Journal of Molecular Graphics and Modelling 56 (2015) 31–42

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Rational design of hyper-glycosylated interferon beta analogs: A computational strategy for glycoengineering

Mojtaba Samoudi[a], Fatemeh Tabandeh[a], Zarrin Minuchehr[a,*], Reza Ahangari Cohan[b], Davoud Nouri Inanlou[b], Mahvash Khodabandeh[a], Mohammad Sabery Anvar[a]

[a] Institute of Industrial and Environmental Biotechnology (IIEB), National Institute of Genetic Engineering and Biotechnology (NIGEB), P.O. Box 14965/161, Tehran, Iran

[b] Department of Research and Development, Pasteur Institute of Iran, P.O. Box 1316942551, Tehran, Iran

ARTICLE INFO

Article history:
Accepted 1 December 2014
Available online 9 December 2014

Keywords:
Glycoengineering
Interferon beta
Comparative modeling
Molecular dynamics
In silico design

ABSTRACT

Glycoengineering has been successfully used to improve the physicochemical and pharmaceutical properties of therapeutics. One aspect of glycoengineering is to introduce new N-linked glycosylation consensus sequences (Asn, X, Thr/Ser) into desirable positions in the peptide backbone by mutational insertion to generate proteins with increased sialic acid content. In the current work, human interferon beta (huIFN-β) was used as a model to identify the potential positions for the addition of new N-glycosylation sites. A computational strategy was employed to predict the structural distortions and functional alterations that might be caused by the change in amino acid sequence. Accordingly, three-dimensional (3D) structures of the designed huIFN-β analogs were generated by comparative modeling. Molecular dynamics (MD) simulation was carried out to assess the molecular stability and flexibility profile of the structures. Subsequently, for the purpose of glycoengineering huIFN-β, analogs with 3D structures more similar to the wild-type huIFN-β and exposed Asn residue in the new N-glycosylation site were identified. These modeling procedures indicated that the addition of the new N-glycosylation site in the loop regions had lower constraining effects on the tertiary structure of the protein. This computational strategy can be applied to avoid alterations in the 3D structure of proteins caused by changes in the amino acid sequences, when designing novel hyper-glycosylated therapeutics. This in turn reduces labor-intensive experimental analyses of each analog.

1. Introduction

Interferons (IFNs) are considered important cytokines which have a multiplicity of biological actions including anti-viral, anti-proliferative, and immunomodulatory activities. These biological properties have been leveraged into clinical usage to treat various diseases such as hepatitis, multiple sclerosis (MS) and several cancers [1]. In the case of MS, the recombinant huIFN-β is considered as the first line of treatment [2]. Although fibroblasts were originally identified as a source of IFN-β when responding to a number of stimuli *in vitro*, but other cell types have also been found to produce IFN-β, which include endothelial and epithelial cells, and various leukocytes [3]. IFN-β belongs to type I interferon family along with IFNαs, IFNω and IFNτ. Type I interferons exert their biological activities by binding to the type I interferon receptor, which is composed of two subunits, IFNAR1 and IFNAR2 [1]. IFNγ is the only known member of the type II interferon family. HuIFN-β adopts the standard fold of type I interferons, which consists of five α helices labeled A–E. A long overhand loop, known as the AB loop, connects the A and B helices and can be conceptually subdivided into three segments: the AB1 (residues 23–35), the AB2 (residues 36–40), and the AB3 (residues 41–50) loops. Three shorter loops (named BC, CD, and DE) connect the rest of the helices [4]. Fig. 1 shows the 3D structure of huIFN-β with respect to regions that are involved in receptor binding interactions. It is created from the PDB file of a crystal structure using the Discovery Studio visualization tool [4–6]. The blue residues (B, BC, C2, D and DE1) correspond to the regions interacting with the IFNAR1 subunit, and the green residues (A2, AB1, AB2 and E) correspond to the regions

* Corresponding author at: Department of Systems Biotechnology, Institute of Industrial and Environmental Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), P.O. Box 14965/161, Tehran, Iran. Tel.: +98 21 44787363.

E-mail addresses: m.samoudi@nigeb.ac.ir (M. Samoudi), taban_f@nigeb.ac.ir (F. Tabandeh), minuchehr@nigeb.ac.ir (Z. Minuchehr), cohan_r@yahoo.com (R. Ahangari Cohan), d_nouriinanlou@yahoo.com (D. Nouri Inanlou), saba@nigeb.ac.ir (M. Khodabandeh), mohammad_sa@nigeb.ac.ir (M. Sabery Anvar).

http://dx.doi.org/10.1016/j.jmgm.2014.12.001
1093-3263/

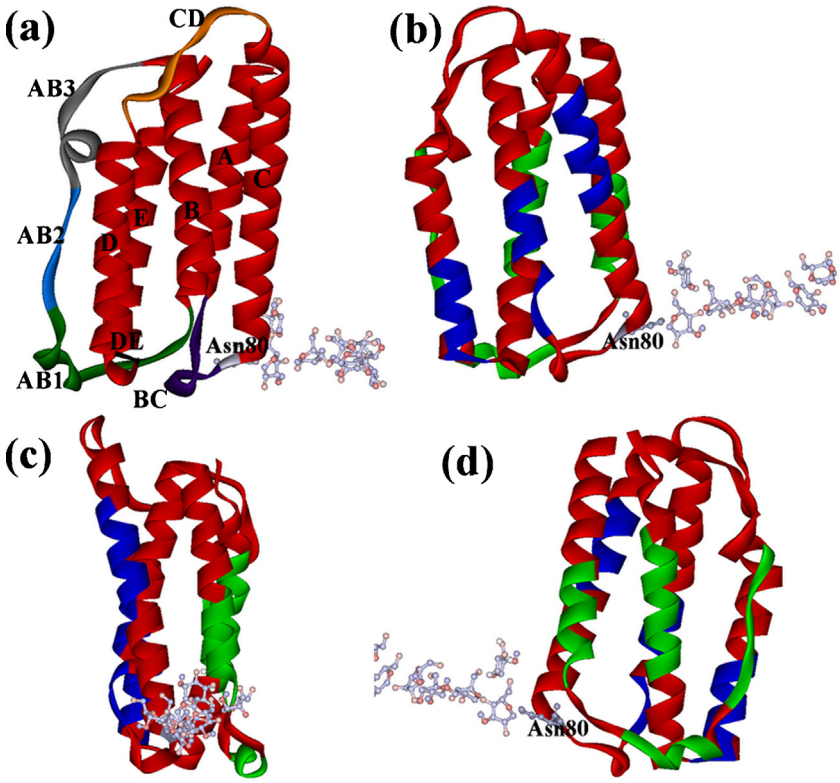

Fig. 1. (a) Crystal structure of huIFN-β visualized by the discovery studio program (Accelrys Software Inc., San Diego) from the PDB file of huIFN-β (PDB code: 1AU1, chain A). HuIFN-β adopts the standard fold of type I interferons and consists of five α helices (A–E) and connecting loops. AB loop connects helices A and B and is conceptually subdivided into three segments: the AB1 (green), the AB2 (blue), and the AB3 (gray) loops. Three shorter loops named BC (purple), CD (orange) and DE (black) loops connect the rest of the helices. A single N-glycosylation site exists at residue Asn-80 that involves a biantennary complex-type carbohydrate chain. (b) Front, (c) lateral and (d) back views of the crystal structure of huIFN-β with respect to the receptor binding regions. The blue regions correspond to the domain interacting with the IFNAR1 subunit and the green regions correspond to the domain interacting with the IFNAR2 subunit of the receptor.

interacting with the IFNAR2 subunit of the receptor. As illustrated in Fig. 1, the regions interacting with the IFNAR1 and IFNAR2 subunits are located on the opposite sides of the molecule.

HuIFN-β is a 166 amino acid glycoprotein with a non-glycosylated molecular weight of 18 kDa, which can increase to 23–25 kDa due to the presence of N-linked glycans at the single glycosylation site, Asn-80. The glycosylated IFN-β1a and non-glycosylated IFN-β1b produced in *Escherichia coli* and CHO cells, respectively, are both commercially available for clinical usage [7]. IFN-β1a is more potent and has been reported to have higher specific activity due to the stabilizing effect of the carbohydrate [7,8]. Additionally, glycosylation of huIFN-β reduces *in vitro* aggregation [7]. In spite of benefits resulting from glycan addition, the therapeutic usage of huIFN-β is limited. A major problem in MS patients treated with recombinant huIFN-β is the development of neutralizing antibodies (NAbs) against both IFN-β forms. This is primarily due to the low clinical effectiveness which increases the frequency of IFN-β administration [9]. Protein-based drugs routinely display suboptimum therapeutic efficacies due to their poor physicochemical and pharmacological properties [10]. Different strategies have been employed to improve therapeutic behavior of protein drugs. Glycoengineering is one of the most promising procedures as it has been shown to simultaneously improve most of the parameters necessary for optimization the *in vivo* efficacy of protein drugs. This approach has been successfully used to improve the biological activity of a number of therapeutics by increasing molecular stability and serum half-life [11,12]. The addition of glycans to a protein can also decrease aggregation by increasing solubility and/or masking the hydrophobic patches on the protein surface [7]. Specifically, protein aggregation is a challenging problem for recombinant huIFN-β due to the high degree of hydrophobic interactions among exposed residues located on the molecular surface [13].

One aspect of glycoengineering is to introduce new N-glycosylation consensus sequences (Asn, X, Thr/Ser) into desirable positions in the peptide backbone, so as to generate proteins with increased carbohydrate content. The glycan moiety can attach to the protein through the Asn residue in the consensus tripeptide sequence. However, the addition of a new tripeptide sequence will change the amino acid sequence of the protein. Such a modification must not alter the 3D structure, stability and receptor–ligand interaction of the protein. Inappropriate positions can affect folding and destroy the 3D structure of the active protein [11]. Additionally, factors such as surface accessibility of the N-linked Asn residue and probability of the enzymatic glycosylation of the new N-glycosylation site must be considered. Experimental screening of all possible analogs would be time-consuming and costly. Therefore, rational selection of suitable positions for introducing new N-glycosylation sites is very valuable. The main goal of this research is to develop a computational procedure for designing, evaluating and screening novel hyper-glycosylated huIFN-β analogs in order to find molecules in which changes in amino acid sequences have the least effect on the 3D structure, stability and receptor binding, whilst ensuring the glycosylation process is carried out efficiently. Due to the high degree of aggregation encountered during the production of huIFN-β and associated poor clinical outcomes, huIFN-β was selected as the model protein. Analogs of huIFN-β carrying one additional N-glycosylation site were examined with regard to molecular stability, flexibility regions, similarity to the wild-type huIFN-β and surface area accessibility of the new Asn residue. This computational procedure will determine if the designated analogs have increased numbers of carbohydrate chains

while retaining their structure and function, and can be applied to other bio-pharmaceutically important proteins early in the development process to avoid negative clinical outcomes.

2. Methods

2.1. Online servers and databases

Prediction of Asn glycosylation efficiency was carried out by theNetNGlyc1.0 server [14] (http://www.cbs.dtu.dk/services/NetNGlyc/). Accessible Surface Area (ASA) measurements of residues were carried out using the ASAview software/server, a graphical program for solvent accessibility of proteins [15] (http://www.abren.net/asaview/). Model validation was carried out using the Structural Analysis and Verification Server (SAVES) from the NIH MBI Laboratory for Structural Genomics and Proteomics (http://nihserver.mbi.ucla.edu/SAVES/). Multiple Sequence Alignment (MSA) was performed using the Clustal Omega software/server, and alignments were examined manually [16] (http://www.ebi.ac.uk/Tools/msa/clustalo/). Protein sequences were obtained from the UniProt database (http://www.uniprot.org/). The X-ray crystallography structure of huIFN-β was obtained from RCSB protein data bank (PDB: 1AU1) (http://www.rcsb.org/pdb/home/home.do).

2.2. Software

The comparative-based modeling was carried out by the MODELLER 9v10 software [17]. MD simulation and energy minimization were performed using GROMACS version 4.5.3 [18] with the OPLS-AA force field [19]. Discrete Optimized Protein Energy (DOPE) score profile of each analog was checked using MODELLER 9v10. The whole-structure RMSD (Root Mean Square Deviation) value was measured using the Swiss-PDB Viewer software [20]. Protein structure visualization and animations were rendered by VMD [21] and Accelrys Discovery Studio [22].

2.3. Design and selection of potential N-glycosylation sites

The primary sequence of huIFN-β was obtained from UniProt (UniProtKB: P01574). The huIFN-β sequence contains 187 amino acids. The signal peptide, comprising the first twenty one amino acids located at the N-terminal was removed from the sequence before theoretical analysis, as it does not contribute to protein folding. The choice of possible locations for introduction of new N-glycosylation sites was based on the analysis of the 3D crystallographic structure of huIFN-β and a literatures review for finding non-structural and non-functional residues [4–6]. Analog sequences were obtained by replacing corresponding residues in the huIFN-β sequence. All constructed analogs were then analyzed by the NetNGlyc1.0 server to identify potential glycosylation probability prior to 3D structure generation.

2.4. Comparative modeling and structure validation

A comparative modeling technique was applied to construct 3D structures of selected sequences. The crystal structure of huIFN-β (chain A) was used as a template and the MODELLER 9v10 software was utilized for creating the 3D structures. The input to the program was the alignment file of the target-template amino acid sequences. Sequence alignments were obtained from the align2d command in MODELLER, and used as input data in the MODELLER 9v10 software. Once a target-template alignment was constructed, MODELLER calculated 3D models of the target automatically. For each target sequence, 10,000 molecules were generated, and the molecule corresponding to the lowest value of the Probability Density Function (PDF) was selected as the best model and used as an input for the MD simulations. For internal evaluation and energy analysis, DOPE score profiles of the models were generated by MODELLER [23]. Evaluation methods checked whether a model satisfied standard steric and geometric criteria. External evaluation of the protein structures was performed by using the Structural Analysis and Verification Server (SAVES). Several programs from the SAVES (PROCHECK [24], WHAT_CHECK [25], ERRAT [26] and VERIFY_3D [27]) were used to analyze the stereochemical parameters and the quality of the models. The RMSD value indicates the degree to which two 3D structures are similar. The RMSD value between the template and each model was measured by the Swiss-PDB Viewer program.

2.5. Molecular dynamics simulation

Dynamic behavior of the generated models was investigated by MD simulation, using the GROMACS 4.5.3 package on RedHat Linux with the OPLS-AA force field. In order to prepare the topology from the PDB file, the pdb2gmx tool from the GROMACS software package was used. The protein was solvated with the Simple Point Charge (SPC) water molecule [28] in a cubic box type using the genbox program of the GROMACS software package. The minimum distance between any atom in the protein and the box walls was set to 1.0 nm. Different numbers of Cl^- ions were added to neutralize the system based on the net charge of each protein. In order to improve the steric conflict between the protein and water molecules, the solvated system was subjected to energy minimization using the steepest descent integrator with a maximum force smaller than 1000 kJ mol^{-1} nm^{-1} and maximum number of 50,000 steps. Minimization was followed by equilibration dynamics to bring the system to the desired temperature and pressure. Equilibration was performed by applying position-restrained MD on the energy-minimized protein under NPT conditions, keeping the number of particles (N), the system pressure (P) and the temperature (T) constant. This was carried out for 100,000 steps for a total of 200 ps. The reference temperature for coupling was set at 300 K *via* the Berendsen temperature coupling [29], and pressure of one atmosphere (atm) was maintained by the Parrinello–Rahman algorithm [30]. After equilibrating the system to the desired temperature and pressure, the unrestrained MD simulations were continued under NPT conditions for 15–30 ns. The Particles Mesh Ewald (PME) method was used to treat long-range electrostatic interactions [31]. A cutoff of 1.0 nm, both for the Coulomb and van der Waals interactions and the neighborlist generation was employed, and the list was updated every five steps. All simulations were run with periodic boundary conditions and a time step of 2 fs. Atomic coordinates were recorded every 1 ps for the trajectory analysis during MD simulation. Trajectory analysis was carried out using the relevant tools in GROMACS 4.5.3. The RMSD values over the course of the simulation were calculated using g_rms. The radius of gyration (R_{gyr}) during simulation was calculated by g_gyrate. The g_rmsf tool was used for computing the RMSF (Root Mean Square Fluctuation) values during simulation. The program g_cluster (GROMACS tool) was used to determine the conformations that best represent the structures of the entire trajectory obtained during the simulation. The Gromos method was employed to perform a cluster analysis, and RMSD cut-off for two structures to be neighbor was set to 0.1 nm. The central structure of the most dominant cluster for each of the simulations was saved for the subsequent studies.

Table 1
Potential analogs of huIFN-β with one additional N-glycosylation site. The probability of glycosylation was measured by the NetNGlyc1.0 server. The functional and structural residues were determined with respect to the crystal structure of huIFN-β and literatures review. The selected analogs used for subsequent analyses are indicated in bold.

Analog	Potential glycosylation site	Position of mutation(s) in the huIFN-β structure	Glycosylation probability (%)	Likelihood of receptor binding and/or conformational changes	Identification method
L6T	N4	A Helix	79	NO	1
Q10N	N10	A Helix	77	YES – Conformational changes	1
R11N	N11	A Helix	64	NO	1
Q16T	N14	A Helix	70	YES – Receptor binding	1
R27T	N25	AB1 Loop	74	YES – Receptor binding	1 & 3
D39T	N37	AB2 Loop	71	YES – Conformational changes/Receptor binding	1
A56N	N56	B Helix	67	NO	1
F67T	N65	B Helix	76	YES – Receptor binding	1
Q72N	N72	BC Loop	54	YES – Receptor binding	1
D73N	N73	BC Loop	69	YES – Receptor binding	1
S74N	N74	BC Loop	38	NO	1
S75N	N75	BC Loop	54	NO	1
L88T	N86	C Helix	70	YES – Receptor binding	1
Y92T	N90	C Helix	71	YES – Receptor binding	1
L98T	N96	C Helix	66	NO	1
L98N	N98	C Helix	76	NO	1
D110N	N110	CD Loop	51	NO	1 & 3
L116N	N116	CD Loop	47	NO	1
M117N	N117	CD Loop	66	NO	1
E137N	N137	DE Loop	54	YES – Receptor binding	1 & 3
A142N	N142	E Helix	64	YES – Conformational changes	1
Y155T	N153	E Helix	57	YES – Conformational changes/Receptor binding	1 & 3
L160T	N158	E Helix	33	NO	1
R159N	N159	E Helix	65	NO	1
Q16N-Q18T	N16	A Helix	74	YES – Receptor binding	2
Q18N-L20T	N18	A Helix	76	YES – Receptor binding	2
Q23N-N25T	N23	AB1 Loop	78	YES – Conformational changes/Receptor binding	2
Q46N-Q48T	N46	AB3 Loop	70	NO	2
Q48N-F50T	N48	AB3 Loop	64	NO	2
Q49N-Q51T	N49	AB3 Loop-B Helix	71	NO	2
Q51N-E53T	N51	B Helix	68	NO	2
Q64N-I66T	N64	B Helix	56	YES – Receptor binding	2
Q72N-S74T	N72	BC Loop	61	YES – Receptor binding	2
Q94N-N96T	N94	C Helix	67	YES – Conformational changes/Receptor binding	2
L6N-F8T	N6	A Helix	81	NO	3
E29N-C31T	N29	AB1 Loop	84	YES – Disulfide bond	3
D54N-A56T	N54	B Helix	60	NO	3
D73N-S75T	N73	BC Loop	74	YES – Receptor binding	3
K115N-M117T	N115	CD Loop	75	NO	3
L116N-S118T	N116	CD Loop-D Helix	55	YES – Conformational changes	3

3. Results

3.1. Design and selection of potential hyper-glycosylated analogs

Three methods were used to select positions for insertion of the new consensus tripeptide (Asn-X-Thr/Ser): (1) minimizing sequence changes, (2) minimizing structural alterations and (3) finding potential positions which can be efficiently glycosylated. In the first method, every position that had a Thr or Ser residue located at position +2 was selected as a candidate to be transformed into glycosylation sites through exchange by Asn. Moreover, Asn residues present in the sequence were located and transformed into the glycosylation sites through replacement of the residues at position +2 by Thr. Because tripeptides with Thr in the third position, are glycosylated more efficiently than Ser, Thr mutations were considered first [32]. This method identified 24 potential analogs (Table 1).

Since substitutions with similar amino acids will cause the least amount of changes in protein structure and function [33], for the second method, amino acid residues which were chemically similar to Asn or Ser/Thr were also examined within the consensus tripeptide sequence. This approach identified 10 potential huIFN-β analogs where Gln was replaced by Asn, and the residue at position +2 changed to Thr (Table 1). Ceaglio and colleagues demonstrated that more efficient glycosylation and consequently, full anti-viral and anti-proliferative activities for the glycoengineered IFN-α2 analogs were accomplished when the new N-glycosylation site was aligned to the naturally glycosylated site on huIFN-β [11]. Therefore, natural N-glycosylation sites in the type I interferon family would be assumed potent positions for glycosylation as well as efficiently glycosylated positions on the glycoengineered human IFN-α2. The third method used the alignment strategy, whereby the available mammalian type I interferon sequences (18 species) were downloaded from UniProt database and subjected to Multiple Sequence Alignment. The results, as shown in Fig. 2, indicate that the location of the natural N-glycosylation sites are conserved in the type I interferon family.

Overall, ten conserved positions were identified, and the homologue residues in the huIFN-β molecule were changed to Asn, and those at the +2 positions were changed to Thr residues, so as to create new analogs with one additional N-glycosylation site (Table 1). Perceptibly, four of the ten positions (N25, D110, E137 and N153) had already been proposed by the first method.

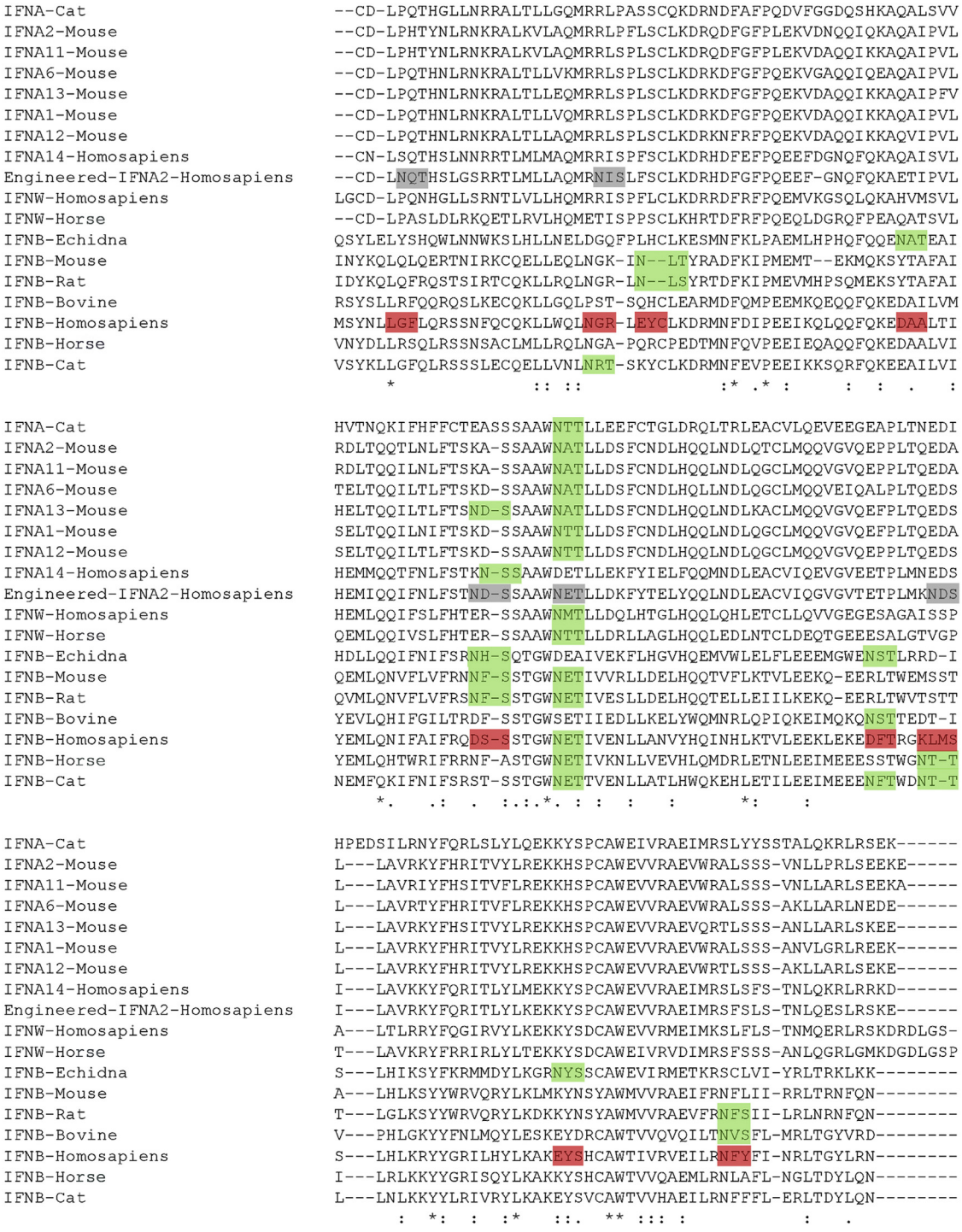

Fig. 2. Multiple Sequence Alignment (MSA) for type I interferon family members. Green boxes are natural N-glycosylation sites that are conserved in the type I interferon family, and can be glycosylated efficiently. Gray boxes indicate efficiently glycosylated sites on the glycoengineered human IFN-α2. Red boxes indicate the corresponding positions on huIFN-β that align to the efficiently N-glycosylated sites.

The glycosylation probability of the 40 mutants was then evaluated by the NetNGlyc1.0 server. The S74N, L116N and L160T analogs that had glycosylation probabilities below 50% were removed from further analyses (Table 1). In order to assure the proposed residue changes would not significantly destroy the receptor binding activity and/or the 3D structure of huIFN-β, a conformational analysis was performed with respect to the crystal structure of huIFN-β and literatures review to determine functional and structural residues [4–6]. In this way, 21mutants in which residue alterations influence structure and/or function of huIFN-β were removed from further analyses (Table 1). Another consideration was the location of Asn or Asn-substituted residues relative to the proline residue. Specifically, proline at position +1 or +3 of the N-glycosylation tripeptide sequence has been known to reduce the glycosylation efficacy. [34]. However, the identified analogs did not have proline at position +1 or +3. Finally, 16 remaining analogs with one additional N-glycosylation site were selected, and their 3D structures were then generated by comparative modeling.

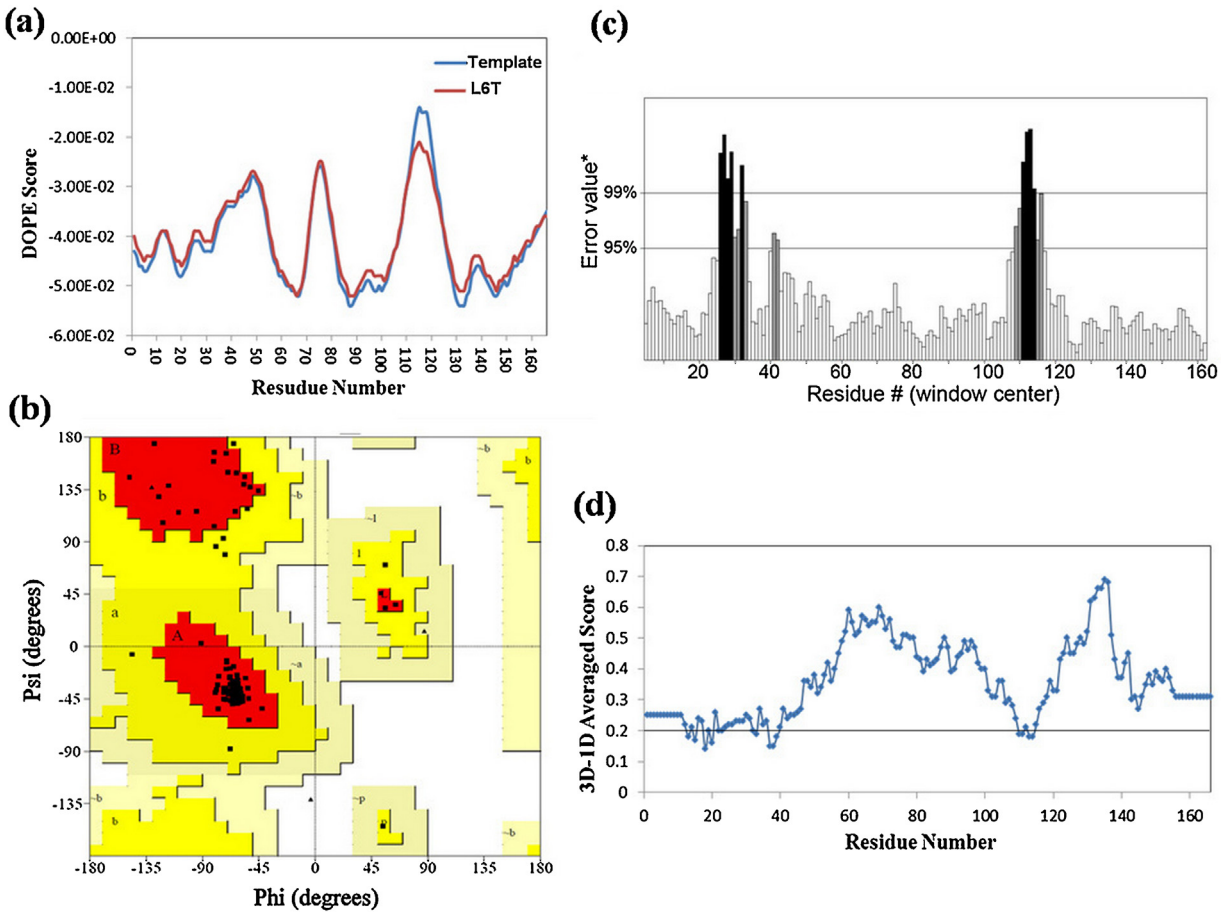

Fig. 3. (a) Comparison of DOPE score profiles of the template and L6T analog did not show any regions of relatively high energy and significant energy differences. (b) The Ramachandran plot calculation of L6T was carried out by PROCHECK, and showed that all the residues are in the core regions of the Ramachandran plot. (c) The quality factor profile of L6T analog was verified by ERRAT, and showed the acceptance of the model. (d) Profile of the side chain environments of the L6T analog was verified by VERIFY_3D. 92.8% of the residues have a compatibility score >0.2, indicating they are reasonably folded.

3.2. 3D structure generation and validation of the potential hyper-glycosylated analogs

3D structures of the analogs and wild-type hulFN-β (control) were generated from the primary sequences. The comparative modeling protocol was used successfully due to the high degree of sequence identity between the template and target sequences. Respective sequence-template structure alignments were performed using the align2D module in the MODELLER software. For each target sequence, 10,000 models were generated and the best model was selected based on PDF values. Energy analysis of models was verified by the DOPE score profiles and presented graphically (Supplement 1). The DOPE score profiles of all models did not show any regions of relatively high energy, and the energy profile of each analog was within the range when compared to the template. An example of this plot for the L6T analog is shown in Fig. 3(a).

Further validation of the generated structures was conducted using the SAVES package. Ramachandran plots were then built by PROCHECK. As shown in Table 2, over 93% of residues for each generated model were located in the most favored regions, and no residue was observed in the disallowed regions. This is satisfactory based on the ψ and ϕ angles, where a good quality Ramachandran plot has over 90% of the residue in the most favored regions [35].

This trend continued in the data obtained by the validation program WHAT_CHECK that calculated Ramachandran *Z*-score as well as bond length and bond angle RMS *Z*-score. As shown in Table 2, all generated models were predicted to have acceptable values where the RMS *Z*-scores should be close to one and the positive value for the Ramachandran *Z*-score is better than average. Side chain environments were assessed by VERIFY_3D for the generated models. The percentage of residues with a 3D-1D score greater than 0.2 should be more than 80% for a reliable structure [36]. The VERIFY_3D results (Table 3) showed that in most generated models, above 80% of the residues had average 3D-1D scores of greater than 0.2, indicating the acceptance of the environment profiles of models. The percentage of residues with 3D-1D scores greater than 0.2 for the four analogs (R11N, Q51N-E53T, L6N-F8T and D54N-A56T) were slightly below the 80% threshold. The residues with 3D-1D scores lower than 0.2 were mainly located at AB loop although in some analogs residues with small compatibility scores were also observed at CD loop and C-terminal of helix A.

ERRAT was used to assess the overall quality factor of the generated models. The ERRAT plots give a measure of the structural error for each amino acid residues in the 3D structure model. These plots showed values relatively close to the 95% cut-off value for all models (Table 3), thus indicating the acceptance of the models. Examples of the validation analyses plots for the L6T analog are presented in Fig. 3(b)–(d).

To measure the average distance between the backbone atoms of the generated models and template, the Cα-based superimposition RMSD values were calculated using the Swiss-PDB Viewer software. The RMSD values are shown in Table 3. These results showed that all the generated models had reasonable low RMSD values when compared to the template, where lower

Table 2
Ramachandran plot and WHAT_CHECK *Z*-scores statistical analyses for quality assessment of the generated models by MODELLER.

Analog	Ramachandran plot			WHAT_CHECK results		
	Most favored regions (%)	Allowed regions (%)	Disallowed regions (%)	Ramachandran plot appearance *Z*-score	Bond length RMS *Z*-score	Bond angle RMS *Z*-score
L6T	94.3	5.7	0.0	2.513	0.905	1.191
R11N	94.3	5.7	0.0	2.372	0.901	1.175
A56N	94.3	5.7	0.0	2.460	0.907	1.18
S75N	94.3	5.7	0.0	2.381	0.907	1.166
L98T	94.3	5.7	0.0	2.378	0.906	1.175
L98N	94.9	5.1	0.0	2.561	0.908	1.191
D110N	93.6	6.4	0.0	2.434	0.906	1.181
M117N	94.3	5.7	0.0	2.277	0.909	1.178
R159N	94.3	5.7	0.0	2.347	0.902	1.200
Q46N-Q48T	93.6	6.4	0.0	2.499	0.911	1.147
Q48N-F50T	93.6	6.4	0.0	2.335	0.908	1.189
Q49N-Q51T	94.3	5.7	0.0	2.639	0.908	1.188
Q51N-E53T	93.6	6.4	0.0	2.292	0.909	1.174
L6N-F8T	93.6	6.4	0.0	2.241	0.908	1.164
D54N-A56T	93.6	6.4	0.0	2.33	0.907	1.189
K115N-M117T	94.3	5.7	0.0	2.503	0.907	1.173
Wild-type IFN-β	94.3	5.7	0.0	2.482	0.908	1.172

values indicated better fits. Such low RMSD values alongside data from protein evaluation imply that the generated model provides reasonable initial coordinates for further MD calculations [37].

3.3. Molecular dynamics simulations of the potential hyper-glycosylated analogs

3.3.1. Molecular stability

MD simulations were performed in order to investigate the effect of amino acid substitutions on the overall stability of the analogs. The initial structures for MD simulations were created by comparative modeling as described above. The OPLS-AA force field was assigned, the proteins were solvated with explicit water, minimized by using the steepest descent integrator, and 15 ns simulations were subsequently performed. The simulations were extended to 30 ns for some analogs to make sure that the equilibration phase has been achieved because the structures remain intact only in that phase and are reliable for subsequent analyses. The equilibrated state was considered achieved when the RMSD over the MD trajectory stabilized in a plateau for at least 10 ns.

One of the most important criteria to analyze the stability of the protein structure is to measure the RMSD during the simulation.

Table 3
VERIFY_3D, ERRAT and whole RMSD calculations relative to the template for quality assessment of the models generated by MODELLER.

Analog	Residues with averaged 3D-1D score >0.2 (%)	Overall quality factor by ERRAT	Whole RMSD (nm)
L6T	92.81	88.608	0.018
R11N	79.04	86.076	0.021
A56N	93.41	87.342	0.020
S75N	85.03	93.671	0.018
L98T	82.63	88.608	0.024
L98N	92.81	87.342	0.020
D110N	91.62	90.506	0.022
M117N	81.44	87.342	0.024
R159N	91.02	86.076	0.019
Q46N-Q48T	97.60	89.241	0.021
Q48N-F50T	87.43	93.038	0.021
Q49N-Q51T	95.21	84.810	0.023
Q51N-E53T	73.65	85.443	0.024
L6N-F8T	79.04	81.646	0.022
D54N-A56T	79.64	86.709	0.025
K115N-M117T	85.63	85.443	0.021
Wild-type IFN-β	89.82	88.608	0.019

The RMSD values of the all atoms were calculated over the trajectory between the initial structure of the MD production phase ($t = 0$) and conformations taken every 1 ps for generated models and then pictured as a function of time (Supplement 2).

Fig. 4(a) shows the RMSD plot for the wild-type IFN-β. The RMSD increases pretty rapidly in the beginning of the simulation, but stabilizes after about 4 ns around 0.27 ± 0.014 nm and remains stable until the end of the simulation. Subsequent analysis of the radius gyration for the wild-type (see below for more details) also showed that the average of R_{gyr} of the protein was found to be 1.65 ± 0.01 nm which reveals that the structure is

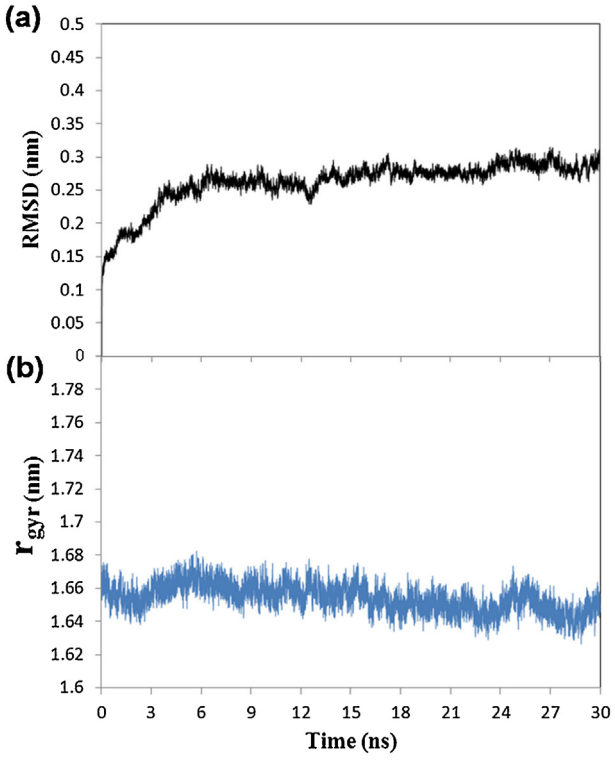

Fig. 4. (a) The RMSD analysis of the wild-type IFN-β over the tragectory showed that the fluctuations equblirated around 0.27 ± 0.014 nm and remained stable until the end of simulation. (b) Analysis of the radius of gyration (Rgyr) for the wild-type IFN-β did not fluctuate much and displayed the average of 1.65 ± 0.01 nm during MD simulation.

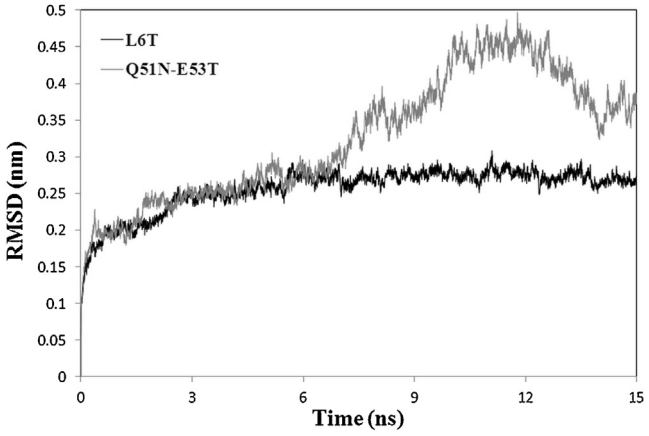

Fig. 5. The RMSD plots for the Q51N-E53T and L6T analogs obtained from MD simulations. The fluctuations of the L6T analog reached plateau and stabilized at approximately 0.27 ± 0.01 nm, while the Q51N-E53T analog did not stabilize and deviated by >0.45 nm.

compact (Fig. 4(b)). The small deviation in both RMSD and R_{gyr} values indicated that the structure keeps stably-folded conformation and revealed the relative stability of the structure for the wild-type IFN-β.

In order to investigate the overall stability of mutated analogs and that the mutation(s) have destabilizing effect on the structures, the RMSDs calculations of the analogs were analyzed with respect to the MD behavior of wild-type system. In this way, the analogs that showed large deviation (>0.3 nm) and/or did not reach to the equilibration during the simulations and increased as time progressed were disqualified.

Table 4 shows the RMSD values from the initial structure during the equilibrated state for each analog. The results showed that the RMSD values for the 11 analogs (L6T, R11N, A56N, S75N, L98N, D110N, M117N, Q46N-Q48T, Q48N-F50T, Q49N-Q51T and L6N-F8T) equilibrated below the 0.3 nm during the simulation and remained stable until the end of the simulations, fluctuating roughly between 0.25 and 0.3 nm. The low values and equilibration of the RMSD as time-dependent functions of the MD simulations indicated that the 11analogues keep stable folding conformations, and that no significant changes caused by the mutation(s) in the structure.

For the five remaining analogs (L98T, R159N, Q51N-E53T, D54N-A56T and K115N-M117T) the RMSDs exceed the 0.3 nm and/or did not reach to the equilibration during the simulations and increased as time progressed. For instance, Fig. 5 illustrates the RMSD plots for an unstable analog (Q51N-E53T) alongside a stable one (L6T) during MD simulation. The fluctuations of the L6T analog reached a plateau at approximately 0.27 nm while the Q51N-E53T analog did not stabilize and deviated at values higher than 0.45 nm.

For analogs R159N, D54N-A56T and K115N-M117T that the RMSD did not equilibrate during the simulations, the structural transitions in the simulations might have caused the protein to enter a highly dynamic path and therefore, some major stabilizing interactions were lost in the protein structure. Accordingly, these analogs were also disqualified and subsequent analyses were performed on the remaining analogs.

The R_{gyr} analysis was also carried out to investigate the permanence of stability for the remaining stable analogs. The R_{gyr} of a protein is a measure of its compactness and is defined as the root mean square distance of the collection of atoms from their common center of gravity. If a protein is stably folded, it will likely maintain a relatively steady value of R_{gyr} [38]. R_{gyr} profiles during the MD production phase are pictured for the stable analogs (see

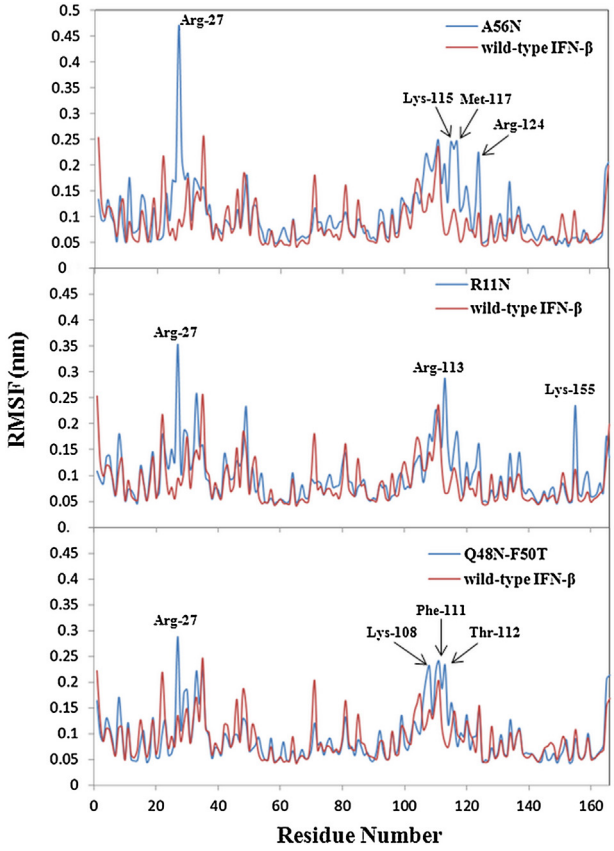

Fig. 6. RMSF calculations for each individual amino acid over the trajectory for R11N, A56N and Q48N-F50T analogs demonstrated regions with elevated values in comparison with the wild-type huIFN-β.

Supplementary data 3). The results of R_{gyr} analysis showed that all the stable analogs display a common R_{gyr} lying roughly between 1.62 and 1.71 nm, at equilibrium without much fluctuation, indicating the stably-folded conformations of the models. An example of this plot for the wild-type IFN-β is shown in Fig. 4(b).

3.3.2. *Flexibility profiles*

The RMSF indicates how much an atom oscillates around its mean position. Certain regions of proteins are more flexible; hence, the RMSF value for each amino acid residue was calculated in order to study the flexibility profile of the stable analogs. A comparison of RMSF profiles based on structure–structure alignments for the wild-type huIFN-β and each individual stable analog was performed to locate the flexible regions. The plots are shown in Supplementary data 4. As expected, the rigid parts of the proteins were located in helices and the flexible parts were mainly located in the loops. There are regions with elevated RMSF values in the A56N, R11N and Q48N-F50T analogs, relative to the wild-type huIFN-β. Fig. 6 shows residues with the main variation for each analog as compared to the wild-type huIFN-β. The residue, Arg-27, demonstrated the highest variation among the three analogs.

Since the major fluctuations for the A56N, R11N and Q48N-F50T analogs were not consistent with the wild-type huIFN-β, they were considered as inappropriate analogs and excluded from the subsequent studies. There were only minor differences in the flexibility profiles for the rest of the stable analogs, when compared to the wild-type huIFN-β. The observed overlapping curves imply that the same flexible regions are located relative to the wild-type huIFN-β, and amino acid changes have not altered flexibility of the regions.

Table 4
Selection of qualified analogs for the glycoengineered huIFNβ. The average and standard deviation of RMSDs over the equilibration phase were calculated to study the overall stability of the mutated analogs. The RMSFs over the trajectory were calculated to study the flexibility profile of the analogs. The average and standard deviation of relative ASA values were calculated to analyze the solvent accessibility of Asn residue at the N-glycosylation site. The whole-structure RMSD was calculated for each stable analog relative to the wild-type huIFN-β to indicate the degree of similarity.

Analog	Average RMSDs over the equilibration phase (nm)	Consistency of flexibility profile	ASA value for Asn residue in new glycosylation site (%)	Whole-structure RMSD relative to the wild-type (nm)
L6T	0.27 ± 0.011	Yes	53 ± 6	0.20
R11N	0.26 ± 0.012	No	N/A	N/A
A56N	0.25 ± 0.022	No	N/A	N/A
S75N	0.25 ± 0.012	Yes	93 ± 8	0.19
L98T	N/A (unstable)	N/A	N/A	N/A
L98N	0.24 ± 0.01	Yes	0	0.17
D110N	0.26 ± 0.011	Yes	91 ± 14	0.16
M117N	0.3 ± 0.013	Yes	86 ± 4.5	0.19
R159N	N/A (unstable)	N/A	N/A	N/A
Q46N-Q48T	0.25 ± 0.013	Yes	80 ± 9	0.15
Q48N-F50T	0.3 ± 0.009	No	N/A	N/A
Q49N-Q51T	0.26 ± 0.006	Yes	84 ± 7	0.18
Q51N-E53T	N/A (unstable)	N/A	N/A	N/A
L6N-F8T	0.26 ± 0.011	Yes	9 ± 7	0.20
D54N-A56T	N/A (unstable)	N/A	N/A	N/A
K115N-M117T	N/A (unstable)	N/A	N/A	N/A

3.4. Selection of qualified stable analogs

Stable analogs were finally screened based on the solvent accessibility of the Asn residue in the new N-glycosylation site and the degree of similarity to the wild-type huIFN-β. The relative solvent accessibility of the Asn residue in the N-glycosylation tripeptide sequence greatly affects folding and the 3D structure of a protein. As glycosylation is a process that occurs prior to folding, the introduced buried sites can be glycosylated, which consequently interfere with proper folding. In order to avoid selecting buried residues, an analysis should be performed on the structures of analogs to determine solvent accessibility of Asn residue in the engineered N-glycosylation site [11]. For this, the PDB files of at least 25 snapshots, corresponding to the last 10–20 ns of the simulations, where the RMSD reached a plateau were generated. The given conformations were then subjected to screening based on surface area accessibility. Table 4 shows the ASA values with variation for Asn residue of new sites in the stable analogs. The engineered Asn residues that had accessibility values less than 50% (L98N and L6N-F8T) were considered buried sites which may interfere with proper folding.

For analysis of similarity, the whole RMSD value between the wild-type huIFN-β and each stable analog was calculated to indicate the degree to which two 3D structures are similar. One particular conformation that represents the set of conformations of the entire trajectory was required, as detected using the g_cluster program. Then, the central structure of the most dominant cluster for each of the simulations was saved and used for the comparative RMSD analysis between the wild-type IFN-β and each analog. As shown in Table 4 all the selected stable analogs showed reasonable low RMSD values relative to the wild-type huIFN-β, so that the RMSDs did not exceed 0.20 nm. This finding suggested that two 3D structures are similar. Finally, based on the results from the MD analysis, ASA% and whole RMSD calculations, six stable analogs (L6T, S75N, D110N, M117N, Q46N-Q48T, and Q49N-Q51T) were identified so that the desired structural properties of the protein are retained despite the addition of the new N-glycosylation site.

4. Discussion

Because of a series of intrinsic physical and chemical instability problems, natural proteins have not evolved for utilization as drugs [10]. HuIFN-β is a therapeutic protein, the application of which is restricted by its low clinical effectiveness [9]. Glycoengineering is a new approach that is currently being applied to improve the physiochemical and pharmacokinetic properties of therapeutics. In this study, a computational methodology was applied to identify suitable new N-glycosylation positions on huIFN-β that minimizes the structural distortion caused by changes in the amino acid sequences. The physics-based energy analysis, MD simulation, was used to investigate the molecular stability of the analogs [39]. The RMSD plots equilibrated with the small deviation and remained stable until the end of simulations for the 11analogues. This suggested that the analogs kept the stable conformations, and a reliable structure was obtained by the end of the simulations. The flexibility profile of the stable analogs was also studied by MD simulation. Such an investigation is important, because variations in the flexibility profile of a protein within important regions can influence the function of the molecule. The RMSF calculations indicated that the Arg-27 residue has the highest variation in some analogs, showing an elevated RMSF (Fig. 6). Mutational analyses have already revealed the important role of Arg-27 for receptor binding, thus, changes in its flexibility profile would abolish the receptor–ligand interaction. Minor variations observed for some analogs, such as Asn-110 in D110N analog, would not be significant as they are in regions that are positioned further away from the interacting sites. In general, the data of this study showed that MD was useful as a means of discriminating the appropriate from inappropriate models through investigating the dynamic behavior of analogs.

To examine whether a time-period used for simulations is sufficient to study the conformational changes, three analogs were selected at random from the stable analogs and the simulations were repeated and extended to 100 ns. The corresponding RMSD plots are presented in Fig. 7. In these three cases, no large structural fluctuations were observed at long time scales. This finding suggests that structural changes induced by mutation(s) can be revealed rapidly and that although longer simulation times are clearly desirable, the time scale used for simulations is sufficient to assess the relative stability of the analogs incorporated in this study.

In order to show the consistency of the simulations and that the differences observed in the behavior of the structures are not simply due to the statistical problems of single MD simulations, the simulations for the three selected analogs were repeated with different starting velocities. Fig. 8 represented a comparison of RMSD from the starting structure over 100 ns for each of the proteins. For each of the structures, similar behavior was observed

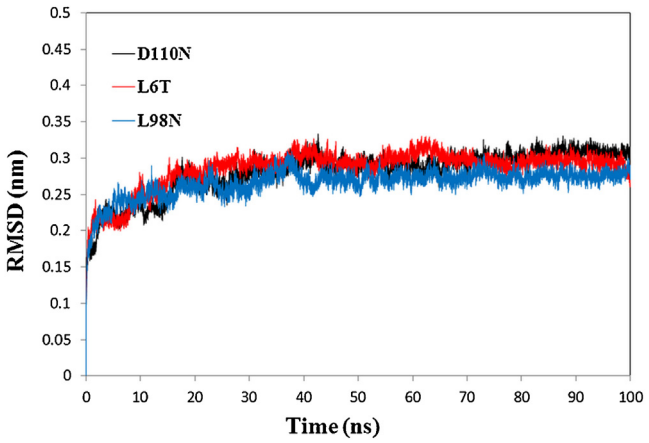

Fig. 7. The RMSD from the starting structure over the trajectory illustrating the effect of extending the simulation time for L6T, L98N and D110N. No large structural fluctuations were observed at extended time.

irrespective of the initial velocities and the choice of initial atom velocities did not result in significant differences in the evolution of the structures in the 100-ns MD simulation. Therefore, it is unlikely that the difference observed in the behavior of structures is simply due to statistical fluctuations.

According to the theoretical analysis of this study, six sites were identified with the highest probability of glycosylation and the least structural distortions caused by the amino acid changes. In all of the qualified analogs, except L6T, the change

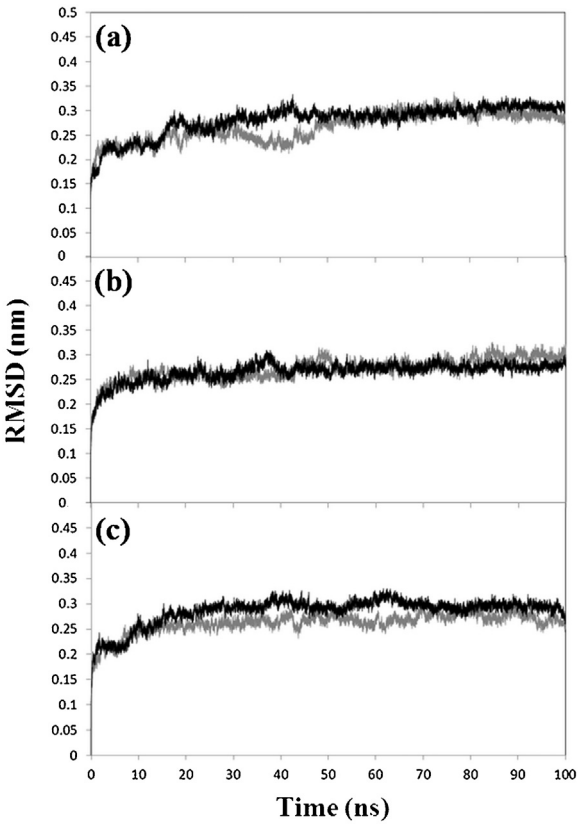

Fig. 8. The RMSD from the starting structure over the trajectory illustrating the sensitivity of the simulations to the changes in the initial velocities. No significant differences were observed for D110N (a), L98N (b) and L6T (c) analogs indicating the consistency of the simulations.

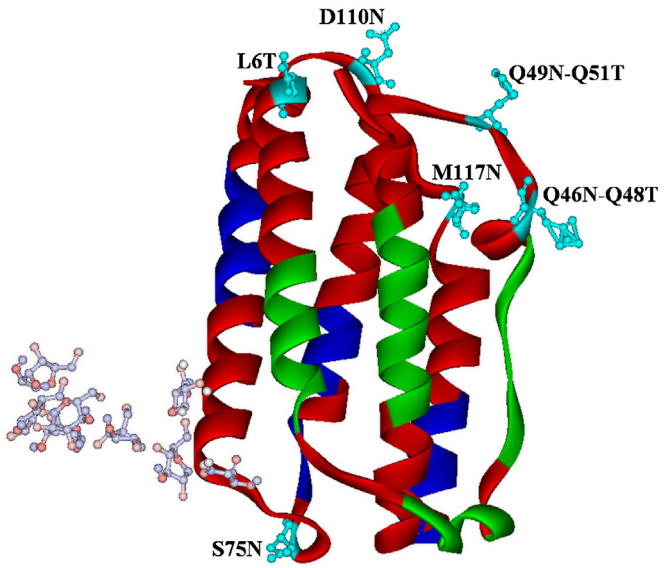

Fig. 9. The location of the new N-glycosylation sites for the qualified analogs, turquoise residues, with respect to the receptor-binding regions on huIFN-β. The blue and green regions correspond to the domains interacting with the IFNAR1 and IFNAR2 subunits of the receptor, respectively. Visualization by Discovery Studio showed that all the new N-glycosylation sites are located away from the receptor binding regions.

in amino acid sequences occurred in the loop regions. By means of the MD simulation, this study also showed that for the unstable analogs, the change in amino acids took place approximately in the helices. Indeed, it was observed that for all the unstable analogs, except for K115N-M117T, mutation(s) occurred in helices. Taken together, these data demonstrated that loops would be favorable regions for introducing a new N-glycosylation tripeptide sequence. This may be due to the fact that structured-regions such as helices are less flexible when compared to disordered-regions like loops and, hence, sequence alterations in such regions cannot be endured. In addition, glycosylation sites in loops will enhance the stability of the protein more effectively than the structured regions [40]. Therefore, five analogs in which the new N-glycosylation tripeptide is located in loops (Q46N-Q48T, Q49N-Q51T, S75N, D110N and M117N) would be preferable candidates due to the stabilizing effect of glycosylation. However, the L6T analog still has the highest probability of glycosylation (Table 1), which represents an important aspect of the glycoengineering strategy. In addition, the existence of a new N-glycosylation site at the N-terminal is beneficial for L6T because it improves the state of post-translational modifications and secretion [41,42].

The N-linked glycan moiety is a negatively charged molecule with a large branched structure [43]. In the huIFN-β molecule, the glycan moiety is not necessary for receptor binding, as the non-glycosylated form of huIFN-β is capable of binding to the receptor. Hence, sugar may cause steric hindrance during the ligand-receptor interaction through contact inhibition and/or charge-based inhibition, if located near the receptor binding domains. The greater the distance between the glycosylation site and the receptor binding regions, the lesser the inhibition imposed on receptor–ligand interaction. By means of visualization analysis, it was evident that the introduced new glycosylation sites for all the qualified analogs lie away from the receptor binding regions (Fig. 9).

In the new 3D crystallographic structure of the type I IFN ligand-receptor complex [44], the orientation of the type I IFN between the two receptor subunits (IFNAR1 and IFNAR2) clearly confirms

that the additional N-glycosylation sites are located in regions with distance from the receptor binding regions. Visualization also showed that the glycosylation sites for L6T, Q46N-Q48T, Q49N-Q51T, D110N and M117N are located at one pole of the molecule, while for S75N it is on the opposite pole, located near the naturally glycosylated Asn-80 residue (Fig. 9). Although close proximity to the natural glycosylation site may reduce the enzymatic glycosylation efficiency in S75N, but being located in a region that is naturally glycosylated would be beneficial and avoid possible contact inhibition by the glycan moiety during the receptor-IFN-β binding process. Additionally, the Asn residue is fully exposed in S75N analog (Table 4). This is highly advantageous, because glycosylation of the exposed positions does not have significant effects on folding and the 3D structure of the protein. Altogether, these data demonstrate that S75N can be regarded as one of the most preferable analogs among the qualified ones.

In order to identify the residues of huIFN-β that are involved in receptor binding, Runkel and colleagues mutated the residues not individually but in groups. Therefore, a higher resolution scan in which the functionally important regions of the molecule probed by single point mutations is needed to give a better picture of the receptor interaction sites [6]. Hence, it may be possible that some other suitable sites for introducing a new N-glycosylation site have been excluded only because of low resolution. To evaluate this assumption, we considered analog D73N for further analyses, in which the proposed position is located in the regions that were mutated in group. The MD analyses showed that D73N had normal behavior during simulation (Supplements 2, 3 and 4). The relative ASA value for the engineered Asn-73 was 65 ± 5% and the whole RMSD relative to the wild-type huIFN-β was 0.24 nm, both of which are acceptable values. However, this analog was excluded because of low resolution applied for identification of the functionally important residues. More effort is therefore needed to identify the interacting residues on the huIFN-β surface involved in the receptor binding, so as to introduce more possible sites by glycoengineering.

In conclusion, the results of this study introduced six qualified analogs. The experimental evaluation of the two qualified analogs (L6T and S75N) were also carried out and confirmed that both analogs were glycosylated at the new sites while retaining their functions, and therefore validated the results of this methodology (unpublished data). This computational approach can be considered as a valuable means of introducing a new N-glycosylation site with the least structural distortion. This is of great importance because construction, expression and characterization of any possible analog can be labor-intensive and time-consuming. For instance, Elliot and colleagues constructed 62 analogs of the human erythropoietin and assessed them experimentally to elicit new hyper-glycosylated analogs [45]. Hence, it seems that such rational evaluation is necessary before designing glycoengineered analogs of therapeutically important proteins.

Acknowledgments

We hereby thank Dr. Parvin Shariati, Dr. Armin Madadkar-Sobhani, Dr. Sarah Harcum and Dr. Mehriar Amininasab for their scientific editing of the work. This work was supported by grant number 346 of National Institute of Genetic Engineering and Biotechnology (NIGEB).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.12.001.

References

[1] J. Bekisz, H. Schmeisser, J. Hernandez, N.D. Goldman, K.C. Zoon, Human interferons alpha, beta and omega, Growth Factors 22 (4) (2004) 243–251, http://dx.doi.org/10.1080/08977190400000833.
[2] R.A. Rudick, S.E. Goelz, Beta-interferon for multiple sclerosis, Exp. Cell Res. 317 (9) (2011) 1301–1311, http://dx.doi.org/10.1016/j.yexcr.2011.03.002.
[3] T. Taniguchi, A. Takaoka, The interferon-alpha/beta system in antiviral responses: a multimodal machinery of gene regulation by the IRF family of transcription factors, Curr. Opin. Immunol. 14 (1) (2002) 111–116.
[4] M. Karpusas, M. Nolte, C.B. Benton, W. Meier, W.N. Lipscomb, S. Goelz, The crystal structure of human interferon beta at 2.2 Å resolution, Proc. Natl. Acad. Sci. U. S. A. 94 (22) (1997) 11813–11818.
[5] M. Karpusas, A. Whitty, L. Runkel, P. Hochman, The structure of human interferon-beta: implications for activity, Cell. Mol. Life Sci. 54 (11) (1998) 1203–1216.
[6] L. Runkel, C. deDios, M. Karpusas, M. Betzenhauser, C. Muldowney, M. Zafari, A. Whitty, Systematic mutational mapping of sites on human interferon-beta-1a that are important for receptor binding and functional activity, Biochemistry 39 (10) (2000) 2538–2551.
[7] L. Runkel, W. Meier, R.B. Pepinsky, M. Karpusas, A. Whitty, K. Kimball, S.E. Goelz, Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-beta (IFN-beta), Pharm. Res. 15 (4) (1998) 641–649.
[8] P.S. Sorensen, C. Ross, K.M. Clemmesen, K. Bendtzen, J.L. Frederiksen, K. Jensen, Danish Multiple Sclerosis Study Group, Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis, Lancet 362 (9391) (2003) 1184–1191, http://dx.doi.org/10.1016/S0140-6736(03)14541-2.
[9] P.S. Sorensen, Neutralizing antibodies against interferon-beta, Ther. Adv. Neurol. Disord. 1 (2) (2008) 125–141, http://dx.doi.org/10.1177/1756285608095144.
[10] R.J. Sola, K. Griebenow, Effects of glycosylation on the stability of protein pharmaceuticals, J. Pharm. Sci. 98 (4) (2009) 1223–1245, http://dx.doi.org/10.1002/jps.21504.
[11] N. Ceaglio, M. Etcheverrigaray, R. Kratje, M. Oggero, Novel long-lasting interferon alpha derivatives designed by glycoengineering, Biochimie 90 (3) (2008) 437–449, http://dx.doi.org/10.1016/j.biochi.2007.10.013.
[12] A.M. Sinclair, S. Elliott, Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins, J. Pharm. Sci. 94 (8) (2005) 1626–1635, http://dx.doi.org/10.1002/jps.20319.
[13] J. Rodriguez, M. Spearman, T. Tharmalingam, K. Sunley, C. Lodewyks, N. Huzel, M. Butler, High productivity of human recombinant beta-interferon from a low-temperature perfusion culture, J. Biotechnol. 150 (4) (2010) 509–518, http://dx.doi.org/10.1016/j.jbiotec.2010.09.959.
[14] R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function, Pac. Symp. Biocomput. 7 (2002) 310–322.
[15] S. Ahmad, M. Gromiha, H. Fawareh, A. Sarai, ASAView: database and tool for solvent accessibility representation in proteins, BMC Bioinformatics 5 (2004) 51, http://dx.doi.org/10.1186/1471-2105-5-51.
[16] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539, http://dx.doi.org/10.1038/msb.2011.75.
[17] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, A. Sali, Comparative protein structure modeling using Modeller, Curr. Protoc. Protein Sci. Chapter 5 (2006), http://dx.doi.org/10.1002/0471250953.bi0506s15, Unit 5.6.
[18] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1) (1995) 43–56.
[19] W.L. Jorgenson, J. Tirado-Rives, The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, J. Am. Chem. Soc. 110 (6) (1988) 1657–1666, http://dx.doi.org/10.1021/ja00214a001.
[20] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PDB Viewer: an environment for comparative protein modeling, Electrophoresis 18 (15) (1997) 2714–2723, http://dx.doi.org/10.1002/elps.1150181505.
[21] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1) (1996) 33–38, 27-38.
[22] Accelrys Software Inc., Discovery Studio Modeling Environment, Release 3.5, Accelrys Software Inc., San Diego, 2012.
[23] M.Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci. 15 (11) (2006) 2507–2524.
[24] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (2) (1993) 283–291, http://dx.doi.org/10.1107/S0021889892009944.
[25] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (6580) (1996) 272.
[26] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (9) (1993) 1511–1519, http://dx.doi.org/10.1002/pro.5560020916.
[27] R. Liithy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (6364) (1992) 83–85.
[28] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to proteins hydration, in: B.E. Pullman (Ed.),

Intermolecular Forces, Reidel Publishing Company, Dordrecht, The Netherlands, 1981, pp. 331–342.
[29] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690, http://dx.doi.org/10.1063/1.448118.
[30] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190, http://dx.doi.org/10.1063/1.328693.
[31] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (19) (1995) 8577–8593.
[32] J. Jones, S.S. Krag, M.J. Betenbaugh, Controlling N-linked glycan site occupancy, Biochim. Biophys. Acta 1726 (2) (2005) 121–137, http://dx.doi.org/10.1016/j.bbagen.2005.07.003.
[33] S. Henikoff, J.G. Henikoff, Amino acid substitution matrices from protein blocks, Proc. Natl. Acad. Sci. U. S. A. 89 (22) (1992) 10915–10919.
[34] S.H. Shakin-Eshleman, S.L. Spitalnik, L. Kasturi, The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency, J. Biol. Chem. 271 (11) (1996) 6363–6366.
[35] J. Xiao, Z. Li, M. Sun, Y. Zhang, C. Sun, Homology modeling and molecular dynamics study of GSK3/SHAGGY-like kinase, Comput. Biol. Chem. 28 (2004) 10.
[36] M. Li, B. Wang, Homology modeling and examination of the effect of the D92E mutation on the H5N1 nonstructural protein NS1 effector domain, J. Mol. Model. 13 (12) (2007) 1237–1244, http://dx.doi.org/10.1007/s00894-007-0245-0.
[37] L. Zhang, H. Zhu, Q. Wang, H. Fang, W. Xu, M. Li, Homology modeling, molecular dynamic simulation and docking studies of cyclin dependent kinase 1, J. Mol. Model. 17 (2) (2011) 219–226, http://dx.doi.org/10.1007/s00894-010-0710-z.
[38] D.P. Homem, R. Flores, P. Tosqui, T. de Castro Rozada, E.A. Basso, A.G. Junior, F.A.V. Seixas, Homology modeling of dihydrofolate reductase from *T. gondii* bonded to antagonists: molecular docking and molecular dynamics simulations, Mol. Biosyst. 9 (6) (2013) 1308–1315, http://dx.doi.org/10.1039/c3mb25530a.
[39] H. Fan, A.E. Mark, Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study, Proteins 53 (1) (2003) 111–120, http://dx.doi.org/10.1002/prot.10496.
[40] D. Shental-Bechor, Y. Levy, Folding of glycoproteins: toward understanding the biophysics of the glycosylation code, Curr. Opin. Struct. Biol. 19 (5) (2009) 524–533, http://dx.doi.org/10.1016/j.sbi.2009.07.002.
[41] Y. Liu, A. Nguyen, R.L. Wolfert, S. Zhuo, Enhancing the secretion of recombinant proteins by engineering N-glycosylation sites, Biotechnol. Progr. 25 (5) (2009) 1468–1475, http://dx.doi.org/10.1002/btpr.241.
[42] C.M. Sagt, B. Kleizen, R. Verwaal, M.D. de Jong, W.H. Muller, A. Smits, C.T. Verrips, Introduction of an N-glycosylation site increases secretion of heterologous proteins in yeasts, Appl. Environ. Microbiol. 66 (11) (2000) 4940–4944.
[43] P. Hossler, S.F. Khattak, Z.J. Li, Optimal and consistent protein glycosylation in mammalian cell culture, Glycobiology 19 (9) (2009) 936–949, http://dx.doi.org/10.1093/glycob/cwp079.
[44] C. Thomas, I. Moraga, D. Levin, P.O. Krutzik, Y. Podoplelova, A. Trejo, C. Lee, G. Yarden, S.E. Vleck, J.S. Glenn, G.P. Nolan, J. Piehler, G. Schreiber, K.C. Garcia, Structural linkage between ligand discrimination and receptor activation by type I interferons, Cell 146 (4) (2011) 621–632, http://dx.doi.org/10.1016/j.cell.2011.06.048.
[45] S. Elliott, T. Lorenzini, S. Asher, K. Aoki, D. Brankow, L. Buck, J. Egrie, Enhancement of therapeutic protein in vivo activities through glycoengineering, Nat. Biotechnol. 21 (4) (2003) 414–421, http://dx.doi.org/10.1038/nbt799.