In Silico Investigation of a Chimeric IL24-LK6 Fusion Protein as a Potent Candidate Against Breast Cancer

Hafiz Muhammad Rehman[1,2], Hafiz Muzzammel Rehman[3,4], Muhammad Naveed[5], Muhammad Tahir Khan[6], Muhammad Aqib Shabbir[5], Shakira Aslam[1] and Hamid Bashir[1]

[1]Centre for Applied Molecular Biology (CAMB), University of the Punjab, Lahore, Pakistan. [2]University Institute of Medical Lab Technology, Faculty of Allied Health Sciences, The University of Lahore, Pakistan. [3]School of Biochemistry & Biotechnology, University of the Punjab, Lahore, Pakistan. [4]Department of Human Genetics and Molecular Biology, University of Health Sciences, Lahore, Pakistan. [5]Department of Biotechnology, Faculty of Science & Technology, University of Central Punjab, Lahore, Pakistan. [6]Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan.

ABSTRACT: Targeted delivery of therapeutic anticancer chimeric molecules enhances the efficacy of drug by improving cellular uptake and circulation time. Engineering the molecules to facilitate the specific interaction between chimeric protein and its receptor is critical to elucidate biological mechanism as well as accuracy in modeling of complexes. A theoretically designed novel protein-protein interfaces can serve as a bottom-up method for comprehensive understanding of interacting protein residues. This study was aimed for in silico analyses of a chimeric fusion protein against breast cancer. The amino acid sequences of the interleukin 24 (IL-24) and LK-6 peptide were used to design the chimeric fusion protein via a rigid linker. The secondary and tertiary structures along with physicochemical properties by ProtParam and solubility were predicted using online software. The validation and quality of the fusion protein was confirmed by Rampage and ERRAT2. The newly designed fusion construct has a total length of 179 amino acids. The top-ranked structure from alpha fold2 showed 18.1 KD molecular weight by ProtParam, quality factor of 94.152 by ERRAT, and a valid structure by a Ramachandran plot with 88.5% residues in the favored region. Finally, the docking and simulation studies were performed using HADDOCK and Desmond module of Schrodinger. The quality, validity, interaction analysis, and stability of the fusion protein depict a functional molecule. The fusion gene IL24-LK6 after cloning and expression in a suitable prokaryotic cell might be a useful candidate for developing a novel anticancer therapy.

Keywords: Interleukin 24, LK-6, fusion protein, molecular docking

Introduction

The term cancer describes more than 100 malignancies which affect different tissues and cells characterized by uncontrolled proliferating abnormal cells.[1] According to the World Health Organization, the mortality rate will increase up to 80% within a decade.[2] Among different types of cancers, breast cancer is the most common cause of deaths after lung cancer. The underlying mechanism of heterogeneous breast cancer development, both histopathologically and genetically, is still uncertain.[3] The monotherapeutic treatment selection regarding cancer includes surgery, hormonal therapy, radiotherapy, and chemotherapy. Despite the success of the applied monotherapies, the chemotherapeutic drugs for cancer treatment have serious drawbacks of nonspecific toxicity. These monotherapies' success is largely limited due to its adverse side effects and remarkable increase in resistance.[4] The conventional chemotherapeutic drugs against cancer have some limitations such as damaging the healthy growing cells and development of resistance against the drugs.[5] Some of the drugs cause secondary tumors in the body, although they do not affect the slow growing and resting cells.[6] Cytokines have potential antitumor activities and play important roles in the regulation of the immune system. Among these cytokines, the most potent interleukins (ILs) such as IL-2, IL-15, and IL-24 have been evaluated extensively for their promising anticancer drug potential.[7,8] But the use of these cytokines to boost the immune activity against the cancer cell had limited success due to the requirement of high-dose administration. The vast and existing knowledge of structural and functional characteristics of protein can be exploited to design a novel peptide. An optimistic approach in cancer treatment is targeted therapy. In anticancer targeted therapy, the drugs with reduced side effects selectively target the cancer cells, without being affected by chemoresistance mechanisms[9] which lead to enhanced drug efficiency.[10] The peptide alone or in fusion with other proteins could be a choice of drug to target tumor

Table 1. Primary sequence of molecules.

PEPTIDE	AMINO ACID SEQUENCE	NUMBER OF AMINO ACIDS	REFERENCE
Interleukin 24	QEFHFGPCQVKGVVPQKL	155	http://www. uniprot.org/
	WEAFWAVKDTMQAQDNIT		(Accession # Q13007)
	SARLLQQEVLQNVSDAES		
	CYLVHTLLEFYLKTVFKN		
	YHNRTVEVRTLKSFSTLA		
	NNFVLIVSQLQPSQENEM		
	FSIRDSAHRRFLLFRRAFK		
	QLDVEAALTKALGEVDIL		
	LTWMQKFYKL		
Linker	PAPAPAPAPAP	11	
LK-6	KLLKKIKSLIKKI	13	

cells.[11] The advantages of using peptide over small molecules are lower toxicity profiles and high specificity to their targets.[12] Using a chimeric protein, ie, killer peptide along with the tumor-targeting peptide, is a good approach in targeted therapy against cancerous cells.[13,14] Fusion of tumor-targeting peptide and IL to make a recombinant protein may have a potential to treat the human cancer more effectively due to their synergistic effects. The benefits to use peptide therapeutics as a mean of anticancer agent include the ease of peptide modification, rapid peptide synthesis, and 10- to 20-fold high therapeutic potential compared with the individual monomeric protein when joined through recombinant DNA technology. An existing and vast knowledge base of structure and function of protein can be exploited to design a novel peptide.[15,16]

Interleukin 24, a member of IL-10 family, plays an important role in autoimmune diseases, inflammation, and action against infections.[17] It induces apoptotic signal when interacts specifically with its heterodimer receptors (IL-20R1/IL-20R2 and IL-22R1/IL-20R2) on MCF-7 cell and subsequently activates signal tranducer and activator of transcription (STAT), Janus kinase/signal transducer pathways.[1,18,19]

LK-6 is a lysine/leucine-rich cationic antimicrobial penetrating peptide known for its anticancer activity preferentially against human MCF-7 breast cancer cells. After interaction with MCF-7 cell membrane phosphatidylserine, the LK-6 peptide is internalized via clathrin-independent macropinocytosis without disrupting cell surface and induces a dramatic nuclear damage which leads to MCF-7 cell death. A recombinant protein having tumor-targeting peptide and a cytokine could have a potential to treat human cancer more effectively due to bifunctional domains of the peptide without harming normal cells.[20]

Therefore, we aimed to produce the fusion protein IL24-LK6 to benefit from both selective delivery and selective activity on cancer cells to offer a novel targeted anticancer treatment. A new drug candidate is first evaluated by in silico methods for structural and functional characteristics. Further to understanding the macromolecular structure to function relationships, molecular dynamics simulation is performed. Through MD simulation, intriguing aspects such as molecular interactions, physical processes, protein minimum geometries, drug binding free energies, and precise movements of molecules or atoms over a set length of time can be examined. In addition, this method has been able to evaluate the positioning, wrapping, distance, interactions between residues, and potential involvement of hazardous and targeted moieties.[21-23] So the main objective of this study was to construct a theoretical fusion peptide comprising cytokines IL-24 and a cancer-specific cell-penetrating anticancer peptides LK-6 by silico methods against the IL22RA-IL20RB receptor and evaluate its anticancer therapeutic potential.

Materials and Methods

Construction of IL24–LK6 fusion protein

FASTA format amino acid sequence of *Homo sapiens* IL-24 was retrieved from Universal Protein Knowledgebase (UniProtKB, http://www.uniprot.org/: Accession # Q13007) database. The amino acid sequence of LK-6 peptide was reported previously by Wang.[20] The 3-dimensional (3D) crystalline structure of soluble IL-24 heterodimer receptors (IL-22RA and IL-20RB, PDB ID: 6DF3) in PDB file format was obtained from protein data bank (RCSB PDB: 6DF3) (Table 1). For chimeric molecule designing, we excluded signal peptide sequence (1-56 amino acids) from N terminal of IL-24

and used 52 to 206 amino acids. LK-6, a peptide of 13-amino acid length, was linked to IL-24 via a previously reported rigid linker (PAPAPAPAPAP).[18]

Secondary structure of chimeric protein

For secondary structure prediction of chimeric protein, GOR IV (http://gor.bb.iastate.edu/) online server was used and further functional characteristics such as secondary structure, coiled-coil domain, regions lacking normal structure, solvent-accessible surface area, low-complexity sections, and location of disulfide bridges were assessed.[24]

Homology modeling to predict 3D structure and validation

By using the primary sequence of protein, 3D structure was generated by 3 different software programs, including AlphaFold 2 (https://colab.research.google.com),[25] I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/),[26-28] and trRosetta (https://yanglab.nankai.edu.cn›trRosetta).[29,30] A template (PDB ID: 6DF3) was used by these software programs to build the model. Next, the quality and validity of newly constructed models were evaluated on different programs such as Verify3D (http://services.mbi.ucla.edu/Verify_3D)[31] which determines amino acid sequence compatibility to 3D atomic model, ERRAT2 (https://saves.mbi.ucla.edu/results?job=1033540&p=errat) which checks overall quality factor (OQF), and Rampage server (http://mordred.bioc.cam.ac.ukrapper/rampage.php)[32] which can generate the Ramachandran (RC) plot. Further, the galaxy refine server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE)[33] was employed before docking studies to refine and minimize the energy of modeled proteins to improve the structural geometry and to diminish significant steric clashes produced during the model building.

Physiochemical parameters of IL24–LK6 chimeric protein

The physical and chemical characteristics of chimeric protein were assessed theoretically by ProtParam (https://web.expasy.org/protparam/). This software calculated the isoelectric point (pI), molecular weight, amino acid composition (%), extinction coefficient, aliphatic index, and grand average of hydropathicity (GRAVY) using bioinformatics algorithms.[34] The solubility of chimeric protein was predicted on protein-sol online server (https://prosa.services.came.sbg.ac.at/prosa.php).

Toxicity, antigenicity, and allergenicity prediction

To predict the toxicity, antigenic, and allergenic nature of chimeric protein, the FASTA format amino acid sequence was submitted to online servers of ToxinPred, VaxiJen, and AlgPred, respectively. ToxinPred (https://webs.iiitd.edu.in/raghava/toxinpred/index.html) predicted the regions of toxic amino acid residues in the submitted sequence. AlgPred (https://webs.iiitd.edu.in/raghava/algpred2/batch.html) is based on the similarity index of any protein region with known epitope. VaxiJen server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) calculated antigenicity depending on the physical and chemical properties of the submitted protein sequence. We used a threshold of 0.5 to differentiate between the antigenic and nonantigenic proteins.[35,36]

Molecular docking and interactions

Because IL-24 generates apoptotic signal when binds to its receptor therefore to observe the interaction and the binding pattern between modeled protein and its specific receptor, HADDOCK v.2.4 was engaged for docking.[37] PyMOL molecular graphic system was adopted for removal of water molecules and attached IL-24 from receptor in the PDB downloaded file format of crystalline structure (PDB ID: 6DF3) along with active site prediction of receptor protein before proceeding to docking. The protein-protein interactions were analyzed on PDBePISA and PDBsum which served to identify interacting interfaces, hydrogen bonds, Gibbs free energy (ΔG^{int}, kcal mol^{-1}), salt bridges, nonbonded contacts, tunnels, and pores of complex.[38,39] Finally, the binding affinity was calculated from prodigy online server (https://wenmr.science.uu.nl/prodigy/).[40]

Molecular dynamics simulation

The Desmond module of Schrodinger was exploited to conduct MD simulation studies. The dynamic behavior and stability of the protein-protein complexes were investigated using their docked poses. The protein-protein complex was preprocessed using Protein Preparation Wizard of Maestro, which included complex optimization and minimization. All the systems were prepared using the System Builder tool. Solvation of the complexes was performed with the simple point-charge (SPC) water model with orthorhombic box, along with a 10-Å distance from the edge of the box, and the system was neutralized with Na$^+$/Cl$^-$ ions. To mimic physiological conditions, 0.15-M sodium chloride (NaCl) was added. The potential energy of the protein complex was minimized by employing the constant- temperaure, constant-pressure (NPT) ensemble. The molecular dynamics simulations were performed at 300-K temperature and 1-atm pressure for 100-ns and NPT production ran under the OPLS4 force field. The models were relaxed before the simulation. The short-range electrostatic interactions were calculated using the particle mesh Ewald method.[41] The cutoff radius in Coulomb interactions was 9.0 Å. The water

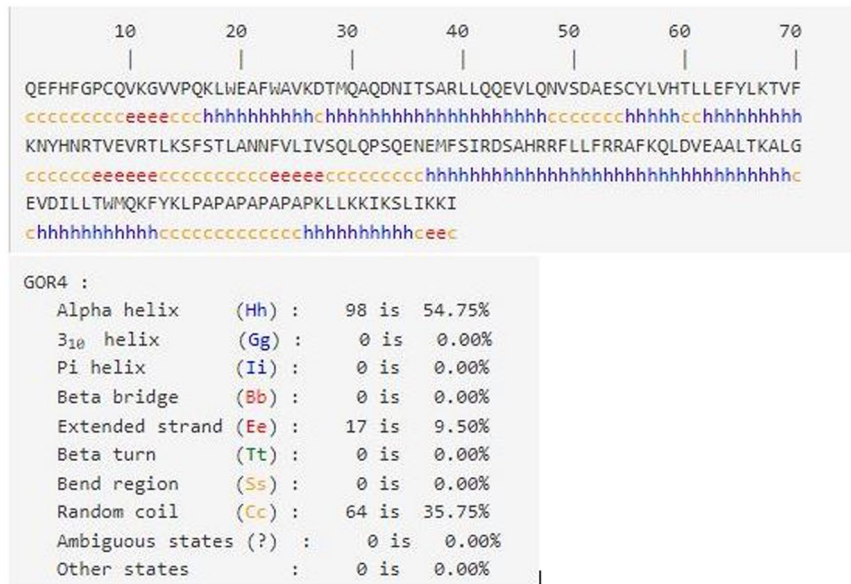**Figure 1.** Schematic representation of IL24-LK6 fusion protein. IL indicates interleukin.

Figure 2. Schematic illustration of primary and secondary structure prediction by GOR IV.

molecules were explicitly described using the SPC model.[42] The Martyna-Tuckerman-Klein chain coupling scheme with a coupling constant of 2.0 ps was used for the pressure control and the Nosé-Hoover chain coupling scheme[43] for the temperature control. The trajectories were saved for examination after every 100 ps, and the simulation's stability was verified by comparing the root mean square deviation (RMSD) of the protein complex over time. The projected changes in their conformation from the initial structure over the entire simulation period were expressed as RMSD and root mean square fluctuation (RMSF) for MD simulations. A detailed description of the methodology can also be found elsewhere.[44-46]

Results and Discussion

Engineering of IL24–LK6 fusion protein

The sequence of IL-24 and LK-6 was retrieved from the National Center for Biotechnology Information and fused via a rigid linker to construct a single peptide of 179 amino acids (Figure 1). The presence of rigid linker ensures the separation of both functional peptides, provides stability, and prevents the formation of disulphide bond. The specific restriction sites were added to both the ends, N and C terminals, of the fusion construct to be cloned into a suitable vector for in vitro expression studies.

Secondary structure analysis

The secondary structure was predicted by GOR IV, that defines the alpha helix and beta sheets of protein, and it is the intermediate structure between primary and tertiary structure. Our modeled protein contains 54.75% (98 residues) alpha helix, 9.5% (17 residues) beta sheets, and 35.75% (64 residues) random coils (Figure 2). The higher percentage of alpha helix imparts structural stability to the molecule. The predicted structure of the rigid linker is a coil that will surely separate IL-24 and LK-6 molecules apart ensuring their functional capabilities.

Tertiary structure evaluation

The tertiary structure of the chimeric protein was prepared by operating homology modeling. FASTA sequence was submitted to 3 different 3D protein builder online servers AlphaFold 2 colab, I-TASSER, and trRosetta to get the most accurate and stable configuration of fusion construct. AlphaFold 2 predicts the 3D protein structure based solely on its primary amino acid sequence with high accuracy of side chains. The output file of AlphaFold 2 contains 5 different models ranked on the basis of local model quality of the protein structure. The top AlphaFold 2 model from the output file was selected for comparison with other models. I-TASSER produced 5 top models based on the C-score with a higher model value signifying more confidence.[47] So the top I-TASSER model with

the highest C-score of –0.39, TM-score of 0.66, and RMSD of 5.9 was selected for comparison with other models. Like I-TASSER, trRosetta also constructed 5 top models based on highest TM-scores with restrains from both homologous templates and deep learning. A model with the highest TM-score of 0.745, Z-score of 37.8, and E-score of 1.2E-4 was selected for comparison with other models. After comparison of all top models obtained from 3 different servers, such as I-TASSER, AlphaFold 2 colab, and trRosetta, the best 3D model was presented by AlphaFold 2 so that the structure was selected and proceeded for further analysis (Figure 3).

Reliable 3D structure selection, quality check, and validation

The Ramachandran plot provides a distribution of phi and psi angles of each amino acid residue with a higher percentage of
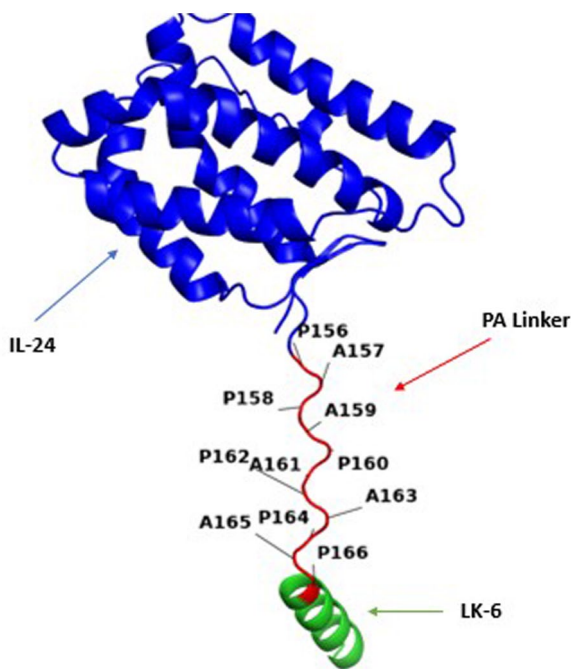

Figure 3. The ribbon structure visualization of IL24-LK6 fusion protein on PyMOL illustrates the 2 protein domains are completely separated by a rigid linker with no interaction which ensures the integrity and functional capability of both segments. IL indicates interleukin.

amino acid in the preferred region and in the additionally allowed region suggests the overall structural reliability, accuracy, and validation.[48] A reliable hit from 3 predicted models was selected based on the quality score and stereo chemical validation (Table 2).

The selected modeled fusion construct showed 88.5% residues in the allowed region, 10.3% in the additionally allowed region, and 1.2% in the generously allowed region of the RC plot. The only amino acid glycine is known by its symmetrical alpha carbon property and sometimes giving unusual behavior in the RC plot (Figure 4).

Furthermore, Verify 3D program (https://saves.mbi.ucla.edu/results/job=1;032202&p=verify) was used to measure the compatibility of modeled protein with the primary sequence using a 3D profile.[41] Residues having averaged 3D to 1D score ⩾0.2 with the overall result "Pass" are considered for further studies (Figure 5).

Finally, ERRAT2 program (website) predicts the OQF which expresses the percentage of protein for which the calculated error value falls below the 95% rejection limit, with equal or more than 95% indicating good high-resolution structure (Figure 6). According to the results, the most precise 3D structure was modeled by AlphaFold 2 with ERRAT OQF of 98.63, Verify 3D score of 95.48%, and 95.2% residues in the favorable region of the RC plot (Figure 3). The selected model's energy plot was generated by proSA-web online server to indicate the erroneous residues of protein. The positive residue energy represents invalid or wrong amino acid position (Figure 7). The AlphaFold 2 fusion construct was further proceeded to assess solubility and docking studies.

Predicted physicochemical parameters of fusion protein

ProtParam online server computed the physicochemical parameters shown in Table 3. The presence of many basic amino acids (Arg + Lys = 24) as compared with acidic (Asp + Glu = 16) imparts alkaline nature to the chimeric molecule with a theoretical pI of 9.57. The extinction coefficient of the modeled protein is $22\,585\,M^{-1}\,cm^{-1}$, an important factor in protein-protein interaction. The in vitro estimated

Table 2. Quality comparison of homology modeling on different software programs.

PROGRAM	RC	ERRAT QUALITY SCORE	VERIFY 3D SCORE
AlphaFold 2	95.2	98.639	95.48
trRosetta	92.7	95.322	87.71
I-TASSER	92.7	85.96	89.39

Abbreviation: RC, Ramachandran plot values in the favored region.
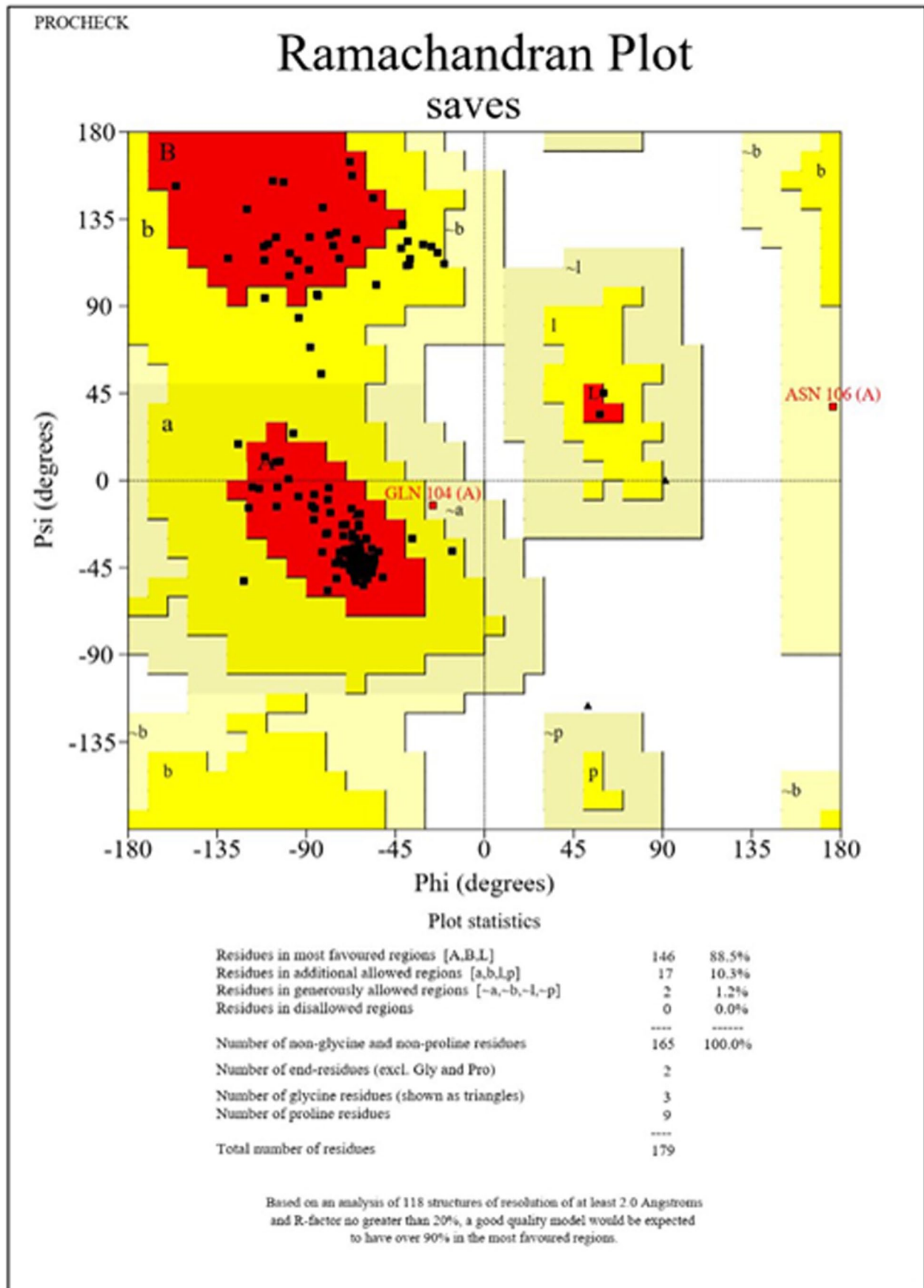All values are expressed in %.

Figure 4. Ramachandran plot of IL24-LK6 fusion protein shows the presence of 88.5% amino acid residues in the favorable region (red), 10.3% in the allowed region (yellow), 1.2% in the generously allowed region (light brown), and no amino acid in the disallowed region (white). IL indicates interleukin.
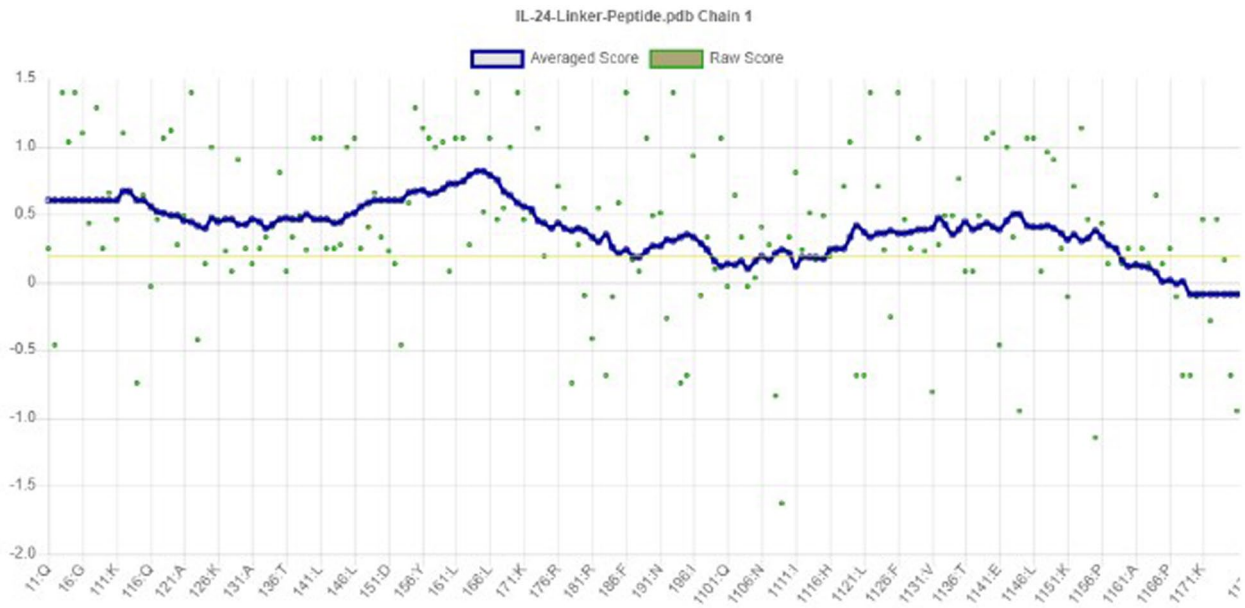
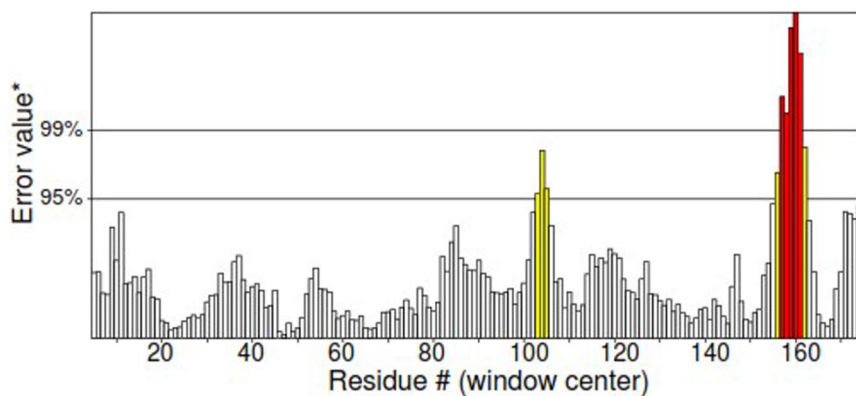**Figure 5.** Verify 3D graphical representation.

Figure 6. Graphical representation of quality error prediction by ERRAT2 program for each amino acid in IL24-LK6 fusion protein. IL indicates interleukin.

half-life in a mammalian cell is 0.8 hour and the instability index is 46.71.

A high aliphatic index of 96.98 indicates a wide temperature range for protein stability. The reactivity of chimeric protein in water is estimated by a low GRAVY index of –0.135. Moreover, ProtPram calculated the highest percentage of Leu (12.8%) in the chimeric construct. For solubility prediction, protein-sol software was employed which compares the solubility of the query FASTA sequence of amino acid to its database and any value greater than 0.45 is predicted as soluble.[49] Our selected model has the scaled solubility value of 0.562 which can be predicted as soluble (Figure 8).

Toxicity, allergenicity, and antigenicity evaluation

ToxinPred, AlgPred, and VaxiJen designated the IL24-LK6 fusion protein as nontoxic, nonallergen, and nonantigenic, respectively. None of the peptide residue exhibits toxicity (data not shown) with an overall score of 0.4478 for protective antigen prediction. The results suggest a nonimmunogenic nature of chimeric protein and propose its use as a good anticancer drug candidate.

Docking analysis

To elucidate the mechanism of interaction between fusion protein and its receptor, HADDOCK protein-protein docking was performed. For interaction studies, the top structure, based on score, electrostatic energies, and van der Waals attraction, has been selected. Among protein-protein docking, the major interactions are electrostatic and van der Waals which mutually find accurate results along with promising levels as scoring function.[50] HADDOCK server clustered a total 159 structures in 13 clusters representing 79% of water-refined models generated by HADDOCK. Table 4 shows the statistical parameters of top 7 clusters.
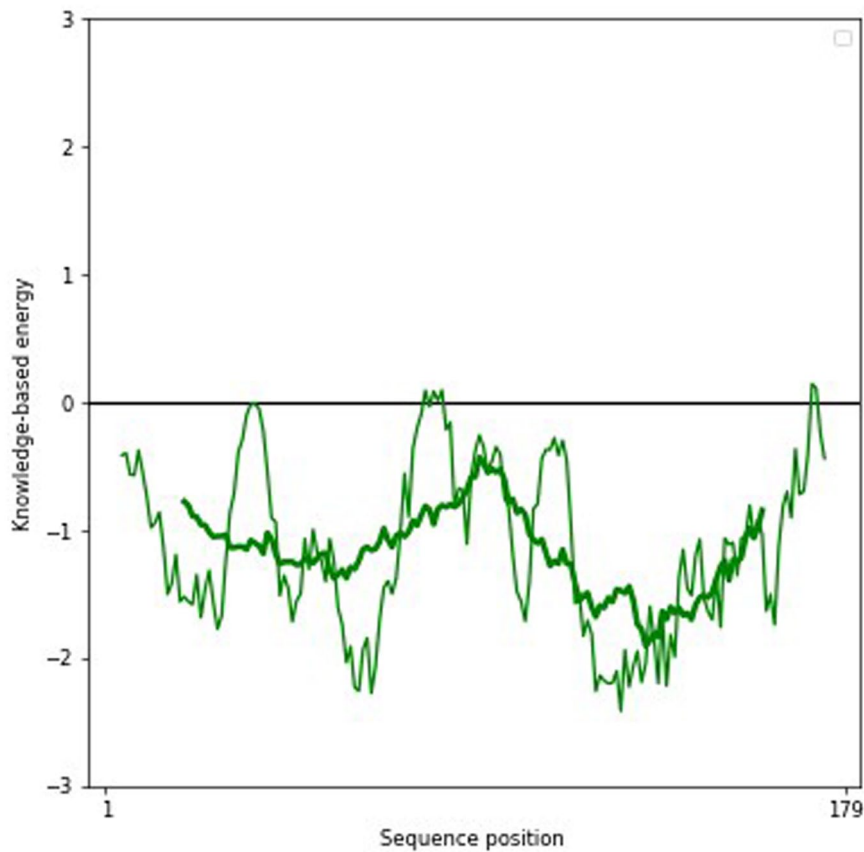
Figure 7. Graphical representation of local quality estimate of each amino acid in fusion protein.

Table 3. Physical and chemical properties of fusion construct.

PHYSIOCHEMICAL PROPERTIES	IL24-LINKER-LK6
Number of amino acids	179
Theoretical pI	9.57
Molecular weight	20636.13
Instability index	46.71
Aliphatic index	96.98
Total positively charged amino acids	(Arg + Lys): 24
Total negatively charged amino acids	(Asp + Glu): 16
GRAVY	−0.135
Coefficient of extinction (in M⁻¹cm⁻¹ at 280nm)	22585

Abbreviations: GRAVY, grand average of hydropathicity; IL, interleukin; pI, isoelectric point.

Figure 8. The population average for the experimental dataset (PopAvrSol) on protein-sol.

The Z-score signifies the number of standard deviations from the mean of the cluster location of score (negative value indicates better results). The more negative value of electrostatic and van der Waals interactions depicts more stable docked complex. Moreover, the binding strength of
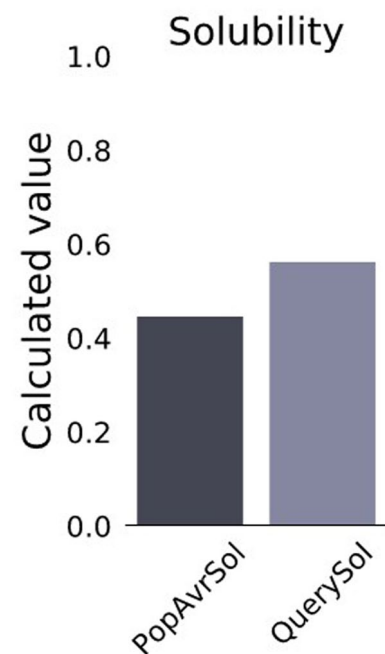
the complex depends on 2 factors, the total number of interacting amino acid residues and area of corresponding interface. Larger interface area depicts larger binding energy

Table 4. Protein-protein docking at HADDOCK server.

	CLUSTER 4	CLUSTER 2	CLUSTER 3	CLUSTER 11	CLUSTER 9	CLUSTER 6	CLUSTER 13
HADDOCK score	-136.3 ± 7.0	-130.4 ± 4.1	-95.5 ± 10.2	-91.7 ± 17.9	-86.1 ± 20.0	-84.5 ± 9.2	-74.2 ± 18.9
Cluster size	13	33	18	5	7	9	4
RMSD	0.7 ± 0.4	1.0 ± 0.1	28.7 ± 0.1	30.2 ± 0.4	2.8 ± 0.3	25.3 ± 0.3	30.8 ± 0.93
Van der Walls energy	-83.5 ± 8.1	-77.2 ± 4.9	-64.4 ± 5.2	-57.9 ± 15.9	-57.4 ± 12.6	-58.4 ± 11.6	-61.6 ± 7.5
Electrostatic energy	-323.5 ± 48.4	-325 ± 17.7	-323.8 ± 24.4	-316.8 ± 84.1	-227 ± 20.9	-288.5 ± 68.4	-236.9 ± 58.4
Desolvation energy	-15.0 ± 1.6	-9.7 ± 3.5	-5.5 ± 2.9	-9.0 ± 6.8	-10.3 ± 0.7	-4.2 ± 8.5	2.1 ± 3.8
Restraint violation energy	269.0 ± 40.2	216.0 ± 28.7	391.8 ± 51.3	384.7 ± 104.9	270.0 ± 28.5	357.4 ± 39.1	368.7 ± 54.3
Buried area	3219.6 ± 115	2984.5 ± 79.8	2465.8 ± 92.1	2435 ± 104.9	2283.4 ± 227	2489.0 ± 184.1	2372.5 ± 42.4
Z-scores	-1.8	-1.6	-0.3	-0.1	0.1	0.1	0.5

Abbreviation: RMSD, root mean square deviation.

Table 5. Interface statistics (chains A, B, and C have been labeled in Figure 9).

CHAINS	IA	IR	SB	DS	HB	NBC	TUNNEL
A:B	180:160	3:2	-	-	-	14	-
A:C	1194:1172	25:20	2	-	5	132	1
B:C	456:437	8:8	-	-	4	38	1

Abbreviations: DS, number of disulphide bonds; HB, number of hydrogen bonds; IA, interface area ($Å^2$); IR, number of interface residues; NBC, number of nonbonded contacts; SB, number of salt bridges.

Table 6. Binding affinities of fusion construct.

FUSION PEPTIDE	ΔG (kcal mol-1)	K_D (M) AT 25.0°C	K_D (M) AT 40.0°C
IL-24-(PS)$_5$P-LK-6	-13.6	1.1E–10	3.2E–10

Abbreviation: IL, interleukin.

(van der Waals + electrostatic + H-bond) and more stable structure.[51]

Residue interaction analysis

Computational estimation of binding energy in protein-protein interface is a challenging problem, although dozens of tools are available based on analytical or empirical approaches; however, the performance of each tool varies according to the particular protein and we cannot pinpoint a single module which provides a reliable quantification.[18] The interaction studies of docked complex analysis were resolved by PDBePISA and PDBsum servers (Table 5). In addition, salt bridges, mapping of interacting hydrogen bonds, and ΔG^{int} were also evaluated. ΔG^{int} value expresses free energy of solvation attained during formation of assembly (total energy of solvation in assembled structure – isolated structure solvation energies) (Table 6).[52] The docked complex revealed a total of 7 interactions mediated by 2 salt bridges and 5 hydrogen bonds with IL-22 alpha 1 subunit of IL-24 heterodimer receptor. According to PDBePISA, the formation of complex is primarily contributed by the interaction between fusion protein and IL-22R1 subunit (Figure 9(B)) which is in agreement with previous studies.[37] The residues at interface between these chains extend over 1181.5 Å with stabilization energy of -8.6 kcal mol^{-1} involving Ser[105], Met[109],
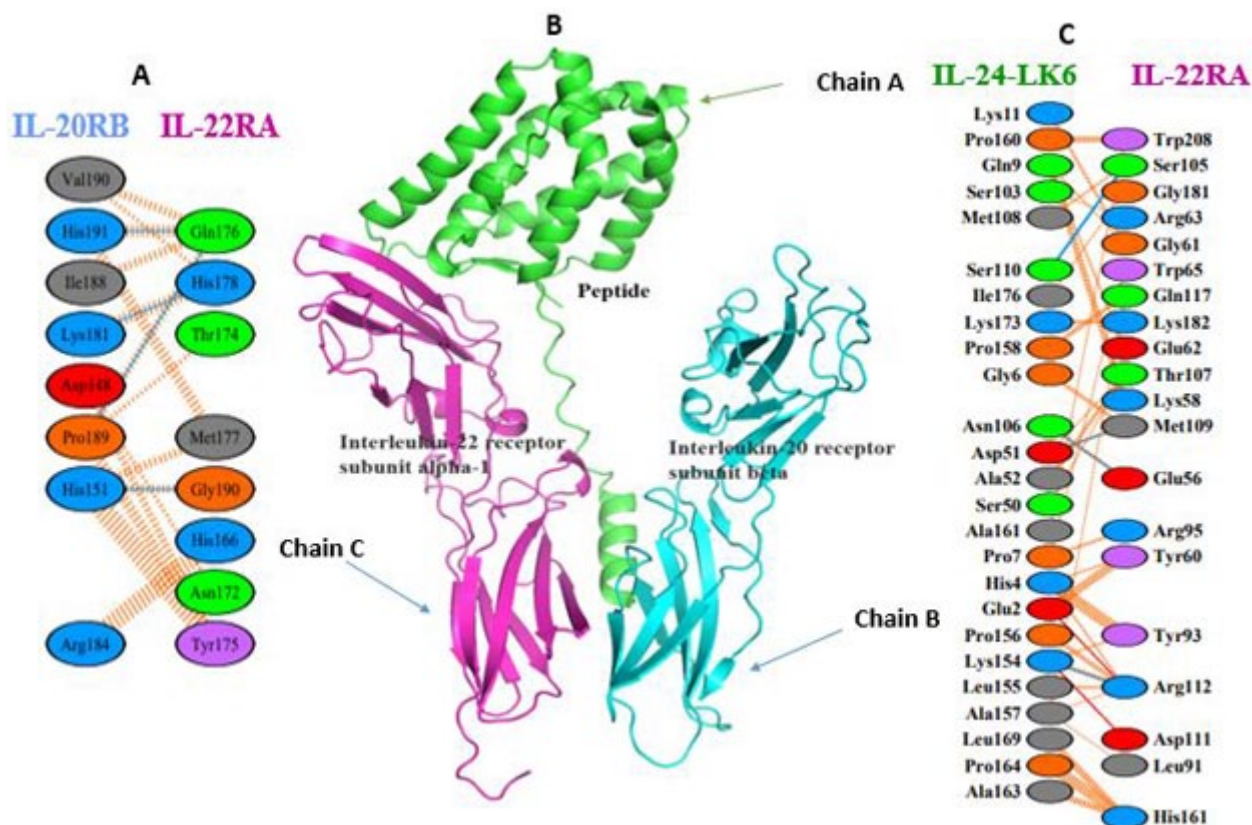
Figure 9. (A, C) PDBsum's interaction plot of chimeric molecule docked with receptor protein. The key indicates hydrogen bond (blue line), salt bridges (red line), disulfide bonds (yellow line), and nonbonded contact (orange tick-mark) for amino acid residues at protein-protein interface. (B) Docked peptide with IL-22 alpha-1 receptor and IL-20 beta receptor. IL indicates interleukin.
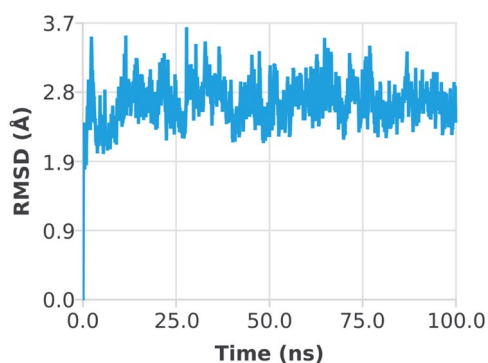

Figure 10. Root mean square deviation plot of IL24-LK6 fusion protein with receptor complex. IL indicates interleukin.

Glu[56], Ser[105], Arg[112] and Ser[110], Asn[106], Asp[51], Ser[110], and Lys[154] in hydrogen bonding contributed by fusion construct and IL-22 alpha 1 subunit, respectively. Surprisingly, no interaction was observed between fusion construct and IL-20 receptor beta subunit. Figure 9(A) and (C) illustrates the binding residues and interacting surfaces. The potential salt bridges were accomplished between Glu[2], Lys[154] and Arg[112], and Asp[111] from fusion construct and IL-22 alpha 1 subunit, respectively. The scores for binding affinity were predicted as a temperature function. The significant increase of Kd value

from 25°C to 40°C indicates a decreased binding affinity for receptor protein having a similarity with the previous study (Table 6).[53]

Molecular dynamics simulation

Molecular dynamics simulations were performed on the top hits containing high binding energies. Over the simulation period, the projected conformational changes from the initial structure were presented in RMSD. Moreover, structural stability, atomic mobility, and residue flexibility at times of interaction of protein hit were expressed with RMSF values. The peaks of RMSF graph represent the fluctuation portion of the protein through the simulation. The N- and C-terminals show more changes than any other portion of the fusion protein. Alpha helices and beta strands show less fluctuation, as they are stiffer than the unstructured part of protein, than loop portion. The RMSD of the complex showed small deviation until almost 10 ns and then there was deviation of 1 Å system throughout the simulation. It indicates the stability of the IL24-LK6 fusion protein/receptor complex and whether the simulation has equilibrated (Figure 10). Similarly, the RMSF is useful for characterizing local changes along the protein chain.
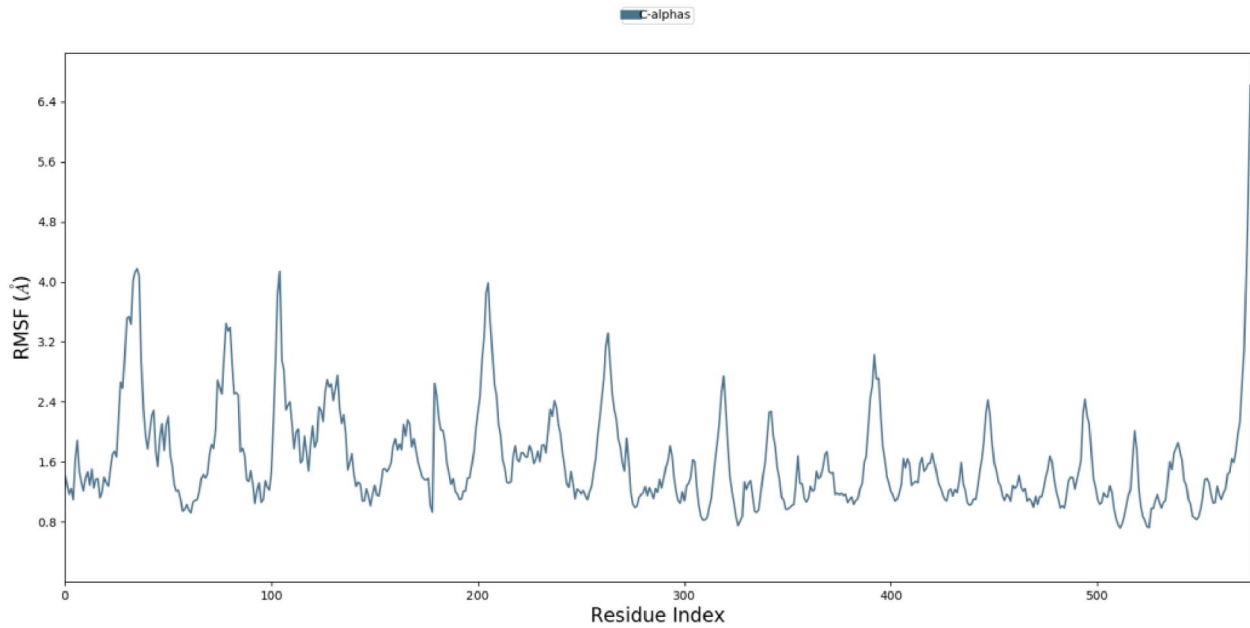
Figure 11. Root mean square fluctuation plot of IL24-LK6 fusion protein with receptor complex. IL indicates interleukin.
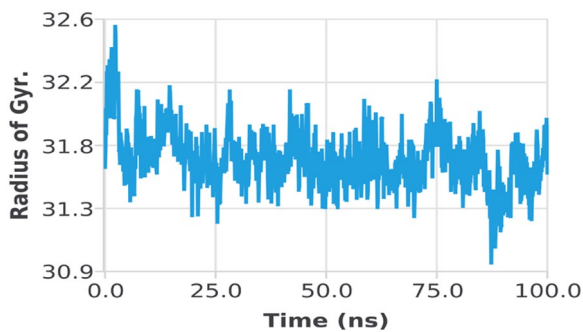

Figure 12. Radius of gyration of IL24-LK6 fusion protein with receptor complex. IL indicates interleukin.

Peaks reflect sections of the proteins that fluctuate the most during the simulation on the RMSF. IL24-LK6 fusion protein tails (both N- and C-terminals) are more likely to change than other regions of the protein. Alpha helices and beta strands, for example, are usually stiffer and less fluctuating than the unstructured component of the protein. The residues with higher peaks, according to MD trajectories, belong to loop areas or the N- and C-terminal zones. For RMSF, there was fluctuation of almost 4.17 Å of ILE 35, 3.44 Å of GLU 79, 4.14 Å of GLU 104, 3.98 Å of GLY 61, and 3.13 Å of SER 119; the remaining structure was made stable comparatively (Figure 11). The radius of gyration is the distribution of atoms in a protein around its axis (Rg). Rg denotes the distance between the rotating point and the location where the energy transfer has the greatest impact. This concept can also be used to identify different polymer types, such as proteins. The calculation of Rg and distance calculations are the 2 most essential markers for predicting a macromolecule's structural activity. The rate at which a protein fold is proportional to its compactness can be measured using a sophisticated computer method for calculating the gyration radius. The value of Rg increases from 31.5 to 32.6 until almost 5 ns; after that, there was almost equilibrium until 100 ns (Figure 12). The analyses of energy parameters for complex indicate that the total energy of the system was decreased and hence the complex gets stable (Figure 13).

Conclusion

In silico approach is deemed as an amenable alternate method for designing antitumor fusion protein. The use of computational methods can fairly decrease the time and cost of peptide designing and it also avoids the ethical aspects of in vivo experiments. Herein, by in silico approach, we have designed IL24-LK6 fusion protein exhibiting the potential of enhanced activity against breast cancer. The predicted model was assessed for quality, validation, physicochemical properties, solubility, docking, interaction studies, and molecular dynamics simulation. The findings suggest that our chimeric protein could be a stable antitumor candidate. The concept is to fuse a targeting peptide with a penetrating peptide via a rigid linker to keep both moieties intact, which could enhance the antitumor activity by dual action. This theoretically presented model would lay the groundwork for our in vitro studies in future to reveal the better insights of its anticancer activity.
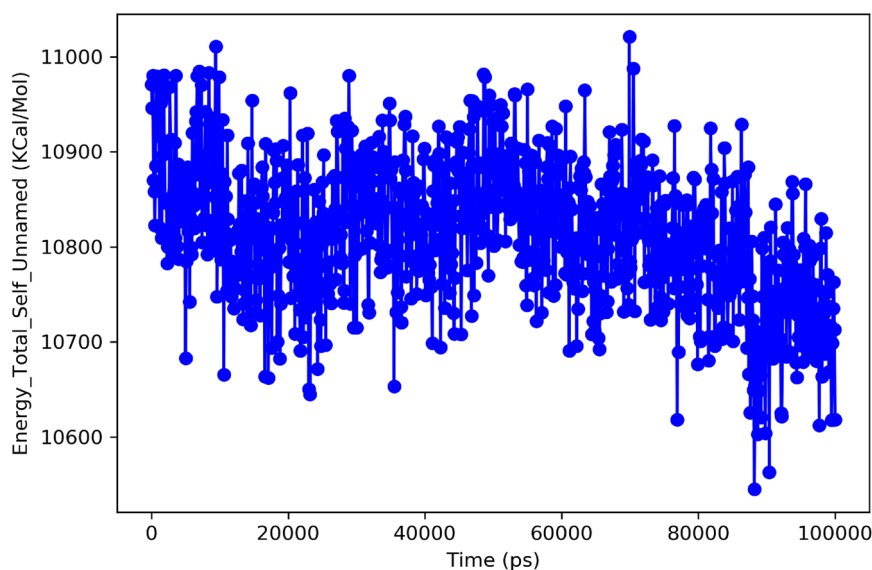
Figure 13. The total energy of docked complex.

Author Contributions

HMR, HB, and HMR helped in methodology, formal analysis, and validation; MN and MTK contributed to investigation resources; SA helped in data curation; and HMR, HB, and HMR prepared the original draft. All authors have read and agreed the published version of the article.

Data Availability

Data could be provided from Dr Hamid Bashir, email: hamid.camb@pu.edu.pk on request.

REFERENCES

1. Ghavimi R, Mohammadi E, Akbari V, Shafiee F, Jahanian-Najafabadi A. In silico design of two novel fusion proteins, p28-IL-24 and p28-M4, targeted to breast cancer cells. *Res Pharm Sci*. 2020;15:200-208.

2. Keshtvarz M, Salimian J, Amani J, Douraghi M, Rezaie E. In silico analysis of STX2a-PE15-P4A8 chimeric protein as a novel immunotoxin for cancer therapy. *In Silico Pharmacol*. 2021;9:19.

3. Sahna KO, Cakir B, Tunali-Akbay T. Antiproliferative activity of whey and casein bioactive peptides on breast cancer: an in vitro and in silico study. *Int J Pept Res Ther*. 2022;28:1-12.

4. Zahreddine H, Borden KL. Mechanisms and insights into drug resistance in cancer. *Front Pharmacol*. 2013;4:28.

5. Chabner BA, Roberts TG. Timeline: chemotherapy and the war on cancer. *Nat Rev Cancer*. 2005;5:65-72.

6. Naumov GN, Townson JL, MacDonald IC, et al. Ineffectiveness of doxorubicin treatment on solitary dormant mammary carcinoma cells or late-developing metastases. *Breast Cancer Res Treat*. 2003;82:199-206.

7. Smyth MJ, Cretney E, Kershaw MH, Hayakawa Y. Cytokines in cancer immunity and immunotherapy. *Immunol Rev*. 2004;202:275-293.

8. Balimane PV, Chong S. Cell culture-based models for intestinal permeability: a critique. *Drug Discov Today*. 2005;10:335-343.

9. Smith LL, Brown K, Carthew P, et al. Chemoprevention of breast cancer by tamoxifen: risks and opportunities. *Crit Rev Toxicol*. 2000;30:571-594.

10. Bae YH, Park K. Targeted drug delivery to tumors: myths, reality and possibility. *J Control Release*. 2011;153:198.

11. Thundimadathil J. Cancer treatment using peptides: current therapies and future prospects. *J Amino Acids*. 2012;2012:967347.

12. Sato AK, Viswanathan M, Kent RB, Wood CR. Therapeutic peptides: technological advances driving peptides into development. *Curr Opin Biotechnol*. 2006;17:638-642.

13. Ellerby HM, Bredesen DE, Fujimura S, John V. Hunter-killer peptide (HKP) for targeted therapy. *J Med Chem*. 2008;51:5887-5892.

14. Boohaker RJ, Lee MW, Vishnubhotla P, Perez JM, Khaled AR. The use of therapeutic peptides to target and to kill cancer cells. *Curr Med Chem*. 2012;19:3794-3804.

15. Chen X, Zaro JL, Shen WC. Pharmacokinetics of recombinant bifunctional fusion proteins. *Expert Opin Drug Metab Toxicol*. 2012;8:581-595.

16. Schmidt SR. Fusion—proteins as biopharmaceuticals—applications and challenges. *Curr Opin Drug Discov Devel*. 2009;12:284-295.

17. Cruz A, Nguyen B, Sauane M, Lopez GE. Structural and functional characterization of interleukin-24 based on atomistic molecular modeling. *Chem Lett*. 2016;45:327-329.

18. Lubkowski J, Sonmez C, Smirnov SV, Anishkin A, Kotenko SV, Wlodawer A. Crystal structure of the labile complex of IL-24 with the extracellular domains of IL-22R1 and IL-20R2. *J Immunol*. 2018;201:2082-2093.

19. Kisseleva T, Bhattacharya S, Braunstein J, Schindler CW. Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*. 2002;285:1-24.

20. Wang C, Dong S, Zhang L, et al. Cell surface binding, uptaking and anticancer activity of L-K6, a lysine/leucine-rich peptide, on human breast cancer MCF-7 cells. *Sci Rep*. 2017;7:8293.

21. Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A*. 2005;102:6679-6685.

22. Mahnam K, Saffar B, Mobini-Dehkordi M, Fassihi A, Mohammadi A. Design of a novel metal binding peptide by molecular dynamics simulation to sequester Cu and Zn ions. *Res Pharm Sci*. 2014;9:69-82.

23. Wadood A, Ahmed N, Shah L, Ahmad A, Hassan H, Shams S. In-silico drug design: an approach which revolutionarised the drug discovery process. *Drug Des Deliv*. 2013;1:3.

24. Mohseni Moghadam Z, Halabian R, Sedighian H, Behzadi E, Amani J, Imani Fooladi AA. Designing and analyzing the structure of DT-STXB fusion protein as an anti-tumor agent: an in silico approach. *Iran J Pathol*. 2019;14:305-312.

25. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589.

26. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Methods*. 2015;12:7-8.

27. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010;5:725-738.

28. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9:40.

29. Du Z, Su H, Wang W, et al. The trRosetta server for fast and accurate protein structure prediction. *Nat Protoc*. 2021;16:5634-5651.

30. Mahtarin R, Islam S, Islam MJ, Ullah MO, Ali MA, Halim MA. Structure and dynamics of membrane protein in SARS-CoV-2. *J Biomol Struct Dyn*. 2022;40:4725-4738.

31. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253:164-170.

32. Pourhadi M, Jamalzade F, Jahanian-Najafabadi A, Shafiee F. Expression, purification, and cytotoxic evaluation of IL24-BR2 fusion protein. *Res Pharm Sci*. 2019;14:320-328.

33. Al-Refaei MA, Makki RM, Ali HM. Structure prediction of transferrin receptor protein 1 (TfR1) by homology modelling, docking, and molecular dynamics simulation studies. *Heliyon*. 2020;6:e03221.

34. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. *Proteomics Protoc Handb*. 2005:571-607.

35. Saha S, Raghava GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res*. 2006;34:W202-W209.

36. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.

37. Mustafa G, Mehmood R, Mahrosh HS, Mehmood K, Ahmed S. Investigation of plant antimicrobial peptides against selected pathogenic bacterial species using a peptide-protein docking approach. *BioMed Res Int*. 2022;2022: 1077814.

38. DeLano WL. Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*. 2002;40:82-92.

39. Laskowski RA. PDBsum new things. *Nucleic Acids Res*. 2009;37:D355-359.

40. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. *eLife*. 2015;4.

41. Toukmaji AY, Board JA Jr. Ewald summation techniques in perspective: a survey. *Comput Phys Commun*. 1996;95:73-92.

42. Zielkiewicz J. Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys*. 2005;123:104501.

43. Martyna GJ, Klein ML, Tuckerman M. Nosé-Hoover chains: the canonical ensemble via continuous dynamics. *J Chem Phys*. 1992;97:2635-2643.

44. Gul R, Hanif MU, Gul F, et al. Molecular cloning, expression, sequence characterization and structural insight of Bubalus bubalis growth hormone-receptor. *Mol Biotechnol*. 2022:1-14.

45. Rehman HM, Sajjad M, Ali MA, et al. Identification of NS2B-NS3 protease inhibitors for therapeutic application in ZIKV infection: a pharmacophore-based high-throughput virtual screening and MD simulations approaches. *Vaccines*. 2023;11:131.

46. Shinwari K, Rehman HM, Liu G, Bolkov MA, Tuzankina IA, Chereshnev VA. Novel disease-associated missense single-nucleotide polymorphisms variants predication by algorithms tools and molecular dynamics simulation of human TCIRG1 gene causing congenital neutropenia and osteopetrosis. *Front Mol Biosci*. 2022;9:879875.

47. Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep Methods*. 2021;1:100014.

48. Wang S-Q, Shi M, Fang L, Xu S-M, Wang C, Yu Z. Design of dual inhibitors of human TNF-α and IL-6 with potentials for the treatment of rheumatoid arthritis. *Trop J Pharm Res*. 2019;18:2305-2312.

49. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*. 2017;33:3098-3100.

50. Mandell JG, Roberts VA, Pique ME, et al. Protein docking using continuum electrostatics and geometric fit. *Protein Eng*. 2001;14:105-113.

51. Khan MT, Zeb MT, Ahsan H, et al. SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. *Arch Microbiol*. 2021;203:59-66.

52. Pons C, Grosdidier S, Solernou A, Pérez-Cano L, Fernández-Recio J. Present and future challenges and limitations in protein-protein docking. *Proteins*. 2010;78:95-108.

53. Idrees D, Kumar V. SARS-CoV-2 spike protein interactions with amyloidogenic proteins: potential clues to neurodegeneration. *Biochem Biophys Res Commun*. 2021;554:94-98.