 which have lower values than WT. D104H has the highest achieved values for residues 660–670, 685–697, and 712–735 among all mutated proteins.

Secondary Structure

The secondary structure has been drawn using DSSP. The secondary structure of WT and the four mutations can be found in Fig. 11. Changes in secondary structure have been found in two domains which are LNS2 and LNS4, while LNS3, LNS5, and LNS6 have no changes. For LNS2 a beta sheet in residues 169–175 with yellow color has disappeared in mutation E428I, and it does exist in WT and the other mutations (Fig. 12). Another beta sheet within residues 596–602 with cyan color in LNS4 has disappeared from mutation E428I, and it exists in WT and the other mutations (Fig. 13).

Rg

Rg is used to measure compactness of the protein. If the Rg values of a protein during MD are high, this implies that the protein is having a less compactness and the low Rg values are indicating high compactness of the protein. The results of the domain comparisons (Fig. 14) have shown that Rg values of LNS2 for D104H, G379E, and R525C have stabilized lower than WT around 1.50 nm, while E427I tend to stabilize higher than WT values near 1.51 nm. For LNS3 the Rg values of the four mutated proteins have stabilized near the stabilization value of WT.

The Rg values of WT for LNS4 are the lowest among all proteins where it has stabilized around 1.49 nm, while

D104H, G379E, and E427I proteins have stabilized around 1.52 nm, and R525C has stabilized higher than all near 1.53 nm.

WT has stabilized near 1.49 nm, similar to D104H and G379E for LNS5. R525C has achieved the lowest stabilized value which is 1.48 nm, while E427I has stabilized little higher around 1.485 nm. D104H and R525C are having higher Rg values than WT to stabilize around 1.52 nm for LNS6, while G379E and E427I are slightly around WT to stabilize around 1.50 nm.

Hbond

The number of internal hydrogen bonds (Hbond) formed between protein residues is an indicator of protein stability. The Hbond of the five domains for WT and 4 mutations are shown in Fig. 15.

G379E is acting very close to WT for LNS2 and both stabilized around 128 bonds, while D104H is having a slightly lower stabilized value than around 127 bonds, and both E427I and R525C have the lowest stabilized values near 126 bonds.

In LNS3 WT has stabilized around 123 bonds, and the stabilized values of the four mutations have varied. The highest stabilized value was for R525C which is around 128 bonds, and the lowest value was for E427I which is 118 bonds. D104 has stabilized around 124 bonds, and G379E has stabilized around 126 bonds.

The Hbond values of WT and E427I for LNS4 have stabilized close to each other near 126 bonds, while D104H and G379E have stabilized around 131 bonds, and R525C has stabilized very near to 122 bonds.

WT has stabilized near 123 bonds, which is the highest value achieved for LNS5 compared to mutations. The four mutations have stabilized close to each other between 116 and 118 bonds.

The WT stabilization value is around 123 bond which matches G379E value for LNS6, while E427I has stabilized higher than WT near 125 bonds. D104H and R525C have lower values than the previous proteins and stabilized around 116 bonds.

SASA

SASA values are an indicator of protein size, and hence smaller SASA indicates a smaller size of the protein. The SASA results of five domains of 4 mutations compared to WT are shown in Fig. 16. For LNS2 the highest SASA values are achieved by E427I which stabilized around 93nm2. WT and the other three mutations have stabilized close to each other around 90 nm2.

The SASA values for LNS3 have shown that R525C has the lowest values among all proteins and stabilized around
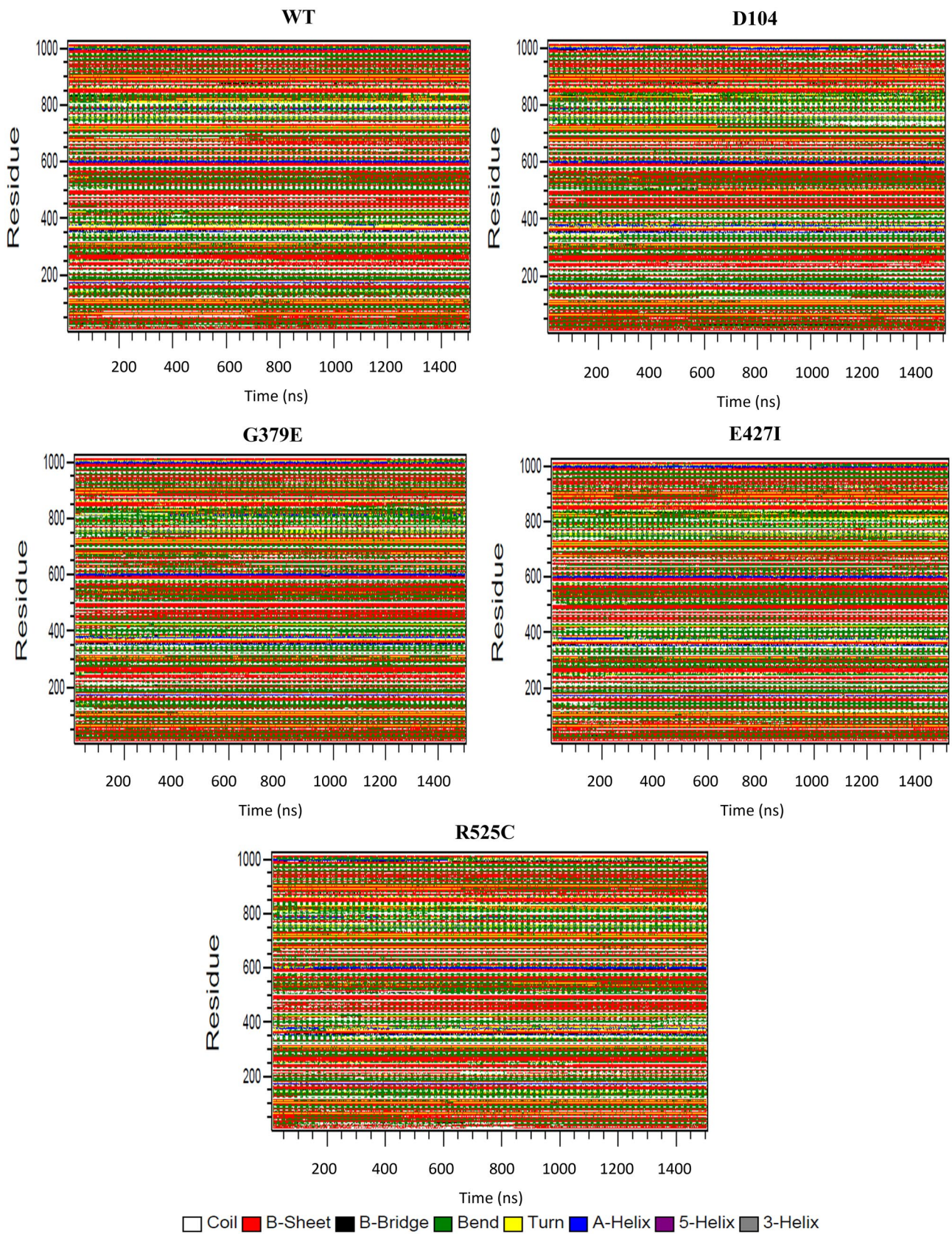
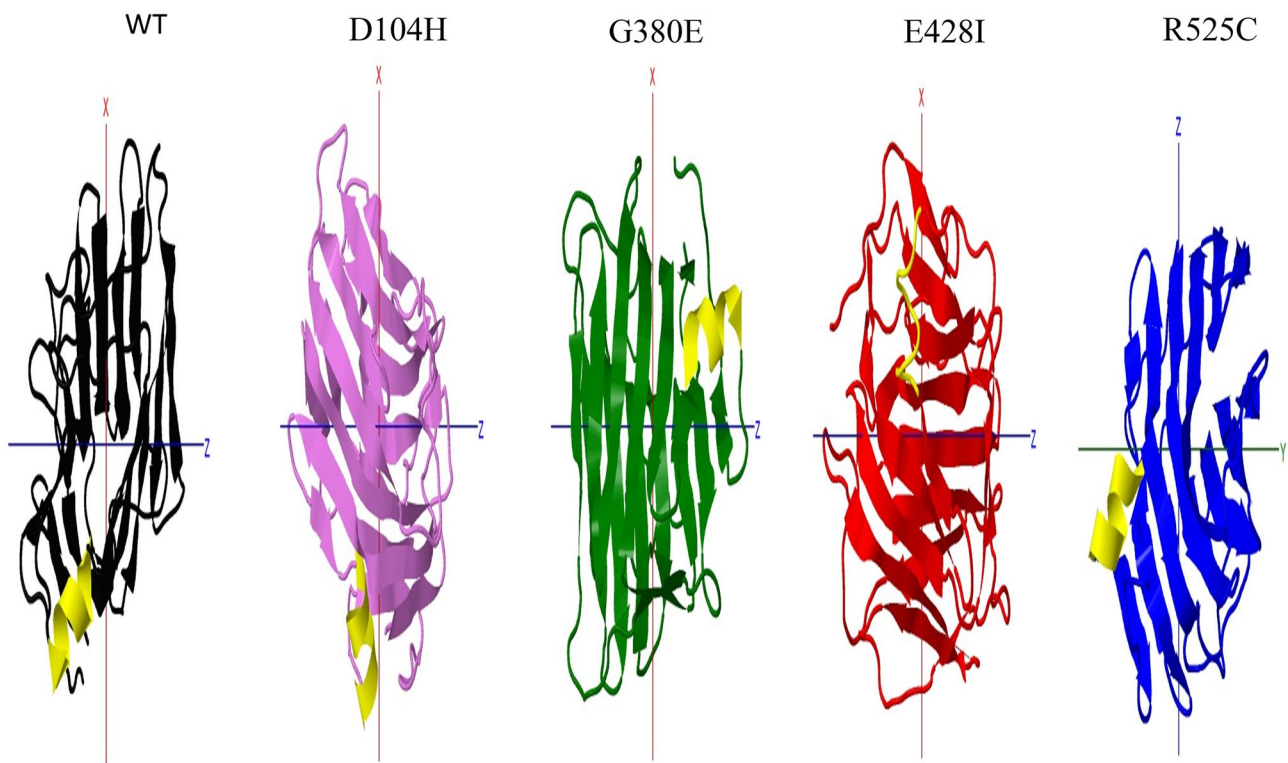**Fig. 11** Secondary structure of WT and 4 mutated whole proteins
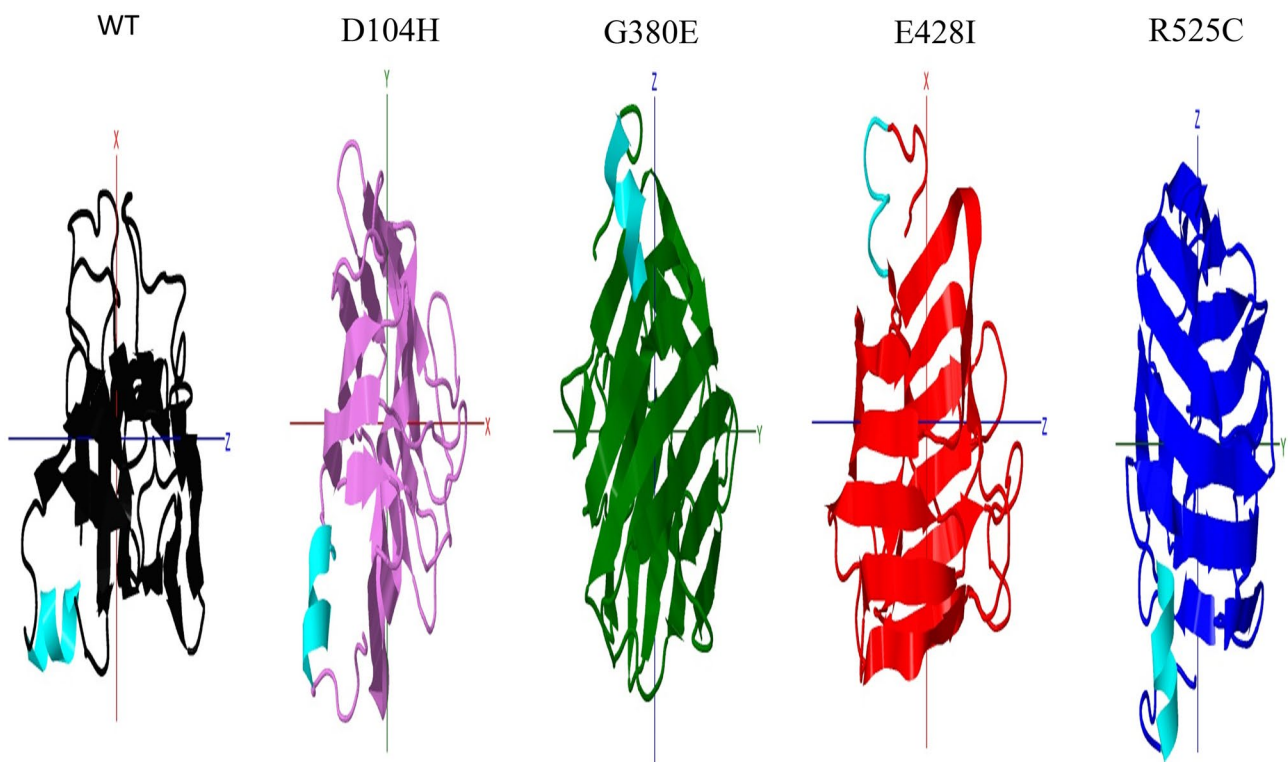
Fig. 12 LNS2 secondary structure changes in residues 169–175

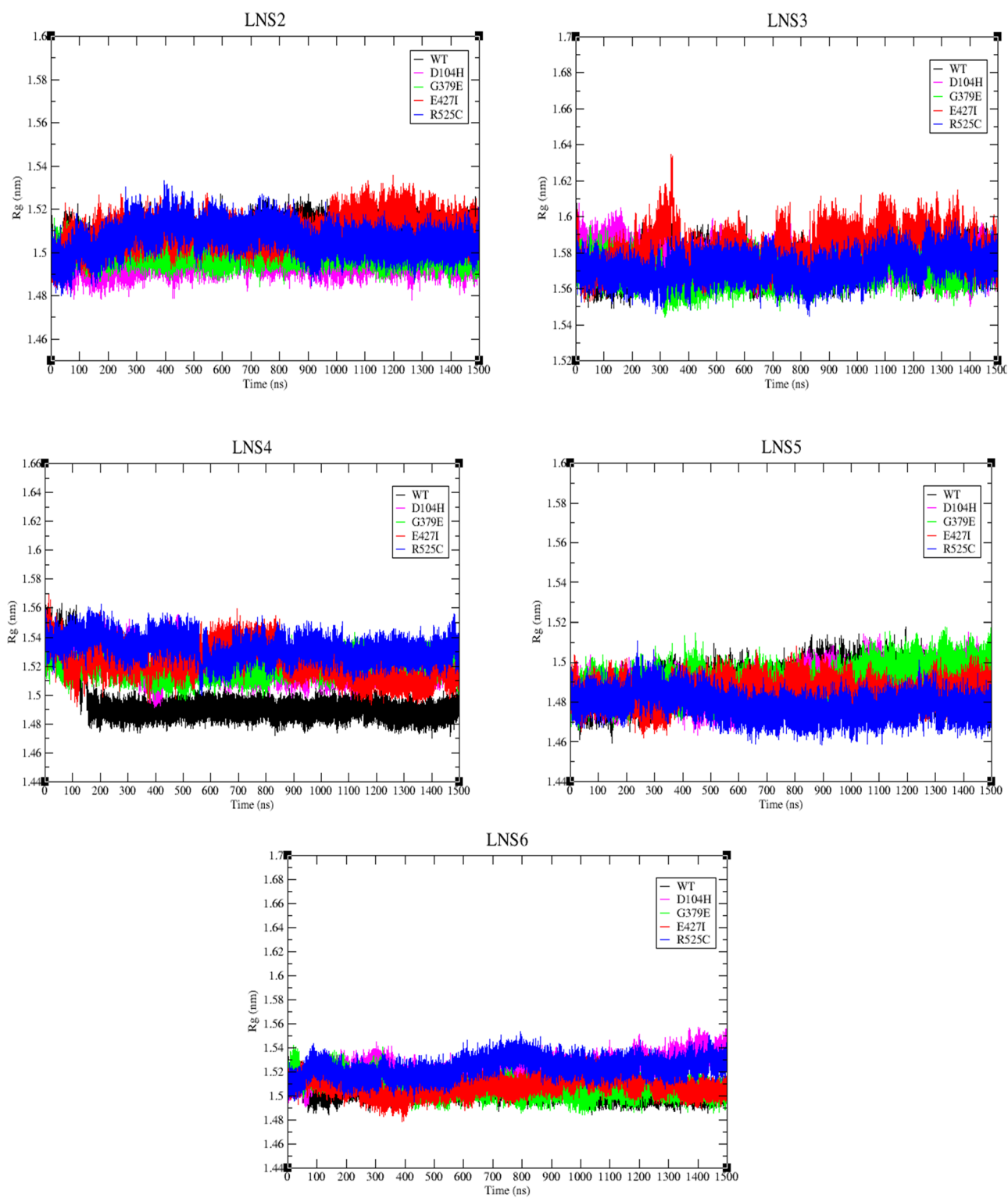**Fig. 13** LNS4 secondary structure changes in residues 596–602

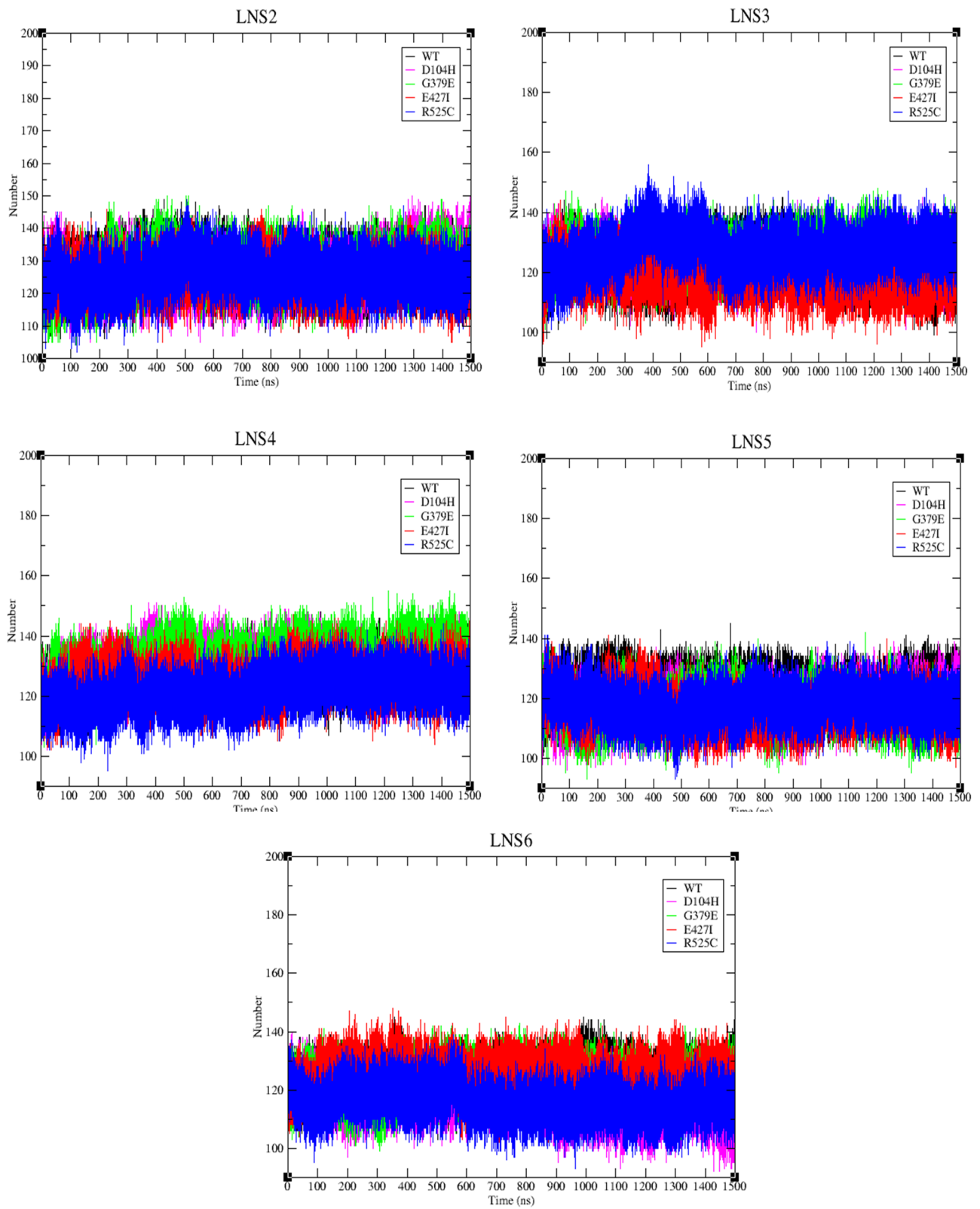**Fig. 14** Rg analysis of WT and mutations domains for 1.5 μs at 300 K

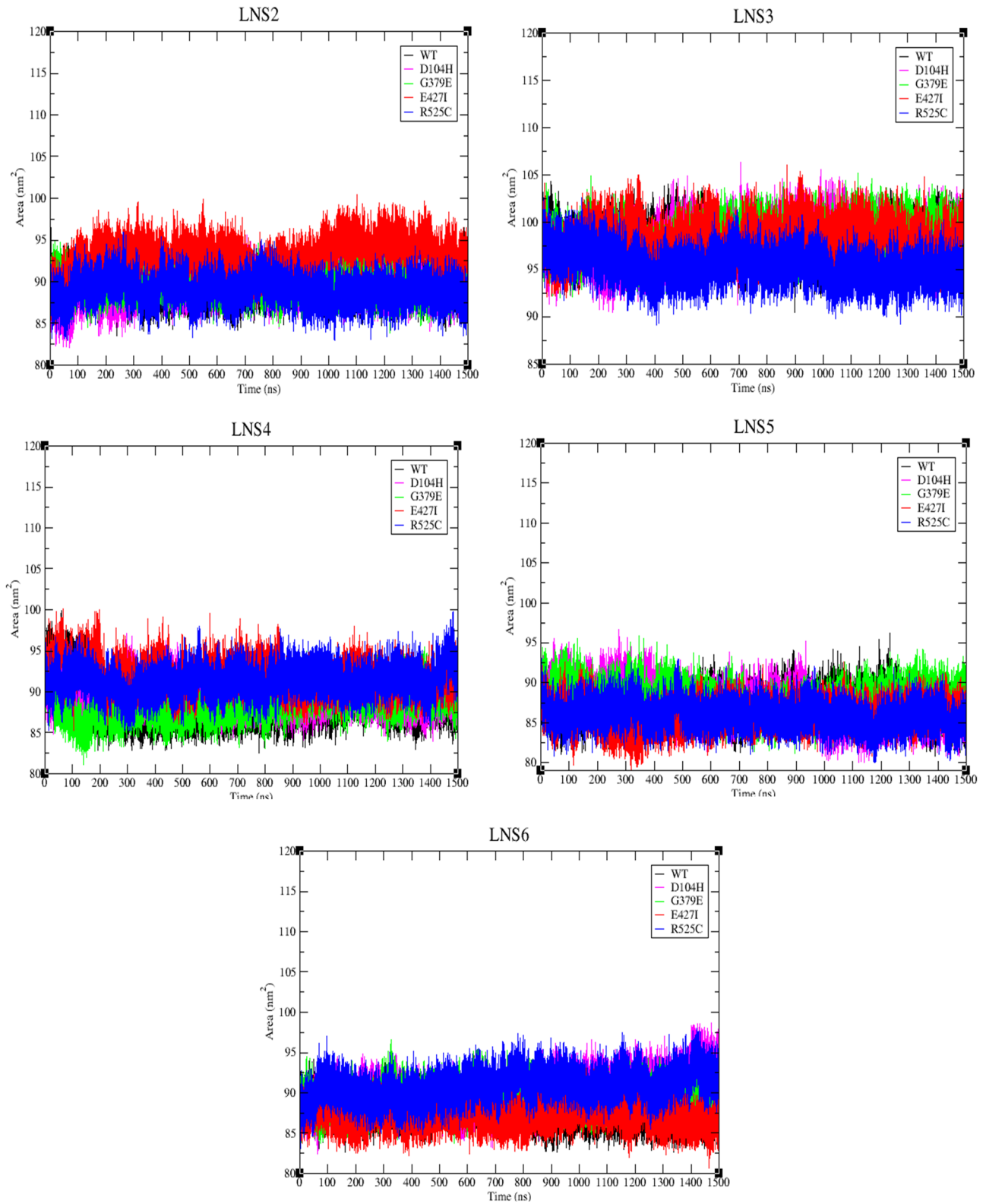**Fig. 15** Hbond of five domains for WT and 4 mutations

Fig. 16 SASA for the five domains of WT and 4 mutations

96nm2, while WT and the other three mutations have stabilized at a closer level near 98nm2.

WT has the lowest values for LNS4 compared with all mutations and stabilized around 88nm2. The highest stabilized values are for E427I and R525C around 92nm2, while D104H and G379E tend to act close to WT near 89nm2.

The values for LNS5 for WT and four mutations are close to each other. E427I and R525C have stabilized near 86nm2, while WT, D104H, and G379E have stabilized close to 88nm2.

R525C and D104H have stabilized close to 91nm2 which is the highest value for LNS6, while WT and E427I have stabilized near 88nm2, and G379E has stabilized in an average value near 90nm2.

PCA

PCA is determining the important modes of the protein and then defines the system motion in the form of motion scales (eigenvalues) and motion planes (eigenvectors). The projection of the first two eigenvectors (PC1 vs PC2) which have the biggest eigenvalues can describe the necessary subspace where protein dynamics can happen.

The comparisons of the five domains are shown in Fig. 17. E427I and R525C are covering larger areas of phase space compared to WT especially on PC1 plane in LNS2. For LNS3 E427I is covering a larger region of phase space for both PC1 and PC2 planes. E427I and R525C have covered a larger area of phase space in PC1 for LNS4. The behavior of WT and the four mutations are having closer regions of phase space for both PC1 and PC2 planes in LNS5 and LNS6.

Discussion

Whole protein comparison has been done through calculating four parameters for WT and four mutated proteins. The four parameters are RMSD, Rg for measuring protein compactness, Hbond for protein stability, and SASA for measuring protein size. The purpose of this comparison was to seek agreement or disagreement with previously used computational prediction tools.

The MD results have confirmed the prediction of the computational tools for E427I with autism observation and R525C mutations with NSID observation. Both mutations have caused great differences in protein stability, compactness, and size, but both are in the opposite direction of each other.

This different behavior could be the cause of different symptoms of each associated disease, since NSID causes mental retardation in persons with no abnormal physical appearance and has problems with adaptability, while autism

is a spectrum of complicated symptoms related to problems in behavior, communication, and ability to learn.

D104H has the most similar behavior to WT, and its changes are minor. This mutation has been found in two persons, one with SCZ, and the other is control. Minor changes with different recorded observations, including control, may indicate that this mutation by itself cannot cause an effect but rather needs another factor.

Autism mutations G379E and E427I have different behavior from each other; while G379E tends to act close to WT, the other mutation E427I has acted differently, which confirms the previous results of Table 1 in determining the severity of mutations despite having the same observation.

The LNS domains of Neurexin1α play an important role in interacting with other proteins to perform desired tasks. Neurexin1α interacts with dystroglycan which is a cell surface protein, through LNS2 and LNS6 to form cell adhesion in the intercellular region (Sugita et al. 2001). The interaction with dystroglycan occurs in the absence of insertion at splicing sites SS2 and SS4 (Reissner et al. 2014). Neurexophilin is interacting with Neurexin1α by binding to LNS2 domain (Missler et al. 1998b). Neuroligin proteins bind with Neurexin1α through LNS6 domain (Boucard et al. 2005) to form a complex that exists in the cleft between two neurons (Missler et al. 2012). The complex's function is not only to connect post- and presynaptic neurons, but also to aid in synapse formation and plasticity (Thalhammer and Cingolani 2014).

Neurexin1 also plays an important role as an organizer of the synaptic where the absence of the protein will affect some phenotypes of the synaptic such as the number and distribution of synapses between excitatory and inhibitory synapses (Chen et al. 2017). A network of Neurexin1 partner proteins is shown in Fig. 18.

Comparisons of the five domains for WT and four mutations have been done by calculating seven parameters, and a diversity of behaviors has been found for mutations. This will lead to a possible affected interaction with partner proteins.

D104H and G379E have similar behavior compared to WT confirmed by similarity in most residue flexibility for LNS2. The other two mutations have acted differently. Significant changes in E427I have been related to protein size, compactness, and stability, confirmed by high residue flexibility. R525C has acted differently from WT especially in higher residue flexibility, confirmed by low protein stability. Changes in E427I and R525C suggest that these mutations may have a major influence on the domain and can affect its function in binding with partner proteins dystroglycan and neurexophilin. Both proteins are competing with each other to bind with Neurexin1α (Reissner et al. 2014) which indicate that dystroglycan have been acting as synaptic cell adhesion (Jahncke and Wright 2022).
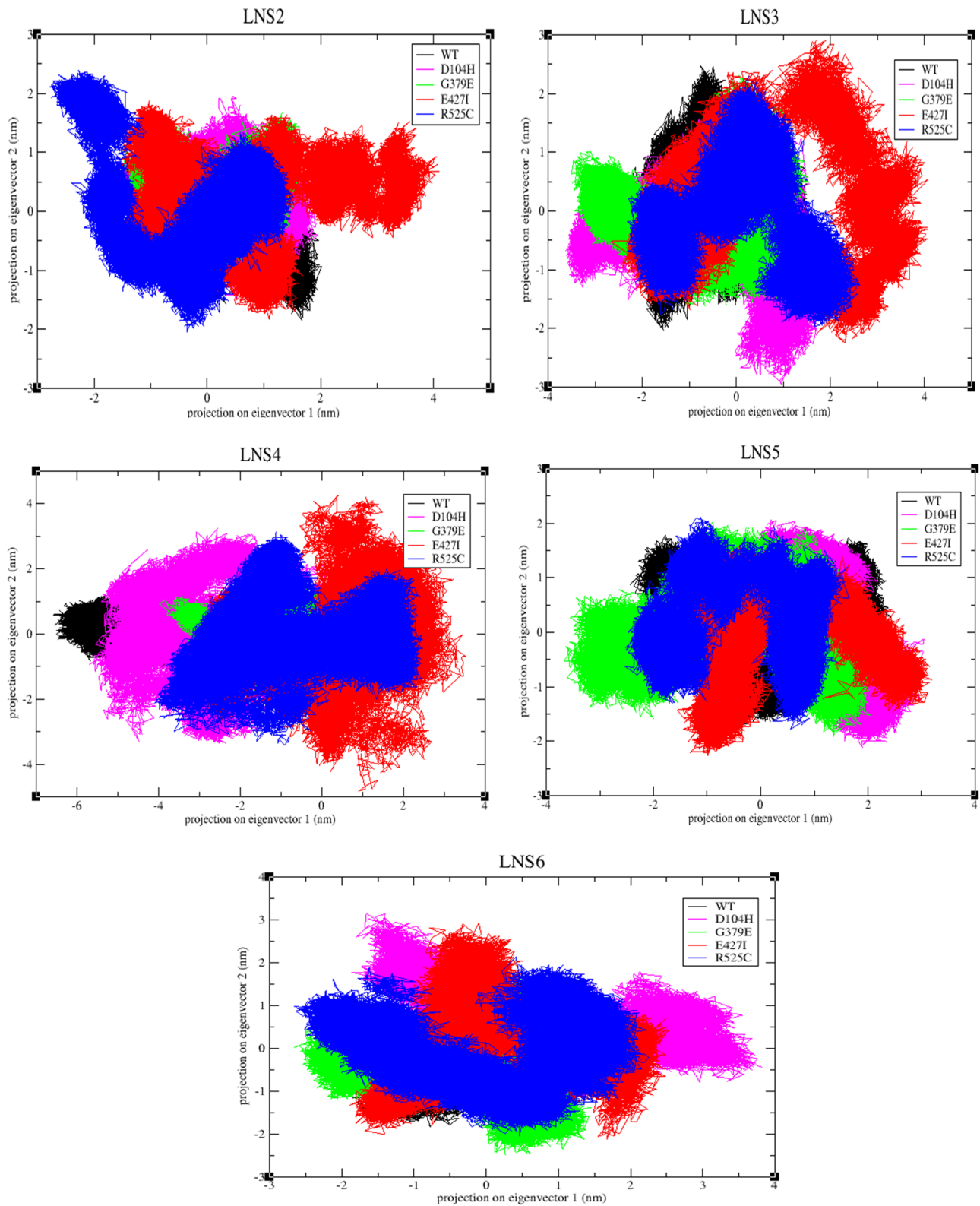
Fig. 17 Projection of the first two principal eigenvectors of the protein in phase space at 300 K
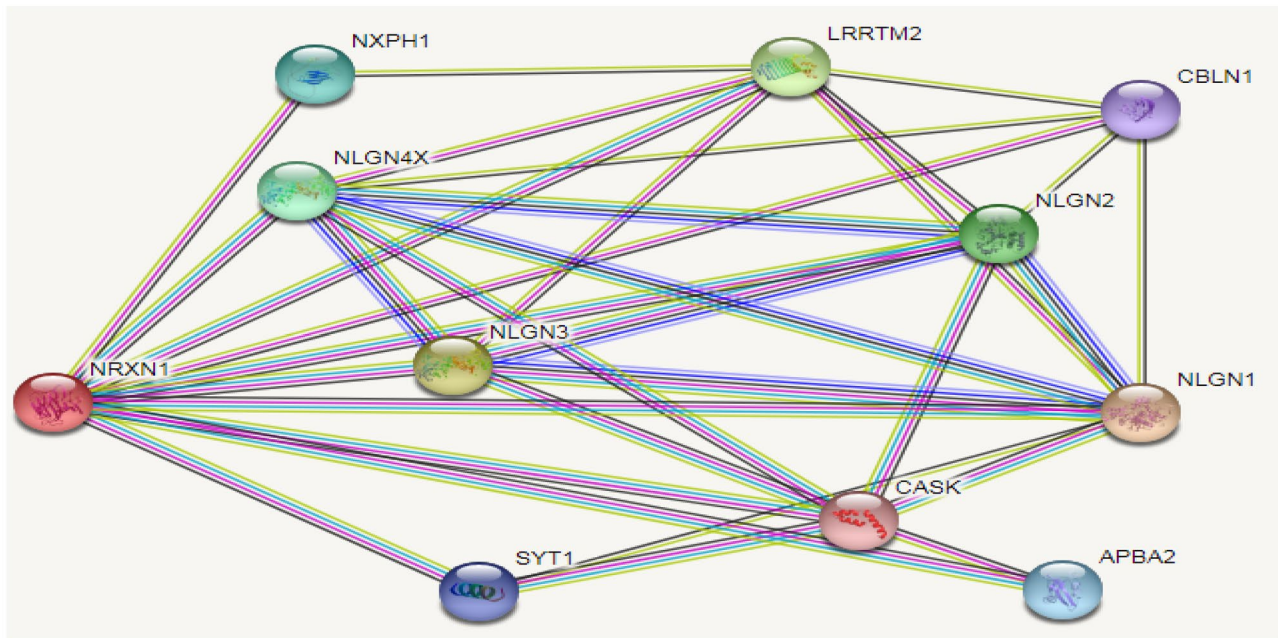
Fig. 18 Neurexin1 partner protein network

Mutations in the LNS3 domain have similar results as in the previous domain, where D104H and G379E have acted similarly to WT especially in protein compactness, stability, size, and PCA, while E427I and R525C have acted differently. E427I and R525C have distinctive behaviors according to changes in the calculated parameters, and again, both have acted differently from each other. E427I has the lowest protein stability and compactness, while R525C has the highest protein stability and smallest size.

The comparison of LNS4 has shown different behavior for all the four mutations compared to WT. D104H and G379E have acted similarly to each other with the highest protein stability, while D104H's residues have higher flexibility opposite to G379E. E427I tends to have the same protein stability as WT, while R525C has the lowest protein stability.

LNS5 of D104H has acted close to WT, and the other three mutations have acted differently. G379E has a larger protein size, while E427I has higher compactness and a smaller protein size. R525C has different harmonized values of the parameters where it has smaller size and highest compactness, and its residues have lower flexibility.

For LNS6 domain, D104H and R525C have acted similarly to each other for all the parameters, with different behavior from WT and the other two mutations.

G379E has the same compactness and stability as E427I, higher stability, and the smallest protein size. A summary of the affected domains in each mutated protein is shown in Table 3. This domain contains the binding site of dystroglycan and neuroligins.

From Table 3, the only mutation in which all domains have been affected is R525C mutation. This mutation has NSID observation, which is a mental retardation disorder. For this mutation, all its bindings to partner proteins, including dystroglycan which binds with LNS2 and LNS6 domains, neurexophilin which binds with LNS2 domain, and neuroligins which binds with LNS6 could be affected. These bindings have functions related to the development of synapses for regions inside the brain, which affect many functions of the brain. The severity of this mutation may be due to its effect on binding of all protein partners since Neurexin1α genetic variations are widely associated with neuropsychiatric disorders (Hu et al. 2019).

Table 3 Summary of the affected domains for the four mutations		LNS2	LNS3	LNS4	LNS5	LNS6	Observation
	D104H	X	X	√	X	√	SCZ or control
	G379E	X	X	√	√	X	Autism
	E427I	√	√	√	√	X	Autism
	R525C	√	√	√	√	√	NSID

LNS6 is the domain that differentiates between the effects of E427I with autism observation and the effects of R525C with NSID since E427I causes no changes in this domain. Both mutations have affected the rod-shaped assembly of NRXN1α which consists of the four domains LNS2–LNS5.

In a recent study (Ishizuka et al. 2020), a hybrid method has been used that combines an in silico approach and in vitro confirmation method. Through in silico approach, three mutations T737M, D772G, and R856W associated with autism and SCZ have shown a significant predicted damaging effects on the protein. The three mutations exist in LNS4. Also, 3D structural modelling has shown that two mutations could affect the rod-shaped assembly of the four domains LNS2-LNS5. In vitro confirmation method has revealed a low interaction level between NRXN1α and Neuroligin1 for two mutations. These results comply with MD results of E427I and R525C where both have affected LNS2-LNS5 domains, which constitute the rod-shaped assembly of the protein. In addition R525C exists in LNS4 and has affected LNS6. R525C could affect the formation of NRXN1α and Neuroligin1complex according to the previous study.

D104H and G379E have the lowest number of affected domains which are two domains.

The domain comparison has supported the whole protein comparison and gives an idea about the possible effect caused by mutations.

Conclusion

A simulation of WT and for mutated proteins has been conducted for a period of 1.5 µs in order to study the effect of the mutations on the whole protein and which domain(s) of the protein have been affected. Four parameters have been calculated for analyzing the whole protein results. E427I and R525C have shown distinctive behaviors across the four parameters, especially E427I which has the lowest RMSD and stability, while it has the highest compactness and largest size. D104H and G379E tend to act close to WT especially D104H where it has the same values as WT for RMSD, compactness, and size, while it has lower stability than WT. The study of the protein domains has been conducted using the previous four parameters on the domain level plus three new parameters. A common different behavior of LNS4 has been found for all mutations. R525C and E427I have different behaviors across domains, LNS2 to LNS5, while E427I tends to act similar to WT for LNS6 contrary to R525C which continues to have different behavior. D104H and G379E have acted similarly to WT for LNS2 and LNS3. LNS5 and LNS6 have differentiated between the two mutations, while D104H tends to act like WT for LNS5 G379E

has acted differently. On the other hand, G379E behavior for LNS6 tends to be similar to WT, while D104H has acted differently.

A real experiment is needed to confirm the results of R525C and E427I mutations on the binding partners of Neurexin1α which can lead to better understanding of autism and NSID.

Author Contribution Ashraf Hendam: Conducted the experience and wrote the draft, Ahmed Farouk: Contributed to ideas and revise, Hesham Hefny: Contributed to ideas and revise.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1:19–25

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7(4):248–249

Amir M, Mohammad T, Kumar V, Alajmi MF, Rehman M, Hussain A, Hassan M (2019) Structural analysis and conformational dynamics of STN1 gene mutations involved in coat plus syndrome. Front Mol Biosci 6:41

Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC (2005) A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α-and β-neurexins. Neuron 48(2):229–236

Capriotti E, Fariselli P, Casadio R (2005) 'I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure', *Nucleic Acids Research*, 33(Web Server), pp. W306–W310

Chen LY, Jiang M, Zhang B, Gokce O, Südhof TC (2017) Conditional deletion of all neurexins defines diversity of essential synaptic organizer functions for neurexins. Neuron 94(3):611–625

Cheng J, Randall A, Baldi P (2005) 'Prediction of protein stability changes for single-site mutations using support vector machines', *Proteins: Structure*. Function, and Bioinformatics 62(4):1125–1132

Chikhale RV, Gupta VK, Eldesoky GE, Wabaidur SM, Patil SA, Islam MA (2020) Identification of potential anti-TMPRSS2 natural products through homology modelling, virtual screening and molecular dynamics simulation studies. J Biomol Struct Dyn 1–16

Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinform 31(16):2745–2747

Dehury B, Raina V, Misra N, Suar M (2021) Effect of mutation on structure, function and dynamics of receptor binding domain of human SARS-CoV-2 with host cell receptor ACE2: a molecular dynamics simulations study. J Biomol Struct Dyn 39(18):7231–7245

Feng J, Schroer R, Yan J, Song W, Yang C, Bockholt A, Sommer SS (2006) High frequency of neurexin 1β signal peptide structural variants in patients with autism. Neurosci Lett 409(1):10–13

Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, Fathalli F (2011) Truncating mutations in NRXN2 and NEUREXIN1 in autism spectrum disorders and schizophrenia. Hum Genet 130(4):563–573

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. ELP 18(15):2714–2723

Hendam A, Al-Sadek AF, Hefny HA (2020) In silico deleterious prediction of nonsynonymous single nucleotide polymorphisms in Neurexin1 gene for mental disorders. Int J Bioinform Res Appl 16(1):1–24

Hosseini FS, Amanlou M (2020) Anti-HCV and anti-malaria agent, potential candidates to repurpose for coronavirus infection: virtual screening, molecular docking, and molecular dynamics simulation study. Life Sci 258:118205

Hu Z, Xiao X, Zhang Z, Li M (2019) Genetic insights and neurobiological implications from NRXN1 in neuropsychiatric disorders. Mol Psychiatry 24(10):1400–1414

Hunter NH, Bakula BC, Bruce CD (2018) Molecular dynamics simulations of apo and holo forms of fatty acid binding protein 5 and cellular retinoic acid binding protein II reveal highly mobile protein, retinoic acid ligand, and water molecules. J Biomol Struct Dyn 36(7):1893–1907

Ishizuka K, Yoshida T, Kawabata T, Imai A, Mori H, Kimura H, Ozaki N (2020) Functional characterization of rare NRXN1 variants identified in autism spectrum disorders and schizophrenia. J Neurodev Disord 12(1):1–16

Jahncke JN, Wright KM (2022) The many roles of dystroglycan in nervous system development and function. Dev Dyn

Kumar P, Henikoff S, Ng P (2009) 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm'. Nat Protoc 4(8):1073–1081

Kumar S, Sharma PP, Shankar U, Kumar D, Joshi SK, Pena L, Rathi B (2020) Discovery of new hydroxyethylamine analogs against 3CLpro protein target of SARS-CoV-2: molecular docking, molecular dynamics simulation, and structure–activity relationship studies. J Chem Inf Model 60(12):5754–5770

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26(2):283–291

Luan B, Wang H, Huynh T (2021) Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. FEBS Lett 595(10):1454–1461

Missler M, Fernandez-Chacon R, Südhof TC (1998a) The making of neurexins. J Neurochem 71(4):1339–1347

Missler M, Hammer RE, Südhof TC (1998b) Neurexophilin binding to α-Neurexins a single lns domain functions as an independently folding ligand-binding unit. J Biol Chem 273(52):34716–34723

Missler M, Südhof TC, Biederer T (2012) Synaptic cell adhesion. Cold Spring Harb Perspect Biol a005694

Nadvorny D, Soares-Sobrinho JL, de La Roca Soares MF, Ribeiro AJ, Veiga F, Seabra GM (2018) Molecular dynamics simulations reveal the influence of dextran sulfate in nanoparticle formation with calcium alginate to encapsulate insulin. J Biomol Struct Dyn 36(5):1255–1260

Padhi AK, Rath SL, Tripathi T (2021) Accelerating COVID-19 research using molecular dynamics simulation. J Phys Chem B 125(32):9078–9091

Perilla JR, Schulten K (2017) Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. Nat Commun 8:15959

Persico AM, Napolioni V (2013) Autism genetics. Behav Brain Res 251:95–112

Piao L, Chen Z, Li Q, Liu R, Song W, Kong R, Chang S (2019) Molecular dynamics simulations of wild type and mutants of SAPAP in complexed with Shank3. Int J Mol Sci 20(1):224

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Hess B (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinform 29(7):845–854

Reissner C, Runkel F, Missler M (2013) Neurexins. Genome Biol 14(9):213

Reissner C, Stahn J, Breuer D, Klose M, Pohlentz G, Mormann M, Missler M (2014) Dystroglycan binding to α− neurexin competes with neurexophilin-1 and neuroligin in the brain. J Biol Chem jbc-M114

Rowen L, Young J, Birditt B, Kaur A, Madan A, Philipps DL, Graveley BR (2002) Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. Genomics 79(4):587–597

Rujescu D, Ingason A, Cichon S, Pietiläinen OP, Barnes MR, Toulopoulou T, Murray R (2008) Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet 18(5):988–996

Ryckbosch SM, Wender PA, Pande VS (2017) Molecular dynamics simulations reveal ligand-controlled positioning of a peripheral protein complex in membranes. Nat Commun 8(1):6

Shaik NA, Bokhari HA, Masoodi TA, Shetty PJ, Ajabnoor GM, Elango R, Banaganapalli B (2020) Molecular modelling and dynamics of CA2 missense mutations causative to carbonic anhydrase 2 deficiency syndrome. J Biomol Struct Dyn 38(14):4067–4080

Shen KC, Kuczynska DA, Wu IJ, Murray BH, Sheckler LR, Rudenko G (2008) Regulation of neurexin 1β tertiary structure and ligand binding through alternative splicing. Structure 16(3):422–431

Sugita S, Saito F, Tang J, Satz J, Campbell K, Südhof TC (2001) A stoichiometric complex of neurexins and dystroglycan in brain. J Cell Biol 154(2):435–446

Thalhammer A, Cingolani LA (2014) Cell adhesion and homeostatic synaptic plasticity. Neuropharmacology 78:23–30

Tordjman S, Cohen D, Coulon N, Anderson GM, Botbol M, Canitano R, Roubertoux PL (2017) Reframing autism as a behavioral syndrome and not a specific mental disorder: implications of genetic and phenotypic heterogeneity. Neurosci Biobehav Rev 80:210

Tromp A, Mowry B, Giacomotto J (2021) Neurexins in autism and schizophrenia—a review of patient mutations, mouse models and potential future directions. Mol Psychiatry 26(3):747–760

Ullrich B, Ushkaryov YA, Südhof TC (1995) Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron 14(3):497–507

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Lepore R (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46(W1):W296–W303

William N, Reissner C, Sargent R, Darlington TM, DiBlasi E, Li QS, Coon H (2021) Neurexin 1 variants as risk factors for suicide death. Mol Psychiatry 1–10

Yan J, Noltner K, Feng J, Li W, Schroer R, Skinner C, Sommer SS (2008) Neurexin 1α structural variants associated with autism. Neurosci Lett 438(3):368–370