Research Article

doi:10.34172/ajpr.2020.04

Computational Design for the Production of Secretory Recombinant Codon-Optimized Human Stem Cell Factor (SCF) in Chinese Hamster Ovary (CHO) Cells With an Appropriate Signal Peptide: An Intensive In Silico Study

Ehsan Heidari Soureshjani[1,2*], Maryam Peymani[3*], Kamran Ghaedi[4, 5]

[1]Blood Transfusion Research Center, High Institute for Research and Education in Transfusion Medicine, Tehran, Iran.
[2]Shahrekord Regional Blood Transfusion Center, Shahrekord, Iran.
[3]Department of Biology, Faculty of Basic Sciences, Islamic Azad University, Shahrekord Branch, Shahrekord, Iran.
[4]Department of Biology, Faculty of Sciences, University of Isfahan, Isfahan, Iran.
[5]Department of Cellular Biotechnology, Cell Science Research Center, ACECR, Royan Institute for Biotechnology, Isfahan, Iran.

*Corresponding authors:
Ehsan Heidari Soureshjani, MSc; SaNa Zist Pardaz Co, Member of Chahar Mahal and Bakhtiari Science and Technology Park, Shahrekord, Iran. Tel: +98-38-3226-3854. Email: info@sanazist. ir, Maryam Peymani, PhD; Assistant Professor of Genetics, Department of Biology, Faculty of Basic Sciences, Islamic Azad University, Shahrekord Branch, Shahrekord, Iran. Tel: +98-38-3336-1010

Abstract

Background: Production of the recombinant proteins in mammalian cells is an important issue with a bio-therapeutic purpose. Numerous efforts have been focused on the improvement of the yields of recombinant proteins, which include optimization of conventional biological processes, selection of appropriate signal peptides, codon optimization, and re-engineering of cells to produce more proteins. Stem cell factor (SCF) is a blood cytokine which activates the c-Kit receptor. This factor is crucial not only for the differentiation of hematopoietic progenitor cells but also for the survival, proliferation, and differentiation of mast cells. Recently, its therapeutic role in several diseases such as Alzheimer's and myocardial infarction has been investigated. Therefore, the aim of this study was to design a secretory recombinant human SCF with the maximal yield in an appropriate mammalian host cell as Chinese hamster ovary (CHO) cells using the computational studies.

Methods: As the first step, computational simulation studies were carried out to design the appropriate signal peptide for the human SCF protein. Codon optimized coding sequence of hSCF was transferred into a eukaryotic expression vector (pBudCE4.1). Recombinant vector (pBudCE4.1/SCF) was transfected into CHO cells and the stably transformed cells were screened and isolated. Subsequently, the expression of SCF was defined by reverse transcription polymerase chain reaction (RT-qPCR) in stably transformed cells.

Results: Our bioinformatics studies indicated that Azurocidin signal peptide could be a suitable signal peptide for the production of SCF proteins in the CHO cells. Accordingly, computational studies revealed that the presence of 6×His-tag did not have a significant impact on the three-dimensional structure of the protein. Furthermore, the expression of *hSCF* was significant in the stable CHO cells.

Conclusion: The use of this approach may, therefore, lead to the production of highly efficient recombinant *hSCF,* which would be feasible for the mass production of this factor for therapeutic purposes.

Keywords: Chinese hamster ovary cells, Codon optimization, Signal peptide, Stem cell factor

Introduction

Nowadays, the expression and production of recombinant proteins in the mammalian cells are widely used in medical fields for numerous purposes (1,2). In addition, a variety of investigations have been conducted to improve the production of recombinant proteins through the optimization of conventional biological processes and re-engineering of the host cells with better post-translational modifications (3,4). Notably, there is a growth in the use of more complicated eukaryotic cells as optimal host cells (5,6). Meanwhile, selecting an appropriate signal peptide may lead to an increase in protein secretion (7). Most secretory proteins containing signal peptides in the N-terminal region appear during translation in the ribosome exit tunnel and hence, the signal recognition particle (SRP) recognizes the signal peptides. The sequence of signal peptides includes the N-terminal polar region (N-region), the hydrophobic core (H-region), and the C-terminal region (C-region). In most cases, the N-region includes positively and negatively charged amino acid residues (8,9). It has been shown that increasing the positive charge in the N-region

leads to improved protein secretion (10). However, signal peptides with more negative charges could be transferred efficiently. H-region is very important for signal peptides as this part is recognized by SRP to be connected to the area and inhibits protein synthesis and translation, which can be necessary to prevent its release into the cytosol (11). The H-region can facilitate the protein movement in translocon (12). Therefore, any modification in the length and hydrophobicity of the signal peptides would effectively change the efficiency of the protein movement in Translocon (13-16). Interestingly, any mutation within the regions at the downstream of the signal peptide could affect the protein translocation (17).

Stem cell factor (also known as SCF, KIT-ligand, KL, or steel factor) is a cytokine that binds to the c-KIT receptor (CD117). SCF can exist both as a transmembrane protein and a soluble protein. This cytokine plays an important role in proliferation, migration, survival, and differentiation of hematopoietic stem cells (HSCs), melanocytes, and germ cells, functions as a blood cytokine that binds to the c-Kit receptor and triggers a series of biological effects. The recombinant SCF has been produced and used for clinical and research purposes. SCF may be used along with other cytokines to culture HSCs and hematopoietic progenitors. The expansion of these cells ex vivo (outside the body) would allow advances in bone marrow transplantation, in which HSCs are transferred to a patient to re-establish blood formation. One of the problems of injecting SCF for therapeutic purposes is that SCF activates mast cells. The injection of SCF has been shown to cause allergic-like symptoms and the proliferation of mast cells and melanocytes. Cardiomyocyte-specific overexpression of transmembrane SCF promotes stem cell migration and improves cardiac function and animal survival after myocardial infarction (18-20). To ensure the efficient secretion of a recombinant SCF in mammalian cells, selection of an appropriate signal peptide seems to be necessary. Therefore, the purpose of this study was to examine four different signal peptides already introduced by Kober et al (21).

Materials and Methods
Selection of Appropriate Signal Peptides
As the first step, the sequence of the most appropriate signal peptides was selected according to the criteria introduced by Kober et al (21) for the optimization of the secretion of proteins in the CHO cells. Therefore, four distinct signal

peptides were introduced in this study (Table 1).

Structural Modeling and Molecular Dynamics Simulation
Structural modeling studies were carried out by implementing the I-TASSER server (http://zhanglab. ccmb.med.umich.edu/I-TASSER/). As the first step, the SCF protein sequence, including the original sequence of the signal peptide and four distinct aforementioned signal peptides, was recorded on a server called I-TASSER. In the next step, to facilitate the purification of the secretory SCF, we decided to add both 6×His-tag and c-Myc epitopes at the C-terminus of SCF. It should be mentioned that both epitopes were present in the final expression vector. Therefore, the fusion type of SCF-(His)6-Myc was also recorded on this server to assess whether the addition of His-tag could modulate the structure and function of SCF. The accuracy of the data derived from I-TASSER was also checked by Procheck_NT (http://services.mbi. ucla.edu/PROCHECK) and VADAR (Nihserver.mbi.ucla. edu/ERRATV2) programs. Additionally, ERRAT program (http://services.mbi.ucla.edu/ERRAT) was used to check the quality of the models of all structures and compare them with the original SCF. Meanwhile, the signal peptides cleavage site for each construct was predicted through the Signal IP 4.1 server (http://www.cbs.dtu.dk/services/ SignalP) to ensure the proper separation of the signal peptides.

Finally, the molecular dynamics (MD) simulation was performed using Gromacs software, version 4.5.4 (Department of Biophysical Chemistry at Groningen University, Holland) (22,23). Briefly, by considering the n_ aut.gro of each recorded input derived from the I-TASSER server, the size of simulation box for the original SCF was set to 0.8 A° and for the other constructs, it was taken as 1.6 A°. By considering the size of simulation box for each of the constructs, the number of water molecules, Na^+, and Cl^- was calculated accordingly in 10 nanoseconds (ns) for the simulation process (Table 2). It was used for modeling and MD to examine the effect of poly-histidine tag and the Myc epitope on the structure and function of SCF. Therefore, for the first SCF, the original sequence of signal peptides was recorded on the I-TASSER server. MD simulation was performed in the isothermal-isobaric ensemble under periodic boundary conditions. The LINCS algorithm was also used to constrain all bond lengths during the equilibration step and 5-ns free MDs (24). To

Table 1. Selected Signal Peptides According to Kober et al (21)

Name of the Construct	Signal Peptide Sequence	GenBank Accession Number	Original Protein for the Signal Peptide	Host
A2	MDMRAPAGIFGFLLVLFPGYRS	-	Ig kappa light chain precursor (mutant A2)	*Mus musculus*
B	MKWVTFISLLFLFSSAYS	NP_000468	Serum albumin preproprotein	*Homo sapiens*
E	MTRLTVLALLAGLLASSRA	NP_001691	Azurocidin preproprotein	*Homo sapiens*
M	MASSLYSFLLALSIVYIFVAPTHS	Q766C3	Aspartic proteinase nepenthesin-1	*Nepenthes gracilis*

The sequences of selected signal peptides, their accession numbers and the original protein and the host are indicated.

Table 2. Summary of Simulation Input Data

Three-Dimensional Structure of Proteins	No. of Water Molecules	Distance Between the Protein and the Edge of the Box (A°)	Size of the Simulation Box (A°)	Simulation Time (ns)	No. of Na⁺ Ions	No. of Cl⁻ Ions
SCF	8793	2.64×3.74×3.74	0.8	10	44	44
SCF-(His)6-Myc	15943	3.8×4.95×4.95	1.6	10	80	80
A2-SCF-(His)6-Myc	16708	3.8×4.95×4.95	1.6	10	84	84
B- SCF-(His)6-Myc	16009	3.18×4.49×4.49	1.6	10	81	81
E- SCF-(His)6-Myc	15860	3.17×4.49×4.49	1.6	10	80	80
M- SCF-(His)6-Myc	17278	3.26×4.61×4.61	1.6	10	87	87

The size of simulation box for each construct was derived from n_aut.gro. All of the items were calculated according to the GROMACS software. The number of ions was selected according to their cellular concentrations (140 mM).

reduce artifacts in the calculation of stacking interaction, Amber force field was used with a simple water molecule (25,26). The energy minimization step was performed using the steepest descent method to conjugate the gradient methods. It was followed by 100 picoseconds (ps) of the equilibration step, imposing positional restraints on the non-H atoms. The simulation was conducted at a temperature of 300 K. To reach this threshold temperature, each construct was separately used as an input in a temperature bath using the Berendsen coupling method (27). A cutoff of 1.0 nm was selected for the Coulomb interaction and 0.9 nm was adopted for the Lennard-Jones interaction. The time step was 2 femtoseconds (fs), with coordinates stored after every 2 ps MD simulation was performed for 10 ns.

SCF Coding Sequence
According to the data obtained by the simulation studies, E construct was predicted to be the best signal peptide for the secretion of SCF. Hence, a chimeric version of E fused with human SCF coding sequence (CDS) (GenBank No: M59964.1), termed as E-SCF, was chosen and its nucleotides were codon-optimized through Genscript (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) and IDT server (https://eu.idtdna.com/CodonOpt) for production in the CHO cells. The codon-optimized E-SCF in the type of *BamH* I-E-SCF- *Sal* I was ordered from the Generay Co. (China) for further subcloning of E-SCF. Finally, we used the plasmid encoding *BamH* I-E-SCF- *Sal* I, named as pGHn/SCF, for further experiments.

Construction of the Recombinant pBudCE4.1 (+) Encoding E-SCF
The recombinant pGHn/SCF vector was transformed into *E. coli* Top10 competent cells. Plasmid extraction was carried out on bacterial colonies using Plasmid Mini Extraction Kit (BIONEER, Korea). The BamHI-SalI DNA fragment containing E-SCF was digested from pGHn/SCF using respective endonucleases (Fermentas, Lithuania) and inserted into the same place in the pBud CE4.1 vector, resulting in the production of pBudCE4.1/E-SCF.

CHO Cell Culture and Transfection
CHO-K1 cell line was obtained from Royan Institute for Stem Cell Biology and Technology (Tehran, Iran) and cultured in Ham's F12 (Sigma, USA) containing 10% fetal calf serum and 1% penicillin–streptomycin. Cell culture incubation was performed in a humidified atmosphere of 5% CO2 at 37ºC. Approximately, 6.25×10^5 cells were passaged in each well of a six-well plate dish one day prior to the transfection. Transfection of pBudCE4.1/E-SCF was carried out using Lipofectamine LTX reagent (Invitrogen, Carlsbad, CA, USA) and data were compared with those obtained by transfection of pBudCE4.1.

Isolation of a Stable CHO Cell Line Expressing SCF
Two days post-transfection, the cells were passaged and treated with 60 mg/mL Zeocin (Invitrogen, USA) for 4 weeks. The emerged colonies were isolated for further screening by RT-PCR using specific primers (Table 3) to confirm the presence of SCF expression. Stably transformed CHO-K1 cells with pBudCE4.1/E-SCF were nominated as CHO/SCF cells and compared with the Mock cells, the stable transforming CHO cells with pBudCE4.1.

RT-PCR and Quantitative Real Time PCR for Quantification of Gene Expression
Total RNA was extracted from transfected CHO-K1 cells according to the protocol using TRizol reagent (Invitrogen) based on manufacturer instruction. Approximately, 1 µg of total RNA was used to synthesize cDNA by Revert Aid First Strand cDNA synthesis kit which was purchased from Thermo Scientific (USA) utilizing random hexamer primer. The final step of reverse transcription polymerase chain reaction (RT-PCR) was completed using the specifically designed primers for SCF (Table 3). The

Table 3. List of Primers

Genes	Forward Primer (5′–3′)	Reverse Primer (5′–3′)	Annealing Temperature (°C)	Product Size (bp)
SCF	ATCGACAAGCTGGTGAAC	AACTCTTCGGGTGTGAAC	47	119
GAPDH	TGCCGCCTGGAGAAACC	TGAAGTCGCAGGAGACAACC	58	121

primers were ordered from Bioneer Company (Korea). PCR products were electrophoresed on 1 % agarose gel containing ethidium bromide and bands were visualized with UV light (Uvico, Rapp Optoelectronic, Hamburg, Germany). Real-time RT-PCR was carried out using specific primer pairs (Table 3) in a Step One Plus thermal cycler (Applied Biosystems, Carlsbad, CA, USA) as defined by the manufacturer's protocol. All reactions were carried out in triplicate. The expression level of each target gene was normalized by the expression of the housekeeping gene, Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). Final data were calculated based on $2^{-\Delta\Delta Ct}$.

Western Blot Analysis

To confirm the results of real-time PCR technique**,** the expression of SCF at the protein level was also observed using indirect western blot analyses**.** To do the experiment, the supernatant was condensed with acetone/DTT and alcohol.

An equivalent amount of supernatant was condensed and separated by SDS-PAGE gel and transferred on PVDF membrane. The membrane was incubated in Tris-Buffered Saline containing 3% Bovine Serum Albumin and then the strips reacted with anti-His-tag monoclonal antibody (ab18184) for one hour at 37ºC. Then, the strips were incubated with horseradish peroxides conjugated goat anti-mouse IgG (Dako, P0447) for 1 hour at 37ºC. The strips were visualized after color development in diaminobenzidine/H_2O_2 substrate solution for 15 minutes at room temperature.

Statistical Analysis

Microsoft Excel (2007) and SPSS software (version 17.0) were used to analyze the data which were expressed as mean ± standard error of mean (SEM) and obtained from three independent treatments of replicated observations. One-way analysis of variance (ANOVA) was performed to identify statistical differences between treatments, which were considered to be significant at $P<0.05$.

Results

Signal peptides can have a major effect on the secretion of recombinant proteins expressed in mammalian cells and its performance is interchangeable among different species.

The purpose of this research was to select one of the four signal peptides introduced by Kober et al (21) to efficiently increase the secretion of SCF in the CHO cells. To this end, we performed the modeling of six structures and studied the solubility, polarity, and other kinetic characteristics of the structures (Figures 1 and 2). Moreover, the effect of the 6×His-tag and c-Myc epitopes on the protein structure of SCF was investigated. We carried out MD simulation and then the best signal peptide was chosen for the optimal expression of the recombinant SCF. To synthesize the
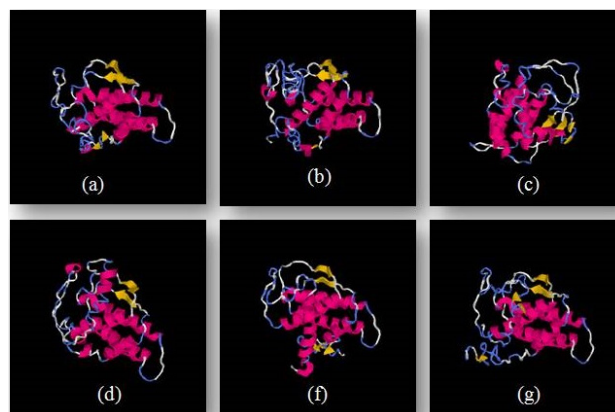
Figure 1. Three-Dimensional Structure of Molecules of SCF (a), SCF-(His)6-Myc (b), A2-SCF-(His)6-Myc (c), B-SCF -(His)6-Myc(d), E-SCF-(His)6-Myc(f), M-SCF-(His)6-Myc (g).
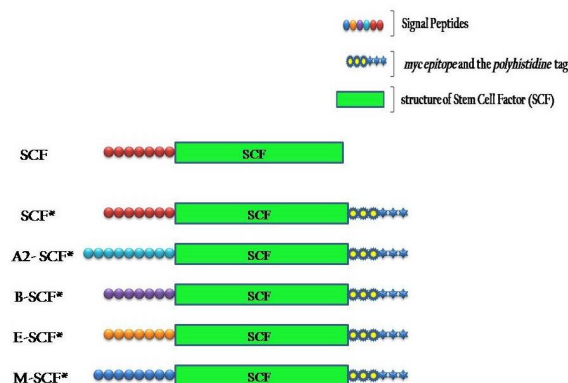
Figure 2. Six Selected Structures in Molecular Dynamics Simulation.
* Poly-histidine tag sequence and myc epitope sequence (EQKLISEEDLNMHTGHHHHHH) were added to the end of SCF protein sequence in three-dimensional structure. A2, B, E and M: Replace the signal peptides (Table 1) with the original signal peptides of SCF in the three-dimensional structure.

chimeric structure of the SCF CDS, we performed codon optimization in order to insert it into a suitable expression vector (pBudCE4.1) (Figure 3).
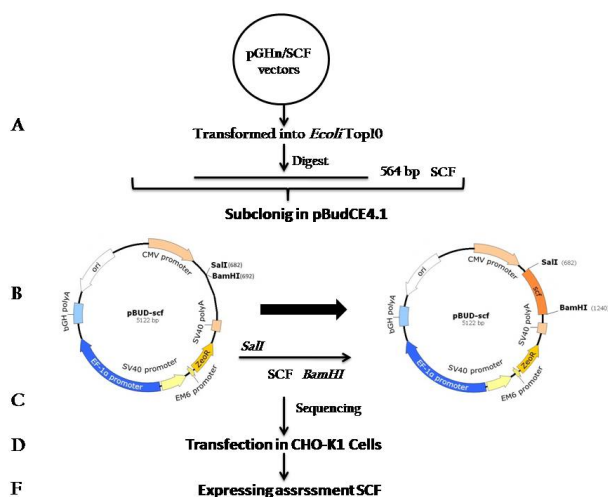
MD Simulation

Figure S1 (see Supplementary file 1) shows the potential energy and the kinetic energy system of six structures in the presence of water and 140 mM of NaCl during 10 ns of the MD simulation (Table 2). The study showed the potential energy of the organization during the simulation in terms of energy balance. As indicated, there were many similarities in terms of the energy level between SCF-(His)6-Myc and E-SCF-(His)6-Myc. The average potential and kinetic energy values during the last 6 ns of the final 10 ns simulation were selected, as shown in Table 4.

The variations in the system temperature (Kelvin) and root-mean-square deviation (RMSD) of the SCF atoms are shown in Figures S2 and 4. The changes in the average of temperatures (Figure S2) showed that the temperature during the simulation system was balanced.

Table 4. The Average Potential Energy, Kinetic Energy, RMSD, RMSF, Rg, Temperature and Number of Hydrogen Bonds of Protein During the Last 6 ns of 10-ns MD Simulation of Protein

Abbreviated Three-Dimensional Structure of Proteins	Potential Energy	Kinetic Energy	Temperature	RMSD	Rg (nm)	RMSF	Number of Hydrogen Bonds
SCF	-418963±663	749943±551	299±2.20	0.40±0.015	1.67±0.008	0.36±0.13	435.8±11.9
SCF*	-743516±897	129263±666	300±1.54	0.46±0.016	1.78±0.013	0.18±0.11	529.8±12.3
A2- SCF*	-777889±976	134566±677	299.9±1.38	0.47±0.011	1.75±0.008	0.18±0.10	514.03±11.27
B-SCF*	-745214±939	129257±665	299.9±1.54	0.45±0.02	1.77±0.011	0.20±0.09	488.14±11.08
E-SCF*	-740427±892	128139±658	300±1.54	0.59±0.009	1.77±0.008	0.23±0.15	509.43±12.34
M-SCF*	-802239±1018	139012±564	300±1.21	0.44±0.032	1.72±0.009	0.16±0.08	509.43±12.34

**Figure 3.** Sub-cloning Into pBudCE4.1/SCF (+) Vector.
A description for stages (A–F) of procedure is detailed in the "Materials and Methods" section.

Moreover, the RMSD results suggested that during 10 ns of simulation, the molecular structures of the studied proteins were at the maximum stability (Table 4). The average RMSD values (Table 4) were 0.40 ± 0.015 and 0.46 ± 0.016 for SCF and SCF-(His)6-Myc, respectively. Therefore, we could conclude that adding a signal peptide E to the SCF protein structure could not change RMSD significantly. As can be seen in Figure 4, Panel g, the structure of M-SCF-(His)6-Myc during simulation was less stable; hence, it could be inferred that this signal peptide caused instability. Figure 5 shows the radius of gyration (Rg) of structures in the presence of water and 140 mM NaCl within 10 ns of MD simulation, under similar conditions expressed above (Table 4). Figure 5 shows the Rg value of M-SCF-(His)6-Myc. As indicated, this structure was less stable when compared with SCF. Therefore, this signal peptide reduced the access level of the protein structure. Among the signal peptides, signal peptides C and E indicated the highest change (1.77) in the Rg, while the structure of E-SCF-(His)6-Myc indicated a minimal change (1.77 ± 0.008).

Since RMSD is not an appropriate parameter to reflect the mobility of the structural elements, we analyzed the root mean square fluctuation (RMSF) (Figure 6 and Table 4) to assess the flexibility of the structures. SCF structure showed little flexibility (Figure 6, panel a). In addition, RMSF analyses (Figure 6, panel f) indicated that the signal peptide E caused the biggest change in the flexibility of SCF (0.23 ± 0.15), which was in a good association with the Rg data (Figure 6, Panel e). The variations in RMSF of M-SCF-(His)6-Myc (Figure 6, Panel g) showed that the signal peptide M could reduce the value of RMSF and cause some flexibility in the structure of SCF.

The average number of hydrogen bonds between six structures and water molecules during the time of simulation was measured to be 435.8±11.9 and 529.8±12.3 for SCF and SCF-(His)6-Myc molecules, respectively. On the other hand, the implementation of signal peptides E or M (Table 1) led to the formation of an equal number of hydrogen bonds between proteins and water
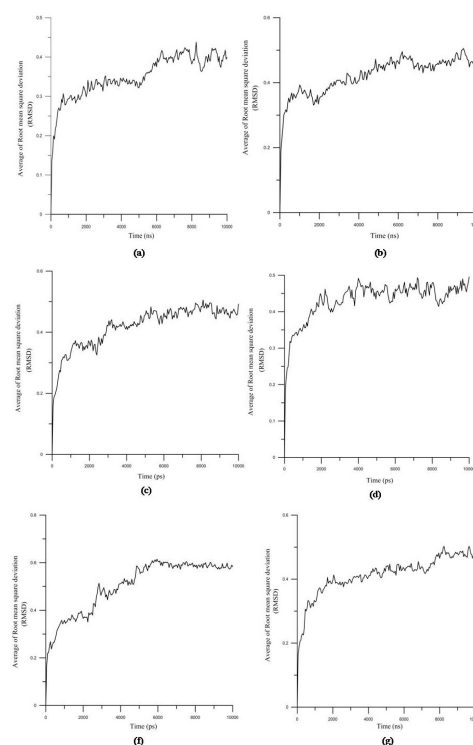
Figure 4. Variations in *Root-mean-square Deviation* of Backbone Atoms of the Molecules of SCF(a), SCF-(His)6-Myc (b), A2-SCF-(His)6-Myc (c), B-SCF-(His)6-Myc (d), E-SCF -(His)6-Myc(f), M-SCF-(His)6-Myc (g).
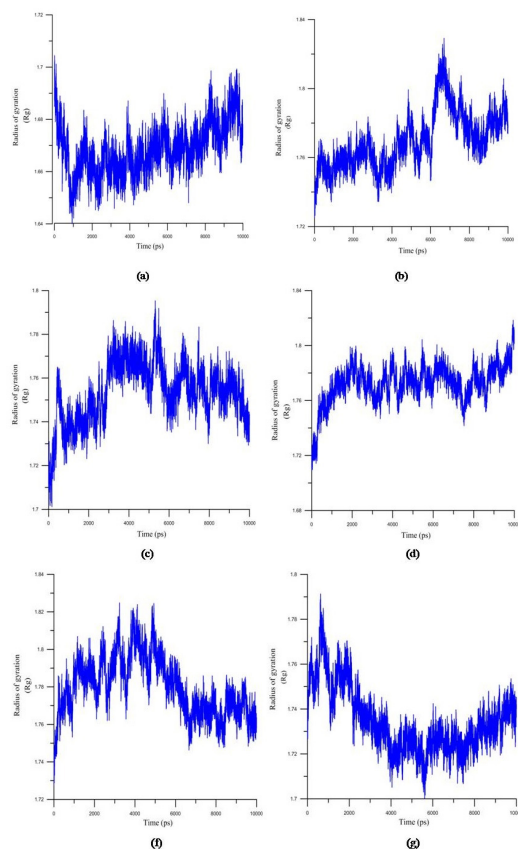
Figure 5. Variations in the Radius of Gyration of Molecules in the Molecular Dynamics Simulation of SCF (a), SCF-(His)6-Myc (b), A2-SCF-(His)6-Myc (c), B-SCF-(His)6-Myc (d), E-SCF-(His)6-Myc (f), M-SCF-(His)6-Myc (g).

(509.43±12.34) (Table 4). The assessment of the changes in the number of hydrogen bonds (Figure 7) between CSF and water showed that SCF dissolved in water. Interestingly, His-SCF had more solubility (Figure 7, Panel b).

(g).

Table 5 shows the kinetic characteristics of the six studied structures in the presence of water molecules during 10-ns MD simulations. The value of electrostatic energy contribution in binding was greater than that of the van der Waals forces. Moreover, the non-polar solvation energy contributed largely to the binding of the SCF molecules to water molecules. Average energy changes in electrostatic binding energy and the binding energy between the atoms of SCF-(His)6-Myc and the molecules of water indicated that there were strong electrostatic forces between these molecules. Interestingly, the levels of variations in the polar solvation energy (-27186.778 kJ/mol), non-polar solvation energy (-461.708 kJ/mol), van der Waals energy (-2327.880 kJ/mol) between water and SCF-(His)6-Myc, and electrostatic interactions (-27186.346 kJ/mol) between SCF-(His)6-Myc and water were greater than those of SCF. The mean changes in the binding energy (Table 5) showed that the electrostatic energy and the binding energy between the atoms and molecules of water and A2-SCF-(His)6-Myc, as compared with the original signal peptide

of SCF, reduced significantly. The van der Waals energy, polar solvation energy, non-polar solvation energy, and the binding energy between water and A2-SCF-(His)6-Myc were equal (-2145.485, 23264.78, -636.629, and -5577.743 kJ/mol, respectively), thereby suggesting that the efficacy of the signal peptide A2 was greater than that of the original signal peptide of SCF in reducing the solubility and polarity of SCF. Meanwhile, the polar and non-polar solvation energy in the presence of the signal peptide B changed considerably, as compared to SCF-(His)6-Myc, showing that this replacement reduced the solubility and polarity of SCF. The assessment of the electrostatic binding energy and the binding energy between the atoms of E-SCF-(His) 6-Myc and the molecules of water indicated that there were strong electrostatic forces between the former and the latter.

In the presence of the signal peptides E, the non-polar solvation energy reached -639.834 kJ/mol, suggesting that the signal peptide E reduced the non-polar solvation energy. Additionally, the change in the electrostatic binding energy and the binding energy between the atoms of M-SCF-(His)6-Myc and the molecules of water indicated that the signal peptide M reduced the amount of the electrostatic force between water and M-SCF-(His)6-My, as compared to the original signal peptide of SCF. The van der Waals energy, polar solvation energy, and the binding energy between water and M-SCF-(His)6-Myc were equal to -2065.76, 2046.5, and -6529.493 kJ/mol,
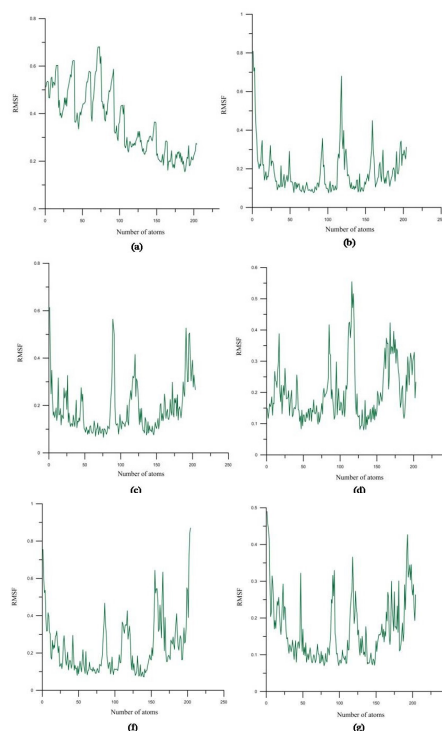
Figure 6. Variations in Root Mean Square Fluctuation of Molecules in the Molecular Dynamics Simulation of SCF (a), SCF-(His)6-Myc (b), A2-SCF-(His)6-Myc (c), B-SCF-(His)6-Myc (d), E-SCF-(His)6-Myc (f), M-SCF-(His)6-Myc (g).

Table 5. Non-polar Solvation Energy, Polar Solvation Energy, Van der Waal Energy, Electrostatic Energy and Binding Energy Between the Atoms of the 6 Studied Structure and Molecules of Water in During 10 ns Dynamics Simulations

Abbreviated Three-Dimensional Structure of Proteins	Non-polar Solvation Energy Between Water Molecules and Protein (kJ/mol)	Polar Solvation Energy Between Water and Protein (kJ/mol)	Van der Waal Energy Between Water and Protein (kJ/mol)	Electrostatic Energy Between Water and Protein (kJ/mol)	Binding Energy Between Water and Protein (kJ/mol)
SCF	-570.688	21011.996	-1841.591	-23292.477	-4692.760
SCF*	-641.708	24264.778	-2327.880	-27186.346	-5891.756
A2- SCF*	-634.629	23264.78	-2145.485	-26238.830	-5577.743
B-SCF*	-639.834	22930.502	-2267.418	-2464.654	-4617.404
E-SCF*	-628.940	23204.242	-2178.110	-26026.685	-6529.493
M-SCF*	-615.433	20406.500	-2065.76	-24902.800	-6529.493
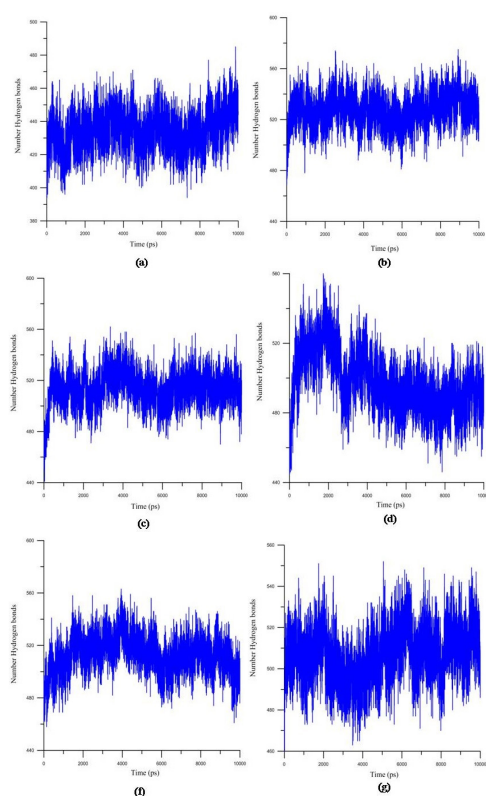
Figure 7. The Number of Hydrogen Bonds Between the Atoms of SCF (a), SCF-(His)6-Myc (b), A2-SCF-(His)6-Myc (c), B-SCF-(His)6-Myc (d), E-SCF-(His)6-Myc (f), M-SCF-(His)6-Myc (g).

respectively, suggesting that replacing the signal peptide M with the original signal peptide of SCF reduced the polar solvation energy.

Ectopic Expression of SCF in CHO Cells

As depicted in Figure 3 and described in the "Materials and Methods" section, after selecting the signal peptide E based on the simulation results and the artificial synthesis of the chimeric gene, the replicating recombinant vector pGHn/SCF was used for further experiments. The recombinant vector was transformed in *E. coli* and plasmid extraction was performed (Figure S3). To ensure that the recombinant plasmid contained the bona fide SCF CDS,

digestion was accomplished by means of *Bam*HI and *Sal*I.

As the next stage, SCF CDS was subcloned into pBudCE4.1 vector under regulation of the CMV promoter, resulting in the generation of pBudCE4.1/SCF (Figure 3, Panel b). Recombinant vector (pBudCE4.1/SCF) was transfected along with the pBud CE4.1 into the CHO-K1 cells (Mock) separately. The transfected cells were further expanded for colony formation under the antibiotic treatment. Numerous resistant colonies were grown and finally, among ten isolated colonies (Mock and SCF), one colony with the highest expression level was selected for further experiments (Data not shown). RT-PCR data showed the expression level of SCF in the pBudCE4.1/SCF cells (Figure 8, Panel a, Lane 4), thereby indicating a basal level of the SCF transcription in the CHO-K1 cells. Quantitative real-time PCR also confirmed the data obtained by RT-PCR (Figure 8, Panel b).

To determine whether SCF protein was successfully produced in the CHO cells, the secretion of the SCF protein was detected by the western blot in the CHO cells media after transfection (Figure 8, Panel c). However, SCF was not detected in the transfected cell (empty vector) media.

Discussion

To achieve the appropriate bioreactor cell lines with a high expression level of the recombinant proteins, we needed an efficient expression system. To our knowledge, the secretion of the protein is mediated by transporting it into the endoplasmic reticulum (28). Accordingly, Zhang et al and Dalton and Barton showed that the secreted protein could be improved by selecting the appropriate signal peptide (7,14). Kober et al also performed a comprehensive study using 16 signal peptides from different species, finding that among the selected signal peptides, albumin and human azurocidin signal peptides were the most appropriate ones feasible for commercial purposes (21).

SCF acts as a survival and growth factor required for the expansion of HSCs, helping to maintain and renew HSCs. Moreover, SCF plays an important role in the proliferation and differentiation of mast cells, melanocytes, and germ cells (29-31). Overexpression of SCF induces the migration of neural stem/progenitor cell through activating the C-kit
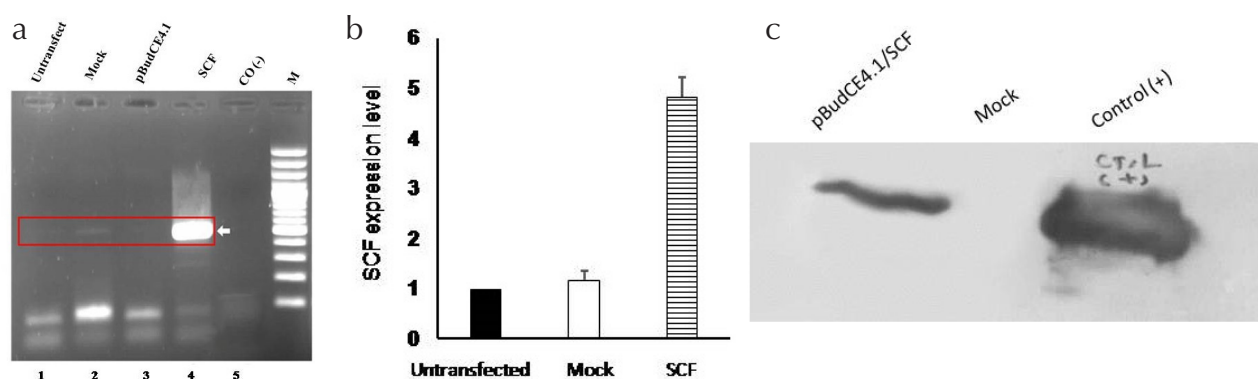
Figure 8. (a) Characterization of stable CHO-K1 cell lines expressing SCF proteins. RT-PCR analysis of CHO-K1 cell lines was used to identify the expression of SCF. (b) M is DNA marker. Lanes 1–5 are referred as untransfected: untransfected cell line, Mock: pBudCE4.1 transformant cell line, SCF: pBudCE4.1-SCF transformant cell line and CO-: no template PCR reaction, respectively. Relative expression levels of SCF of the untransfected cell line (untransfected) and pBudCE4.1 (+) transformant cell line (Mock) and pBudCE4.1- SCF transformant cell line (SCF) were compared. The number of independent repeats was 3 for each experiment (n = 3). Similar alphabets indicate a significant difference between the same samples at $P<0.001$. (c) Western blotting of SCF. SCF protein was detected in CHO transfected cells.

(32).

Additionally, Asghari et al showed the expression of SCF by implementing the vector PET- 266 (+) in *E. coli,* although the yielded protein was not functional (20).

Therefore, the purpose of this research was to design the most suitable signal peptide for the appropriate secretion of the SCF protein in the CHO cell media. We also testified the effects of the His tag on protein structure through MD simulation. Based on the simulation results, we could conclude that signal peptide switching could not influence these parameters and therefore, the least changes in the protein structure occurred; however, it could alter the van der Waals energy levels, electrostatic binding energy, and the polar and non-polar solvation energy between water and SCF, ultimately affecting the solubility and polarity of SCF. On the other hand, by comparing the MD simulation modes of SCF and SCF-(His)6-Myc, we could conclude that adding 6×His tag to SCF resulted in greatest changes in the values of RMSF, RMSD, and Rg. In addition, simulation results suggested that the signal peptide E (azurocidin signal peptide) could be suitably used instead of the original signal peptide of SCF as the binding energy of the signal peptide improved. Our finding was in a good agreement with the data obtained by Kober et al (21). We assumed that this signal peptide could be more accessible to the SRP complex. The experimental results also confirmed our in silico analysis.

Conclusion

MD simulation could serve as a useful tool for the selection of signal peptides and assessment of the function of a protein. In the present study, we testified the human azurocidin signal peptide for the secretion of SCF in the CHO cell media, observing that this signal peptide could act more efficiently than the original signal peptide of SCF.

Conflict of Interests
The authors declare no conflict of interests.

References
1. Francis DM, Page R. Strategies to optimize protein expression in E. coli. Curr Protoc Protein Sci. 2010;61(1):5-24. doi: 10.1002/0471140864.ps0524s61.
2. Zhu J. Mammalian cell protein expression for biopharmaceutical production. Biotechnol Adv. 2012;30(5):1158-70. doi: 10.1016/j.biotechadv.2011.08.022.
3. Andréll J, Tate CG. Overexpression of membrane proteins in mammalian cells for structural studies. Mol Membr Biol. 2013;30(1):52-63. doi: 10.3109/09687688.2012.703703.
4. Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, et al. Overcoming barriers to membrane protein structure determination. Nat Biotechnol. 2011;29(4):335-40. doi: 10.1038/nbt.1833.
5. Xiao S, Shiloach J, Betenbaugh MJ. Engineering cells to improve protein expression. Curr Opin Struct Biol. 2014;26:32-8. doi: 10.1016/j.sbi.2014.03.005.
6. Knappskog S, Ravneberg H, Gjerdrum C, Trösse C, Stern B, Pryme IF. The level of synthesis and secretion of Gaussia princeps luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide. J Biotechnol. 2007;128(4):705-15. doi: 10.1016/j.jbiotec.2006.11.026.
7. Dalton AC, Barton WA. Over-expression of secreted proteins from mammalian cell lines. Protein Sci. 2014;23(5):517-25. doi: 10.1002/pro.2439.
8. von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem. 1983;133(1):17-21. doi: 10.1111/j.1432-1033.1983.tb07424.x.
9. Zheng N, Gierasch LM. Signal sequences: the same yet different. Cell. 1996;86(6):849-52. doi: 10.1016/s0092-8674(00)80159-2.
10. Tsuchiya Y, Morioka K, Shirai J, Yoshida K. Structural requirements of signal peptide in insect cell. Nucleic Acids Symp Ser (Oxf). 2004(48):181-2. doi: 10.1093/nass/48.1.181.
11. von Heijne G. Towards a comparative anatomy of N-terminal topogenic protein sequences. J Mol Biol. 1986;189(1):239-42. doi: 10.1016/0022-2836(86)90394-3.
12. Tsuchiya Y, Morioka K, Shirai J, Yokomizo Y, Yoshida K. Gene design of signal sequence for effective secretion of protein. Nucleic Acids Res Suppl. 2003(3):261-2. doi: 10.1093/

nass/3.1.261.

13. Rapoport TA. Protein translocation across and integration into membranes. CRC Crit Rev Biochem. 1986;20(1):73-137. doi: 10.3109/10409238609115901.

14. Zhang L, Leng Q, Mixson AJ. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J Gene Med. 2005;7(3):354-65. doi: 10.1002/jgm.677.

15. Nilsson I, Johnson AE, von Heijne G. Cleavage of a tail-anchored protein by signal peptidase. FEBS Lett. 2002;516(1-3):106-8. doi: 10.1016/s0014-5793(02)02511-5.

16. von Heijne G. Life and death of a signal peptide. Nature. 1998;396(6707):111-3. doi: 10.1038/24036.

17. Andrews DW, Perara E, Lesser C, Lingappa VR. Sequences beyond the cleavage site influence signal peptide function. J Biol Chem. 1988;263(30):15791-8.

18. Lennartsson J, Rönnstrand L. Stem cell factor receptor/c-Kit: from basic science to clinical implications. Physiol Rev. 2012;92(4):1619-49. doi: 10.1152/physrev.00046.2011.

19. Ropers HH, Craig IW. Report of the committee on the genetic constitution of chromosomes 12 and 13. Cytogenet Cell Genet. 1989;51(1-4):259-79. doi: 10.1159/000132794.

20. Asghari S, Shekari Khaniani M, Darabi M, Mansoori Derakhshan S. Cloning of soluble human stem cell factor in pET-26b(+) vector. Adv Pharm Bull. 2014;4(1):91-5. doi: 10.5681/apb.2014.014.

21. Kober L, Zehe C, Bode J. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol Bioeng. 2013;110(4):1164-73. doi: 10.1002/bit.24776.

22. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29(7):845-54. doi: 10.1093/bioinformatics/btt055.

23. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008;4(3):435-47. doi: 10.1021/ct700301q.

24. Ribeiro AA, Horta BA, de Alencastro RB. MKTOP: a program for automatic construction of molecular topologies. J Braz Chem Soc. 2008;19(7):1433-5. doi: 10.1590/s0103-50532008000700031.

25. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65(3):712-25. doi: 10.1002/prot.21123.

26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926-35. doi: 10.1063/1.445869.

27. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1-3):43-56. doi: 10.1016/0010-4655(95)00042-e.

28. Lai T, Yang Y, Ng SK. Advances in Mammalian cell line development technologies for recombinant protein production. Pharmaceuticals (Basel). 2013;6(5):579-603. doi: 10.3390/ph6050579.

29. Ishikawa K, Fish K, Aguero J, Yaniz-Galende E, Jeong D, Kho C, et al. Stem cell factor gene transfer improves cardiac function after myocardial infarction in swine. Circ Heart Fail. 2015;8(1):167-74. doi: 10.1161/circheartfailure.114.001711.

30. Bernstein ID, Andrews RG, Zsebo KM. Recombinant human stem cell factor enhances the formation of colonies by CD34+ and CD34+lin- cells, and the generation of colony-forming cell progeny from CD34+lin- cells cultured with interleukin-3, granulocyte colony-stimulating factor, or granulocyte-macrophage colony-stimulating factor. Blood. 1991;77(11):2316-21.

31. Migliaccio G, Migliaccio AR, Druzin ML, Giardina PJ, Zsebo KM, Adamson JW. Long-term generation of colony-forming cells in liquid culture of CD34+ cord blood cells in the presence of recombinant human stem cell factor. Blood. 1992;79(10):2620-7.

32. Sun L, Lee J, Fine HA. Neuronally expressed stem cell factor induces neural stem cell migration to areas of brain injury. J Clin Invest. 2004;113(9):1364-74. doi: 10.1172/jci20001.