Computational and nonglycosylated systems: a simpler approach for development of nanosized PEGylated proteins

Hadi Mirzaei¹ Bahram Kazemi¹ Mojgan Bandehpour¹ Alireza Shoari² Vahid Asgary² Mehdi Shafiee Ardestani³ Armin Madakkar-Sobhani⁴ Reza Ahangari Cohan²

¹Department of Biotechnology, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ²Department of Pilot Nanobiotechnology, Pasteur Institute of Iran, Tehran, Iran; ³Department of Radiopharmacy, Pharmacy Faculty, Tehran University of Medical Sciences, Tehran, Iran; ⁴Faculty of Science, University of Ontario Institute of Technology, Oshawa, Canada

Abstract: Cysteine PEGylation includes several steps, and is difficult to manage in practice. In the current investigation, the cysteine PEGylation of erythropoietin analogs was examined using computational and nonglycosylated systems to define a simpler approach for specific PEGylation. Two model analogs (E31C and E89C) were selected for PEGylation based on lowest structural deviation from the native form, accessibility, and nucleophilicity of the free thiol group. The selected analogs were cloned and the expression was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and Western blot using Coomassie blue staining and anti–His monoclonal antibody, respectively. PEGylation with 20 kDa mPEG-maleimide resulted in 79% and 82% conjugation yield for E31C and E89C nonglycosylated erythropoietin (ngEPO) analogs, respectively. The size distribution and charge analysis showed an increase in size and negative charge of the PEGylated forms compared with nonconjugated ones. Biological assay revealed that E31C and E89C mutations and subsequent PEGylation of ngEPO analogs have no deleterious effects on in vitro biological activity when compared to CHO-derived recombinant human erythropoietin. In addition, PEG-conjugated ngEPOs showed a significant increase in plasma half-lives after injection into rats when compared to nonconjugated ones. The development of the cysteine-PEGylated proteins using nonglycosylated expression system and in silico technique can be considered an efficient approach in terms of optimization of PEGylation parameters, time, and cost.

Keywords: site-specific PEGylation, nonglycosylated expression systems, computational methods, erythropoietin

Introduction

Recombinant human erythropoietin (rhEPO) is employed in the treatment of anemia, but like other therapeutic proteins, it suffer from short in vivo half-life. It must be injected 2–3 times/wk for the achievement of therapeutic effect. Erythropoietin (EPO) is structurally a glycoprotein, with 40% of its molecular weight made of glycan (N-glycosylation sites at N24, N38, N83 and O-glycosylation site at S126). Glycosylation plays an important role in maintaining EPO stability and increasing half-life as well as in vivo activity. Since human-like glycosylation only exists in eukaryotic cells, the first-generation of rhEPO was obtained from a CHO cell line transfected with the human erythropoietin gene. rhEPO produced in BHK cells has also been used clinically, in some European Union countries outside Europe and in North America.¹²

Human erythropoietin has been clinically used in the treatment of anemia caused by chronic kidney disease, chemotherapy, and acquired immune deficiency
syndrome-related complications. Due to the pharmaceutical and therapeutic importance of EPO, many investigations have studied its in vivo half-life. PEGylation, one of the most promising tools, performed for in vivo half-life augmentation of EPO, is defined as the attachment of polyethylene glycol (PEG) to the molecule. For successful particular attachment of PEG polymer to a molecule, we need a reactive group on PEG polymer that reacts to a specific site of the molecule. PEGylation is also employed in chemistry for different purposes such as targeting and drug delivery systems, nanotechnology, sustained release of compounds, increasing hydrophilicity, and enhanced in vivo circulation time of therapeutic proteins in the body. PEGylation increases the size of the molecule, leads to lower renal clearance and shielding effect against degradable enzymes, and enhances in vivo half-life. In addition, PEGylation leads to better patient compliance, reduces usage frequency, has lesser side effects, and also decreases medical costs for injecting such drugs.

One of the most important techniques of PEGylation is cysteine PEGylation. In this method, one of the amino acids in the protein, which is not crucial, neither in the activity nor the structure of the protein through genetic engineering, is replaced with cysteine. It is a critical step to define a proper site for the cysteine inside the protein. In the next step, the cysteine analog is expressed and subjected to PEGylation reaction with the PEG polymer that specifically reacts with the free thiol (–SH) group of replaced cysteine residue. As mentioned earlier, cysteine PEGylation includes several steps that are practically difficult to manage and are time-consuming and cost-consuming. Pitfalls include selection of PEG-attached sites (replacement of amino acids with cysteine) and the requirement of large amounts of protein to optimize the PEGylation parameters (eg, size of polymer, time, temperature, pH, and yield). The problem becomes more complicated when the protein is glycosylated. Despite advances in recombinant DNA technology, mammalian expression systems have not produced sufficient amount of protein. In addition, the complexity of attached glycans is not even precisely similar to human, leading to heterogeneity of the finished product. Therefore, to overcome the previously mentioned obstacles, a rational and effortless method for developing of nanosized PEGylated proteins is necessary. Hence, in this study, site-specific PEGylation (cysteine PEGylation) of EPO was examined using computational and nonglycosylated systems to define a simpler approach for site-specific PEGylation of proteins for both therapeutic and nontherapeutic goals.

Materials and methods

In silico analysis
Computational investigation was performed based on a previously described method. Briefly, three-dimensional (3D) structures of analogs were generated by homology modeling using MODELLER and confirmed using discrete optimized protein energy (DOPE) score profile and PROCHECK. The molecular dynamic (MD) simulation of modeled analogs was examined by GROMACS 3.3 package and Gromos96 force field at a constant temperature (300 K) and pressure (1 bar) for 5–15 ns. After calculating the average structures, cysteine analogs were subjected to screening based on the root-mean-square deviation (RMSD) from native EPO, surface area accessibility of engineered cysteine residues (cysteine-SAA), and pKₐ values of the thiol group using Qmol, PROPKA 2.0, and GETAREA, respectively. Finally, two cysteine analogs, E31C and E89C, which had better criteria were selected for experimental investigation.

Cloning of E31C and E89C ngEPO analogs
E31C and E89C analogs were selected for PEGylation on the basis of the results of the computational study. Cysteine analog 31 (E31C) cDNA was previously synthesized, using overlap polymerase chain reaction method, and cysteine analog 89 (E89C) cDNA was ordered to be synthesized by Bioneer Inc. (Daejeon, South Korea). Nde I and Hind III restriction sites and His-tag sequence were introduced using 5′-AAGCTTCAATGATGATGATGATGGT and 5′-ACTCATATGGCCCCACCACGCCTCATCTGTGA as forward and reverse primers.

Expression, identification, and purification of E31C and E89C ngEPO analogs
The constructs were transferred into the BL21 (DE3) Escherichia coli host cell (Gene Bank) using CaCl₂ transformation method. The expression of analogs was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich, Munich, Germany) and analyzed using 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using Mini-PROTEIN® Tetra Cell system (Bio-Rad Laboratories, Hercules, CA, USA); the gel was developed using the Coomassie blue staining technique. The identification of expressed analogs was carried out by Western
PEG-conjugation and purification of PEGylated E31C and E89C ngEO analogs

Thirtyfold molar excess of 20 kDa methoxypoly(ethylene glycol) (mPEG)-maleimide (SunBio Inc.) was added to each analog solution (0.5 mg/mL) in 50 mM Tris buffer pH 8.0, incubated for 3 hours at room temperature. The reaction mixture was then diluted with 20 mM phosphate buffer (pH 7.0) and loaded onto a 6 mL pre-equilibrated cation exchange Resource S column (GE Healthcare, Piscataway, NJ, USA). After washing with 20 mM phosphate buffer (pH 7.0), bonded proteins were eluted using a gradient 20 mM phosphate buffer containing 500 mM NaCl adjusted to pH 7.0. Finally, PEGylated forms were analyzed by SDS-PAGE, Coomassie blue staining, using Mini-PROTEIN® Tetra Cell system (Bio-Rad Laboratories).

Characterization of PEGylated E31C and E89C ngEO analogs

The purity and yield of the conjugates were also determined using size-exclusion high-performance liquid chromatography (using a Shodex column PROTEIN KW-802.5, Showa Denko, Kanagawa, Japan). A solution of 50 mM sodium phosphate buffer (pH 7.0) containing 150 mM NaCl was used as an equilibration and elution buffer. The biological activity was assessed based on the proliferation of UT-7 cells. UT-7 cells (ACC 137; DMSZ, Braunschweig, Germany) were grown in α-minimum essential medium supplemented with 20% fetal bovine serum, 40 µg/mL gentamycin, 2 mM glutamine, and 5 ng/mL granulocyte-macrophage colony-stimulating factor (Sigma-Aldrich). UT-7 cells (5×10^4) were seeded in each well of a 96-well plate and incubated at 37°C in a humidified 5% CO₂ incubator. Twofold serial dilutions of the protein samples and standard were prepared in assay media. Approximately 100 µL of the diluted protein samples was added in triplicate in each well of a 96-well microplate. Phosphate-buffered saline and 200 ng/mL rhEPO (R&D Systems, Minneapolis, MN, USA) were added to the cultured cells as a negative and a positive control, respectively. The proliferation of cells was measured by MTT assay at 570 nm, using Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). In addition, size distribution and the charge of macromolecules were assessed using dynamic light scattering technique (Zetasizer Nano ZS, Malvern Instruments, Malvern, UK).

Pharmacokinetic investigation

Animal experiments were carried out once ethical clearance was obtained from the Animal Ethics Committee of the Pasteur Institute of Iran. All animals were housed at room temperature (20–23°C) on a 12 h light/dark cycle, with unlimited access to food and water and adapted for one week before the test. Animal experiments were performed based on the ethical standards in the Declaration of Helsinki. The five Sprague Dawley rats (300±30 g of weight) in each group subcutaneously received a single dose of 100 µg protein/kg. The blood samples were collected from the lateral tail vein, and the plasma was analyzed by enzyme-linked immunosorbent assay method (Roche Diagnostics GmbH, Penzberg, Germany). The terminal half-life and the area under the curve were determined using linear regression of the last three plasma concentrations and a linear trapezoidal method with extrapolation to infinity, respectively.

Molecular drawing and statistical analysis

Statistical analysis was made using Student’s t-test and one-way multivariate analysis of variance to compute the level of significance (P-value), and P<0.05 was considered statistically significant. The schematic representation of cysteine analogs was identified using VMD software version 1.8.7, and Microsoft Excel 2007 was used to create the RMSD and DOPE graphs.

Results

Selection of cysteine analogs

The 3D structures of 115 selected analogs based on literature review (Supplementary materials 1 and 2) were generated
using homology modeling, and the quality of models was verified by Ramachandran plot and DOPE score assessments (Supplementary materials 3–5). RMSD during MD was plotted versus time for stability and average structure calculations (Supplementary material 5). Frothy stable cysteine analogs during MD simulations were then subjected to screening based on the SAA of replaced cysteine residue (Supplementary material 6), pKᵈ values of the thiol group, and whole RMSD versus native EPO (Supplementary material 7). Finally, based on the minimum whole RMSD and maximum cysteine-SSA and pKᵈ, the E31C and E89C analogs were chosen for experimental analysis (Table 1, Figure 1).

### Expression, identification, and purification of E31C and E89C ngEPO analogs

The expression was found to be optimum at 4 hours after induction with 1 mM IPTG. The expression was confirmed by observing a band approximately 18.4 kDa on 12% SDS-PAGE, using Coomassie blue staining (Figure 2A). The bands were identified by Western blot analysis, using anti-His monoclonal antibody (Figure 2B).

In the purification step, cellular proteins were passed through the column and only the E31C and E89C nonglycosylated erythropoietin (ngEPO) analogs containing His-tag attached to the column. The bonded proteins were eluted using elution buffer containing 500 mM imidazole (Figure 3).

### Characterization of PEGylated E31C and E89C ngEPOs

Conjugated E31C and E89C ngEPO analogs were analyzed using 12% SDS-PAGE, and the bands were visualized by Coomassie blue staining. As shown in Figure 4, bands approximately 40 kDa were attributed to PEGylated ngEPO analogs.

The PEGylated and nonPEGylated EPO cysteine analogs’ biological activity were analyzed using MTT assay method. UT-7 cells were cultured in the presence of different concentrations of PEGylated and nonPEGylated analogs. The glycosylated rhEPO (R&D Systems) was used as positive control in the assay (Table 2). EC₅₀ (ng/mL) was defined as the protein concentration that stimulates 50% of maximal stimulation. The conjugation yield was measured to be 79% and 82% for E31C and E89C analogs, respectively. In addition, PEG-conjugated ngEPOs showed approximately tenfold increase in plasma half-lives after injection into rats when compared to nonconjugated ones (Table 2).

The size distribution and the charge of conjugated and nonconjugated ngEPO analogs were also investigated. As seen

### Table 1 Characterization of stable cysteine analogs based on whole RMSD between cysteine analogs and modeled native erythropoietin, pKᵈ, and SAA of thiol groups cysteine analogs

<table>
<thead>
<tr>
<th>Analog</th>
<th>Whole RMSD</th>
<th>pKᵈ value</th>
<th>SAA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4C</td>
<td>2.991</td>
<td>9.00</td>
<td>54.2</td>
</tr>
<tr>
<td>L5C</td>
<td>2.889</td>
<td>9.00</td>
<td>60.6</td>
</tr>
<tr>
<td>T27C</td>
<td>3.250</td>
<td>8.03</td>
<td>57.1</td>
</tr>
<tr>
<td>E31C</td>
<td>2.942</td>
<td>9.00</td>
<td>100</td>
</tr>
<tr>
<td>E89C</td>
<td>2.706</td>
<td>9.00</td>
<td>96.6</td>
</tr>
<tr>
<td>K116C</td>
<td>4.025</td>
<td>8.56</td>
<td>72.6</td>
</tr>
<tr>
<td>D123C</td>
<td>3.628</td>
<td>9.00</td>
<td>88.7</td>
</tr>
<tr>
<td>A125C</td>
<td>3.440</td>
<td>7.12</td>
<td>51.8</td>
</tr>
<tr>
<td>L130C</td>
<td>3.739</td>
<td>9.00</td>
<td>78.1</td>
</tr>
<tr>
<td>T134C</td>
<td>2.985</td>
<td>8.91</td>
<td>52.2</td>
</tr>
<tr>
<td>R162C</td>
<td>3.127</td>
<td>8.38</td>
<td>72.9</td>
</tr>
<tr>
<td>T163C</td>
<td>3.204</td>
<td>8.64</td>
<td>51.9</td>
</tr>
<tr>
<td>R166C</td>
<td>3.116</td>
<td>8.89</td>
<td>70.8</td>
</tr>
</tbody>
</table>

Abbreviations: RMSD, root-mean-square deviation; SAA, surface area accessibility.

---

Figure 1: Computer generated representation of simulated E31C (A), and E89C ngEPO (B) analogs.

Notes: White arrows indicate free incorporated cysteine residues, and disulfide bonds are demonstrated by tube representation (yellow regions).

Abbreviation: ngEPO, nonglycosylated erythropoietin.
in Figure 5, the zeta potential of the E89C ngEPO analog was switched to the negative range due to attachment of 20 kDa PEG polymer. In addition, the size of E89C ngEPO analog was increased when a 20 kDa mPEG polymer attached to the analog due to an increase in hydrodynamic radius of conjugated form (Figure 5). The results were the same for conjugated and nonconjugated form of E31C ngEPO analog (data not shown).
Discussion

Recombinant therapeutic proteins have more limitations for application as a drug than chemical ones, such as sensitivity to temperature, pH, organic solvents, and degradable enzymes that lead to lower activity and stability either in vitro or in vivo. Therefore, protein engineering plays undoubtedly a remarkable role in biopharmaceutical development to improve pharmacodynamic and pharmacokinetic properties of biotechnological products.12,13 The effects of manipulations applied by protein engineering are usually unclear on the structure and function. Moreover, despite the progress in recombinant DNA technology, the production of sufficient quantity of protein(s) for chemical engineering acts as a bottleneck in the engineering process. This limitation becomes clearer when a complex biopharmaceutical drug with post-translational modifications (PTMs) is desired.14 For example, in our previous study, the site-specific PEGylation of EPO was investigated. It should be noted that the possible sites for the conjugation (even on consideration of experimental information in the literature) are very numerous.11,15 The condition is more complicated with variation in PEG sizes and shapes that could employed for PEGylation.16,17 Therefore, the examination of all possible sites is not practically achievable in a lab as it requires considerable time and cost. On one hand, the analysis of engineered protein is a problem. On the other hand, the amount of engineered protein must be sufficient for all in-process quality control tests.

In this study, two analogs, previously purposed by computational analysis for cysteine PEGylation, were subjected to conjugation with 20 kDa mPEG polymer. Two new conjugated ngEPO cysteine analogs had the same bioactivity, better physicochemical properties compared to nonconjugated ones, as well as high yield of PEGylation. Computational systems are the more feasible and less laborious way for high-throughput screening of desired changes in the protein. Our more recent study on glyco-engineering of interferon β reveals that computational systems are reliable, if the condition parameters for engineering are fully known. Computational calculation can decrease the possible sites for modification of proteins, especially for PEGylation and glyco-engineering as it is shown.11,16

In addition, the expression of the protein in eukaryotic cells, like CHO, is difficult, and PTMs made by such systems create many variations in final product. Therefore, the use of simpler expression system, like E. coli, for producing recombinant proteins needed in developmental steps of protein engineering is a promising approach. In this way, the homogenous protein is prepared faster and more efficiently as a starting material for engineering. Although, proteins derived from E. coli lack PTMs, like glycans chains, but screening at this level is easier than most onerous systems. Moreover, the obtained results are extensible for such systems, as we showed for E31C analog, which investigated in both prokaryotic and eukaryotic systems. In case of PEGylation, PTM deficiency can be compensated by a covalent binding of PEG polymer. As shown, by attachment of 40 kDa PEG polymer to ngEPOs, the half-life of ngEPO increases even more than CHO-expressed EPO.19 Also, the increase in size due to attachment of a 20 kDa mPEG polymer enhances the hydrodynamic radius of the macromolecule, potentially incurring repulsion against Bowman’s capsule, leading to lower kidney excretion, and longer in vivo half-life.20

Conclusion

The strategy of creating new forms of protein through computational and nonglycosylated approach can be considered an easy and economical solution to develop new biopharmaceutical products in a more time- and cost-effective manner. In addition, usage of nonglycosylated form of protein is a simpler method for the PEGylation reaction parameter optimization, such as temperature, pH, time, and PEG/protein ratio at industrial scale.

Table 2 In vitro biological activity, conjugation yield (%), and half-life determination of 20 kDa PEGylated and nonPEGylated cysteine analogs

<table>
<thead>
<tr>
<th>Protein</th>
<th>EC_{50} (ng/mL)</th>
<th>Conjugation yield (%)</th>
<th>Half-life (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO-derived rhEPO</td>
<td>0.42±0.02</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>E31C ngEPO analog</td>
<td>0.29±0.09</td>
<td>–</td>
<td>1.35±0.32</td>
</tr>
<tr>
<td>E89C ngEPO analog</td>
<td>0.30±0.03</td>
<td>–</td>
<td>1.41±0.71</td>
</tr>
<tr>
<td>20 kDa-mPEG-ngE31C</td>
<td>0.31±0.02</td>
<td>79</td>
<td>15.4±3.37</td>
</tr>
<tr>
<td>20 kDa-mPEG-ngE89C</td>
<td>0.33±0.06</td>
<td>82</td>
<td>16.3±2.54</td>
</tr>
</tbody>
</table>

Notes: Data are presented as mean ± SD; – values represent not applicable results.

Abbreviations: EC_{50}, half maximal effective concentration; rhEPO, recombinant human erythropoietin; mPEG, methoxypoly(ethylene glycol); ngEPO, nonglycosylated erythropoietin.
Acknowledgment
The authors wish to express their deep gratitude to all who provided support during the course of this research.

Disclosure
The authors report no conflicts of interest in this work.

References

Figure 5 Zeta potential and size distribution analysis of (A) nonconjugated and (B) conjugated E89C analog.