Received: 2 August 2017 | Revised: 13 November 2017 | Accepted: 19 November 2017

DOI: 10.1002/jmr.2695

WILEY Journal of Molecular Recognition

RESEARCH ARTICLE

# Integration of molecular dynamics simulation and hotspot residues grafting for de novo scFv design against *Salmonella* Typhi TolC protein

Siew Wen Leong[1] | Theam Soon Lim[1] | Asma Ismail[2] | Yee Siew Choong[1]

[1] Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia, Minden, Penang, Malaysia

[2] Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia, Health Campus, Kubang Kerian, Kelantan, Malaysia

**Correspondence**
Yee Siew Choong, Institute for Research in Molecular Medicine (INFORMM), Universiti Sains Malaysia, Minden 11800, Penang, Malaysia.
Email: yeesiew@usm.my

## Abstract

With the development of de novo binders for protein targets from non-related scaffolds, many possibilities for therapeutics and diagnostics have been created. In this study, we described the use of de novo design approach to create single-chain fragment variable (scFv) for *Salmonella enterica* subspecies *enterica* serovar Typhi TolC protein. Typhoid fever is a global health concern in developing and underdeveloped countries. Rapid typhoid diagnostics will improve disease management and therapy. In this work, molecular dynamics simulation was first performed on a homology model of TolC protein in POPE membrane bilayer to obtain the central structure that was subsequently used as the target for scFv design. Potential hotspot residues capable of anchoring the binders to the target were identified by docking "disembodied" amino acid residues against TolC surface. Next, scFv scaffolds were selected from Protein Data Bank to harbor the computed hotspot residues. The hotspot residues were then incorporated into the scFv scaffold complementarity determining regions. The designs recapitulated binding energy, shape complementarity, and interface surface area of natural protein-antibody interfaces. This approach has yielded 5 designs with high binding affinity against TolC that may be beneficial for the future development of antigen-based detection agents for typhoid diagnostics.

KEYWORDS

de novo scFv design, hotspot residue grafting, molecular dynamics simulation, *Salmonella enterica* ser. Typhi TolC protein

## 1 | INTRODUCTION

Computer-aided molecular design has become a powerful tool with the progress in structural bioinformatics. Because antibody is a prominent tool in diagnostics and next generation protein therapeutics,[1,2] the applications of computer-aided antibody designs have also been reported.[3-11] In this work, we described the use of molecular dynamics (MD) simulation and de novo hotspot residues grafting in the invention of single chain fragment variable (scFv) specific against TolC protein from *Salmonella enterica* subsp. *enterica* ser. Typhi (*S*. Typhi).

*S*. Typhi is the causative agent for typhoid fever. This enteric fever is a global health burden especially prevalent in South and East Asia, India, South Africa, and South America. A World Health Organization study estimated the incidence and mortality rate to be 21.7 million and 217 000 cases annually.[12,13] Typhoid fever outbreaks in Uganda,[14] North India,[15] South India,[16] Songkhla,[17] and Malawi-Mozambique border[18] in recent years suggested that the disease has yet to be eradicated. *S*. Typhi infection usually occurs via contaminated food and water, implying that supply of clean water and sewage disposal would be the best prevention. World Health Organization has recommended typhoid vaccination to control the disease, as well as health education, improvement of water quality and sanitation, and training health professionals in diagnosis and treatment.[19] However, provisions of such facilities in developing countries were often hindered by budget limitations, rapidly increasing populations and uneven development.

Detection and isolation of *S*. Typhi from blood, urine, and stool are the definitive investigation to confirm typhoid infection, with the bone marrow culture being the gold standard of diagnosis methods.[20] The Widal test which detects IgM and IgG to the O and H antigens to *Salmonella* is widely used due to its low cost but is less reliable due to the sharing of O and H antigens by other *Salmonella* serotypes and also other members of the Enterobacteriaceae family.[21] The major hindrance in relying on the isolation of the bacteria is the need for

  

laboratory equipment and trained personnel which is not feasible due to limited health care facilities in developing countries. Rapid diagnostic kits, for example Multi-Test Dip-S-Ticks (PANBIO INDX, Inc., Baltimore, Md.), TUBEX® (IDL Biotech, Sollentuna, Sweden), and Typhidot® (Malaysian Biodiagnostic Research Sdn. Bhd., Malaysia), have emerged as viable alternatives for diagnostic modality particularly in rural areas. These tests detect for the presence of antibodies in patients' serum or blood in response to anti-O, anti-H, anti-Vi, a specific 50-kDa antigen and the O9 antigen, respectively. Typhidot® was reported to have a specificity of 80%–92% in Malaysia and 60%–80% in the African region, while TUBEX® has a specificity of 69%–88% and a sensitivity of 51%–73%. Antigen-based tests may aid in early detection and the carrier state of the pathogen. Works detailing highly specific monoclonal antibodies targeting the antigens of *S*. Typhi, as well as other serotypes of the pathogen, have been published.[22,23] While some of these works has been developed into simple genus-specific, species-specific, or serotype-specific test kits, they were produced locally in small quantities and no commercial manufacturers have made them widely available.

The study of antigens and immunity of *S*. Typhi would be beneficial for effective and efficient control of the disease. Presently, lipopolysaccharides, capsular antigens, enterobacterial common antigen, flagella, pili or fimbriae, and outer membrane proteins have been investigated for their role in *S*. Typhi infection. Among these, the 50-kDa outer membrane protein which was antigenically specific to *S*. Typhi was subsequently developed into a dot enzyme immunosorbent assay, namely Typhidot® for the detection of IgM and IgG antibodies against the antigen in typhoid patients.[91,92] Comparative modeling showed that this 50-kDa protein is a homolog of the *Escherichia coli* TolC protein,[24] and structural elucidation by crystallography has further confirmed the *S*. Typhi TolC protein is a trimeric channel ~140 Å in length. Previous modeling study showed that the *S*. Typhi TolC protein has a transmembrane domain with 12-stranded β-barrel architecture embedded in the outer membrane and extracellular loops which protrude from the cell (Supporting Information Fig. S1). The equatorial domain has a mixture of α/β-domain, and finally the α-helices form a tunnel which extend 100 Å into the periplasmic space.[24] The TolC protein family functions in tandem with several transport systems to achieve uptake and efflux of diverse molecules.[25]

The ability to manipulate protein function by working on pre-existing structures and designing novel interactions offers promising advances for applications in medicine, research, and industries. De novo computational protein design has been made possible by advances in computational power. One of the approaches in de novo design involved redesigning a scaffold anchored by a hotspot to the target surface. This approach was exemplified by the work describing the design of a high-affinity binder to the conserved stem region on influenza hemagglutinin.[26] Two designed proteins, HB36 and HB80, bind H1 and H5 hemagglutinin with low nanomolar affinity after affinity maturation. The approach can be widely applied to design a binder targeting any desired surface.

In this work, Fleishman's methods were applied to design scFv binder against the extracellular loops of *S*. Typhi TolC protein. Five designs which showed high possibility of forming native-like protein-protein interface with the TolC protein were selected. It was hoped that the designed scFvs may find application in future antigen-based *S*. Typhi diagnostic for typhoid fever.

## 2 | METHODS

### 2.1 | Molecular dynamics simulation

#### 2.1.1 | System setup

The system setup applied the general methodology for membrane protein simulations.[27,28] The starting structure was a model of *S*. Typhi TolC protein obtained from comparative modeling.[24] A model lipid bilayer consisting of POPE lipids was used to mimic the outer membrane environment where *S*. Typhi TolC protein resided. POPE lipids were chosen on the basis of being the most abundant lipid species in the outer membrane Gram-negative bacteria.[29] A CHARMM36 membrane patch which was pre-equilibrated with 40 ns of simulation and consisted of POPE lipids was obtained from http://terpconnect.umd.edu/~jbklauda/research/download.html. This small membrane patch consisting of 80 lipids was duplicated in the x and y direction to make a 180 Å × 180 Å patch in order to accommodate the protein. As the β-barrel was the only part of the protein residing within the membrane, the centre of mass of β-barrel residues was aligned with the centre of mass of the lipid molecules in the bilayer with VMD.[30] The system was subsequently solvated with Gromacs program genbox. Ions were added with Gromacs program genion to obtain a total neutral charge at 0.1 M KCl. The entire system consisted of 526 753 atoms, which included 857 lipids, 132 520 TIP3P water molecules, and 729 counterions.

### 2.2 | Simulation protocol

All simulations including the minimization, equilibration, and production runs were performed with Gromacs version 4.5.4.[31] The CHARMM22 force field augmented with CMAP correction was employed to treat the protein[32,33] whereas CHARMM36 parameters were used for the lipids component.[34] Minimization was performed to remove steric constraints in the system until the maximum force in the system is less than 1000 kJ/mol/nm. An initial 200-ps temperature equilibration with protein and lipid phosphate atoms restrained was performed in NVT ensemble using modified Berendsen thermostat[35] to stabilize the temperature at 310.15 K. This was followed by a 1-ns run in NPT ensemble to equilibrate the pressure at 1.01 bar using the Parrinello-Rahman method.[36] Restraints on lipid phosphate atoms were also relaxed, and temperature coupling was controlled with Nose-Hoover method.[37,38] Finally, a 20-ns production in the NPT ensemble was performed. Analysis was based on the last 15 ns trajectory of the production simulation. In all equilibrations and production simulations, long-range electrostatics was treated with PME. A switch function was used to evaluate the Lennard-Jones potential from 8 to 12 Å. Real-space sum were cut off at 12 Å. LINCS algorithm[39] was used to constrain all bonds and to allow a time step of 2 fs. Simulations were performed on a cluster of 8 Intel® Xeon® X3360 2.8-GHz processors with an average performance of 169.5 h/ns.

## 2.3 | Clustering analysis

Clustering analysis was performed on the last 15 ns of the trajectory, utilizing 15 000 trajectory snapshots of the protein. Root mean square deviation (RMSD)-based clustering was performed with 1-Å cutoff using the single linkage and gromos algorithm as implemented in Gromacs. To minimize variability caused by the flexible C-terminal region, the structures were least square fitted to non C-terminal $C_\alpha$ atoms of the starting conformation. With the single linkage method, a conformation was considered to be a member of a cluster if its distance to any element of the cluster was less than the cutoff value. Meanwhile, the GROMOS algorithm has been described by Daura and colleagues,[40] whereby the RMSD of all pairs of conformations using the cutoff was first determined. The conformation with the highest number of neighbors is to be the center of the cluster and formed the first cluster with all its neighbors. The process was repeated until the pool of conformations was empty and a series of non-overlapping cluster of conformations were obtained.

## 2.4 | De novo design

### 2.4.1 | De novo hotspot docking

Hotspot-centric de novo design method has the advantage to design binders for any target structure without pre-requisite knowledge about the target's binding site. In this case, scFv was targeted towards the exposed extracellular loops on *S*. Typhi TolC protein. This is due to only the extracellular loops are protruded from the cell membrane and exposed to the extracellular environment, therefore able to present to binders.

Potential hotspots that can form high affinity interactions with the target were determined with exhaustive docking of "disembodied" amino acids against extracellular loops of *S*. Typhi TolC protein. Phe, Trp, Tyr, His, and Ile were typical residues which formed high affinity interactions across protein-protein interface, ie, hotspot and were also found with high propensity on the antibody paratope containing surface.[41] The previously mentioned residues as well as Arg, Glu, Met, Asn, Asp, Val, Leu, Thr, Ala, Ser, Gln, and Lys were docked against the surface exposed extracellular loops of TolC to find high affinity interactions which may be incorporated into scFv design. The starting structure for these residues was obtained from http://www.aminoacidsguide.com (accessed 31 January 2016). A global docking was first conducted with ZDock[42] after which the top 50 predictions were refined with Rosetta.[43] Local refinement from ZDock outputs were conducted in Rosetta by randomly perturbing the residue with 3-Å translation and 8° rotation from the starting structure. From each of the 50 predictions, 100 new dockings were generated, yielding 5000 dockings from each residue. The dockings were ranked based on binding energy (Rosetta Energy Unit, R.E.U). This energy as calculated by the scoring function in Rosetta was derived empirically from observed geometries of a subset of proteins in the PDB, including the radius of gyration, packing density, distance/angle between hydrogen bonds, and distance between 2 polar atoms, which are converted into an energy function through Bayesian statistics.[44-46] Even though R.E.U is an arbitrary unit as Rosetta score function is a combination of physics and statistics-based potentials, some studies estimate R.E.U to be between 0.5 and 1.0 kcal/mol.[47-49] Dockings with binding energy within the top 1.0% were selected as potential hotspot residue.

## 2.5 | Scaffold search for scFv design

The RCSB PDB was searched for crystal structures of scFv as scaffolds for design. The scaffolds were required to have a maximum of 2.5-Å resolution[50] without missing residues in the complementarity determining regions (CDRs) and originated from a human source. Based on these criteria, the selected scaffolds were PDB ID: 1JV5,[51] 1X9Q,[52] 2A9M,[53] 2D7T,[54] 2GHW,[55] 2YC1,[56] 2UZI,[57] and 3JUY.[58]

## 2.6 | Shape complementarity docking of scFv scaffold with TolC protein

Shape complementarity between antibody and target was achieved by docking scFv scaffolds to TolC protein with the shape complementarity-based docking program PatchDock.[59] The receptor active site was constrained to residues on the extracellular loops on TolC protein while scFv CDRs were set as the ligand active site. The dockings were run on parameters for antibody-antigen complex to restrict the search space to CDRs. Top 50 docking conformations from each scaffold were carried over to subsequent steps.

## 2.7 | De novo design

The de novo designs were run on batches of 5000 design trails on performed on a cluster of 8 Intel® Xeon® X3360 2.8 GHz processors with an average performance of 140 h/run on single core. The de novo design was performed according to the general protocol from Fleishman and colleagues ([60]) where 2 hotspot residues from the hotspot residue library were placed onto scFv scaffold. Protocol for running this method was available in Rosetta Design Suite using the RosettaScripts module. A total of 200 design trials were attempted for every configuration of TolC-scFv scaffold generated from the shape complementarity docking. In total, there were 10 000 design trials for each scaffold. Hotspot residues were incorporated into CDR of respective scaffold. The CDR were determined by the Paratome online webserver using the scaffold's 3D structure input (http://ofranservices.biu.ac.il/site/services/paratome/index.html).[61] Only residues in the CDR were allowed conformational flexibility and mutation.

During the design, the hotspot residue library was used to restrain the backbone configuration of the scaffold to accommodate as many hotspot residues as possible by means of SetupHotspotConstraints function in RosettaScripts. A coarse-grained centroid level docking was first performed on the target protein-scFv scaffold input to refine the shape complementary pose from the shape-complementarity docking step to a more energetically favorable configuration.

Next, 2 residues from the hotspot library were placed simultaneously onto the scaffold. For a given scaffold position, the scaffold were translated and rotated to align it with the hotspot's position. To maintain the initial shape complementarity between the scaffold and the target, scaffold position with $C_\beta$ atom >4.0 Å from the hotspot residue or C-N and $C_\alpha$-$C_\beta$ vectors that were misaligned with the hotspot residue by 60° were triaged from the design trajectory. The rigid body

orientation and the placed hotspot side chain were minimized in a reduced force field.[60]

Placement of the hotspot was considered successful if its energy is < 1.0 R.E.U. Following successful hotspot placement, scaffold CDR were redesigned using RosettaDesign. The target protein side chains were allowed to repack, allowing for conformational changes that normally accompanied complexation (induced fit). The redesign and minimization steps were repeated 3 times to increase the possibility in getting higher affinity interactions. However, the mutated residues were reverted to the original amino acid identity in the scaffold if their contribution to the binding energy was less than 2.5 R.E.U with revert_to_native application. Resulting models were filtered by binding energy and buried surface area whereby complex with binding energy > −10.0 R.E.U and buried surface area < 800 Å$^2$ were ignored.

## 2.8 | Analysis

Shape complementarity was computed using Sc parameter.[62] Hydrogen bonding, salt bridges, and other intermolecular interactions were calculated and visualized with Accelrys Discovery Studio.[63] Visualization was performed with Accelrys Discovery Studio[63] and VMD 1.9.[30] The work flow is summarized in Supporting Information Figure S2.

# 3 | RESULTS

## 3.1 | The TolC protein stability, flexibility, and dynamics

Root mean square deviation (RMSD) of TolC protein during simulation was calculated to measure the overall drift from its starting structure. RMSD of the protein Cα atoms increased from ~2 to ~5 Å after an initial 5 ns (Figure 1A). The RMSD value remained stable thereafter. To further investigate how each region contributes to the structural drift, RMSD was calculated for the TolC protein transmembrane region (TM), periplasmic region, and extracellular loops (Figure 1A). RMSD of the TM region fluctuated between 1 and 1.5 Å throughout the entire simulation, showing strong stability. Meanwhile, the extracellular loops showed more flexibility with RMSD of 2–4 Å along the simulation. An increase of the RMSD was observed at ~ 7-ns mark, indicating possible interactions among the extracellular loops (Supporting Information Figure S3). From Figure 1A, it can be deduced that TolC protein was mostly influenced by conformational drift from the periplasmic region. RMSD profile of the protein was similar with that of the periplasmic region. To confirm this observation, RMSD of the C-terminal 43 residues were calculated and compared with the other region of

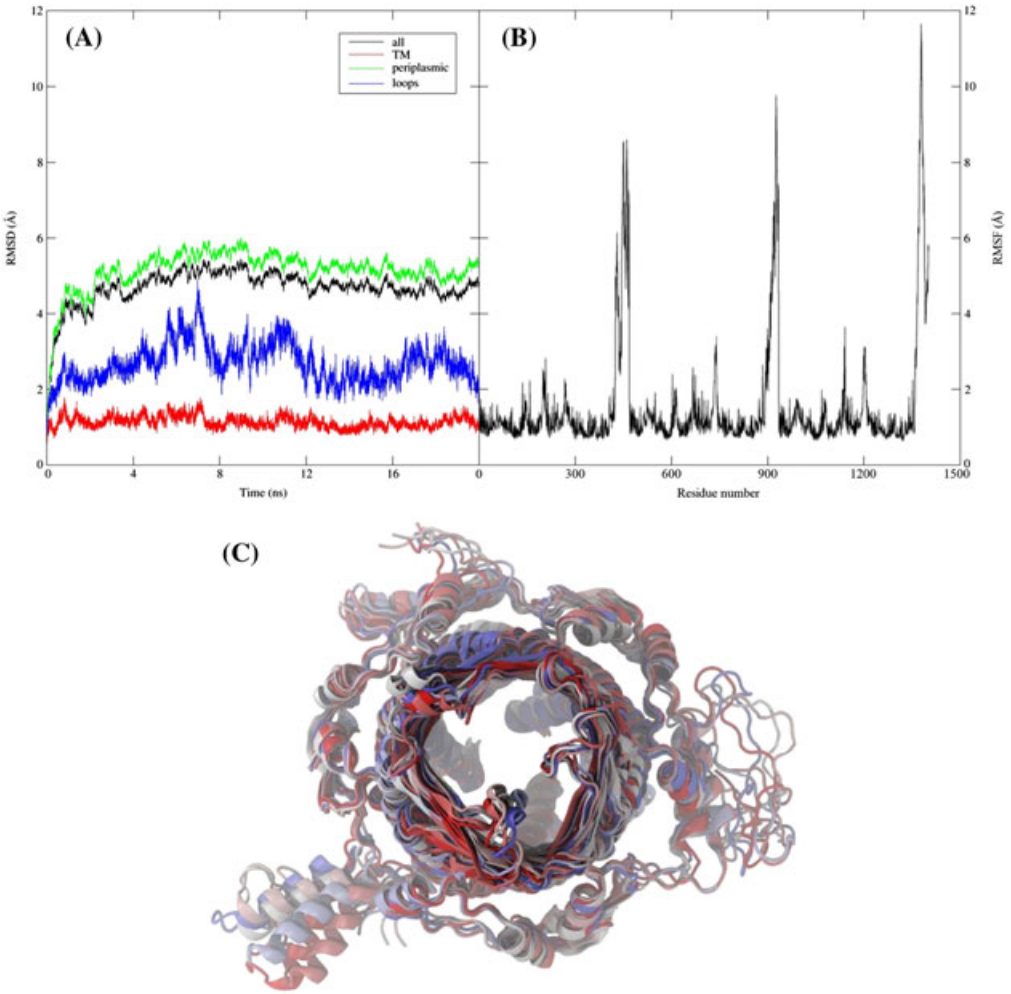


**FIGURE 1** Molecular dynamics simulation of *S*. Typhi TolC protein in POPE. (A) Root mean square (RMSD) of TolC as a whole protein, transmembrane region (TM), periplasmic region and extracellular loops. (B) Individual residue root means square fluctuation (RMSF) of TolC protein. (C) Superimposition of the central structures of the 6 largest clusters (from red to white to blue ribbon presentation)

TolC. In Figure 1A, RMSD of the C-terminal 43 residues reached a high of 15 Å whereas the rest of the residues remained stable at 2-Å deviations. The high RMSD of the periplasmic region was not observed in prior simulations of TolC.[64] This was actually due to the highly mobile C-terminal 43 residues which were truncated in the crystallization of *E coli* TolC.[65] The C-terminal residues was eventually modeled and included in the *S*. Typhi TolC homology structure.[24]

Individual residue flexibility of TolC protein was shown in root mean square fluctuation (RMSF) plot (Figure 1B). A repeating pattern may be found between residue 1 and 467, 468 and 934, and 935 and 1401 as TolC protein consists of 3 identical chains. The peaks with the highest RMSF (8–12 Å) occur on the C-terminal residues, complementing the observation on the periplasmic region from RMSD plot. The peaks with RMSF of 2–3 Å corresponded to extracellular loops on TolC protein. The RMSF plot suggested that residues on the C-terminal region and the extracellular loops have high mobility, as also evidenced from RMSD plot (Figure 1A). This was to be expected considering residues in these regions were exposed to the outer environment and subject to instantaneous interactions with water molecules. In addition, residues in the loops structures were not highly ordered as in α-helix or the β-sheet motif.

Using the single linkage clustering method, the trajectory clustered into a single cluster with the protein conformation at 13.12 ns emerging as the central structure. Table 1 shows the 6 largest clusters found by the GROMOS algorithm. The GROMOS algorithm found agreement with the single linkage method, as the conformation at 13.12 ns appeared to be the central structure of the 7262 conformations in the largest cluster. Superimposition of central structures obtained from the clustering analysis reiterated the relative plasticity of the β-barrel and extracellular loops (Figure 1C). Extracellular loops of representative structures from the MD simulation occupied consistent regions of conformational space.

## 3.2 | De novo hotspot docking and scFv design

A total of 229 residues which has the top 1.0% of binding energy to TolC were identified. These hotspots were made up of Asp, Glu, His, Phe, Ser, Thr, Trp, and Tyr. From 10 000 design trials for each scFv scaffold, 170 designs satisfied the initial constraints of < −10.0 R.E.U and with buried surface area of >800 Å$^2$. Prior to designs evaluation, each mutant was tested with the original amino acid in the scaffold, and should the change in binding energy was less than 2.5 R.E.U in favor of the mutant, the residue was reverted to the wild type residue. Following that, the designs were evaluated by their shape complementarity, hydrogen bonding residues in the interface, and other non-bonded intermolecular interactions visualized in Accelrys Discovery Studio.[63]

**TABLE 1** Six largest groups from the clustering of *S*. Typhi TolC protein from MD trajectory using GROMOS method

| Cluster | No. of Conformations | Central Conformation, ns |
|---|---|---|
| 1 | 7262 | 13.12 |
| 2 | 2687 | 16.68 |
| 3 | 2474 | 5.97 |
| 4 | 704 | 6.23 |
| 5 | 479 | 19.92 |
| 6 | 432 | 10.83 |

## 3.3 | Scaffolds

There were 170 successful designs from 8 scaffolds, with 149 designs based on scaffold 3JUY. Scaffold 2YC1 managed to produce 16 designs, while 2A9M and 1X9Q each produced 2 designs, and 1 design resulted from scaffold 2D7T. Scaffold 1JV5, 2GHW, and 2UZI did not produce any successful design. Based on a total of 80 000 design trials at the initial stage, only 0.2% trials produced positive results. This thus underscored the importance of starting with a diverse set of scaffolds and increasing the number of trials in order to obtain meaningful set of designs for further discrimination.

## 3.4 | General characteristics of the design

The initial designed structures produced complexes with favorable binding energies in the range of −30 to −10 R.E.U. The average predicted binding energy from the 170 scFv-TolC complexes were (−15.75 ± 4.0 R.E.U). These designs also have reasonable buried surface area of 850–2400 Å$^2$, with an average value of 1432.71 ± 329.51 Å$^2$. These results are comparable with the reported analyses on antibody antigen structures which estimated the buried surface area of antigen-antibody complexes at 2071 ± 456 Å$^2$,[41] ~1400 to 2300 Å$^2$,[66] and 1646 ± 442 Å$^2$.[67] The Sc parameter is a measurement of the interface shape complementarity between the designed scFv and TolC. A total of 66 scFv-TolC complexes were found to have a shape complementarity of less than 0.5, whereas the average value was 0.53 ± 0.09. Only 5 designs recorded Sc value of >0.7, in line with studies which claimed that antibody-antigen complexes have imperfect shape complementarity.[62]

Upon revert mutation, it was found that many of the original designs lost their favorable binding energies with *S*. Typhi TolC protein. After the introduction of revert mutations, only 110 designed complexes have a predicted binding energy of < −10 R.E.U. This shows that while some mutations did not contribute greatly to the binding energy, they could be important in maintaining a favorable conformation for hotspot residues to make interactions across the antibody-antigen interface. However, the revert mutations have little effects on the calculated buried surface area of the interfaces. The designed structures retained a buried surface area of 1453.53 ± 316.56 Å$^2$. The designs with revert mutations also maintained the Sc value of 0.53 ± 0.08.

Further analyses were performed with the designs augmented with the revert mutations. This was because of considerations to minimize the number of mutations from the initial scaffold. More importantly, the revert mutations were meant to maintain the stability of the scaffold and ensure that minimal changes were to be made to initial CDR conformation so that the CDR loops will fold into the appropriate conformation upon expression and binding at experimental level.

### 3.5 | Test cases of natural antibody-protein complexes

To further discriminate designs which exhibit native like interfaces, a set of test cases was prepared with solved protein-antibody complexes available in PDB. The set of test cases consisted of 19 complexes (Supporting Information Table S1) with non-redundant antibody-large protein antigens dataset.[41] Properties of the 19 solved protein-antibody complexes were calculated and compared with the designed complexes. The average binding energy was −21.30 R.E.U while Sc value was 0.68. The protein-antibody complexes also buried 1696.73 Å$^2$ surface area. Based on these values, scFv designs with properties which were less than 60% of the test case average value were eliminated from further characterization.

### 3.6 | The properties of selected scFv designs

A total of 74 scFv designs passed the previously mentioned additional screening. A methodology to arrive at a small set of designs worthy of experimental characterization was needed. One of the suggestions would be to select designs with the lowest energy after relaxation or to select lowest energy design based on each scaffold.[68] The best design can also be selected based on the virtue of having the best binding energy to the target. The curated designs which passed the filters during design simulation and additional requirements are listed in Supporting Information Table S2. The list was ranked by the lowest binding energy after revert mutation. From Supporting Information Table S2 it was clear that these designs have comparable energy values. Most designs with the best binding energies originated from scaffold 3JUY. However, this may present a problem for future experimental characterization because if difficulties arise from protein expression and folding, designs with similar sequence and structures are likely to be affected. Therefore, it was reasonable that the designs be chosen based on different scaffolds and orientations be advanced for subsequent analysis.

We have selected and analyzed the top 5 designs from 4 different scaffolds (Figure 2). The selected designs were considered to have the best characteristics based on shape complementarity, binding energy, and presence of favorable interactions in the interface. The designs were named using the following convention; the first 4 letters referred to the scaffold from which the design originates from, followed by a 2-digit number which indicated the docking number between TolC and scFv scaffold, and finally a 4-digit code which referred to the number of trial run. For residue-residue interactions, residues on chain A resided on TolC protein, while residues on chain B resided on the scFv (Note: This designation was meant to facilitate ease of understanding while reading, the TolC functional protein is consisted of 3 identical chains). The *S.* Typhi TolC protein targeted binding interface was largely polar as it exposed to the external environment. Hence, most of the interactions within the designed binding interfaces were dominated by hydrogen bonds and salt bridge interactions. Amino acid sequences of the selected structures were also available in Supporting Information Table S3.

Here, we described the best design based on 3JUY scaffold which passed all the screening. The 3JUY protein is a variant of b12 antibody targeting HIV-1 gp120 envelope protein.[58] Design 3juy-18-0119 has the second best binding energy (−24.08 R.E.U) among the 74 designs which passed all the screening. The calculated interface surface area was 2430.08 Å$^2$, which is highest among the 74 designs while the Sc score was 0.55. The design with the best binding energy, 3juy-18-0004, was not chosen due to its lower Sc score and interface surface area than 3juy-18-0119. Besides, 3juy-18-0119 also made for an interesting case with its variety of interactions (Supporting Information Table S4 and Figure S4). Hydrogen bonding was observed in A:Lys265-B:Thr1503, A:Thr266-B:His1456 and B:Ser1506-A:Gly737. In addition, Arg and Asn also made up for a number of hydrogen bonds, including A:Ser268-B:Asn1512, A:Asn1201-B:Gly1427, B:Arg1432-A:Ala269, B:Asn1453-A:Thr266, B:Asn1453-A:Asn267, B:Arg1581-A:Ala736 and B:Arg1584-A:Asn1201. In addition, pi-pi stacking interactions were formed by aromatic amino acid residues in the interface, namely A:Tyr738-B:Trp1505 and A:Tyr271-B:Tyr1503. Three pi-alkyl interactions were found in the scFv-TolC interface, namely B:Phe1433-A:Ala269, B:His1502-A:Ala269, and B:His1502-A:Ala1203. Descriptions for the other 4 designs may be found in the supplementary discussion. (Further description on another 4 selected designs can be found in Supplementary Information).

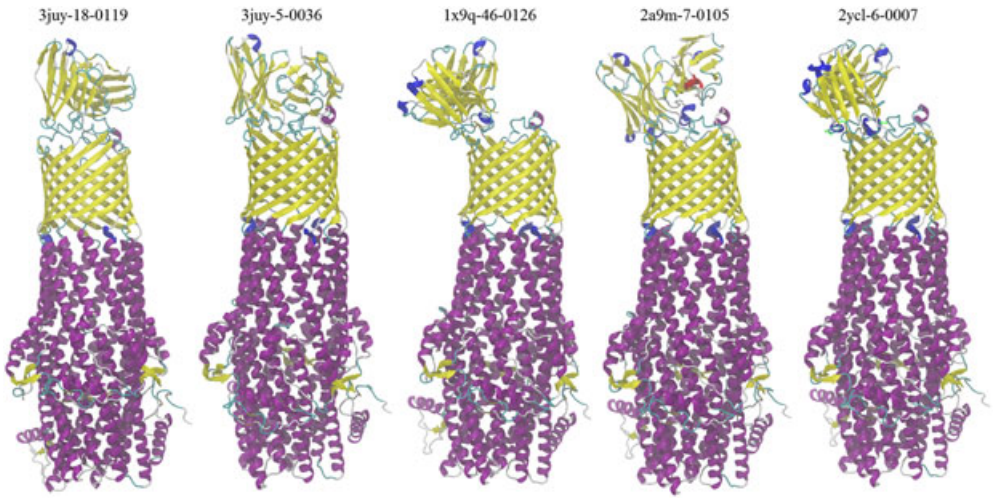


**FIGURE 2** The 5 selected scFv designs (ribbon representation), namely 3juy-18-0119, 3juy-5-0036, 1x9q-46-0126, 2a9m-7-0105, and 2yc1-6-0007; and their orientations with respect to *S.* Typhi TolC protein (ribbon representation)

## 4 | DISCUSSION

MD simulations have been reported to refine protein structures obtained from comparative modeling[69] and crystal structures.[70] In this study, MD simulation was used to refine the comparative model of *S.* Typhi TolC protein in explicit solvent and lipid bilayer model. The templates for *S.* Typhi TolC protein were obtained from X-ray crystal structures of *E coli* TolC protein, which were lacking in the C-terminal 43 residues.[65] Although the lack of these residues did not affect the functional capabilities of the TolC protein, they may be important in the dynamics of the protein and thus need to be included to provide a more comprehensive structural profile of the protein.

The MD simulation included the C-terminal 43 residues which have not been modeled in prior *E coli* TolC protein simulations. Throughout the simulation, *S.* Typhi TolC protein has been found to be relatively rigid with most of the flexibility arising from the C-terminal 43 residues. The β-helices and β-barrel proteins such as TolC TM region were recognized to impose greater restraint and rigidity when embedded in a lipid membrane compared with α-helices proteins. This may contribute towards the TM region's low RMSD profile.[71] In this simulation, the extracellular loops have been found to be relatively flexible compared with the other regions of the protein. In a simulation of *E coli* TolC protein, the extracellular loops were found to open and close spontaneously.[72] Similar observations were also noted from the simulation of *E coli* TolC and *Pseudomonas aeruginosa* OprM protein.[64]

However, as TolC protein and its homologs were associated with other accessory proteins such as AcrA and AcrB to form a working channel for drug transportation, the full dynamics of the protein has yet to be explored.[25] Nevertheless, by means of coarse-grained elastic network analysis (CG-NMA), the conformational couplings between AcrAB accessory proteins and TolC were investigated.[73] Concerted motions were noted between AcrB and TolC not only limited to the contact interface, but remarkably also between distinct parts in both proteins, where motions in the top part of transmembrane domain of AcrB correlate with the top part of TolC. This suggested that energy of proton movement across the inner membrane can drive both the conformational switching of AcrB trimer and the extracellular gate of TolC.[73] The addition of AcrA did not change the correlation map between TolC and AcrB, but it strengthened the coupling between the 2 proteins.[73]

A representative structure was selected from the MD simulation to be the target structure for subsequent binder design. The energy minimized average structure, ie, the representative structure may then be used to design inhibitors or identify binding sites.[74] Additionally, the use of average equilibrated structure as a representative structure from simulation has been utilized to suggest an open conformation for Cx26 hemichannel.[70] The average equilibrated structure may be obtained from selecting the trajectory which has the lowest RMSD to the average structure, thereby producing a conformation that is free from spurious physical effects. In other cases, central structures from clustering algorithm were utilized for downstream applications such as virtual screening[75] or discovering active sites in proteins.[76] Here, it was identified that the central structure from clustering may best represent the most dominant structure of *S.* Typhi TolC protein. Clustering using single-linkage and GROMOS methods converged upon a single frame as the cluster centre for the largest cluster. The cluster centre structure also closely resembled the average structure with a RMSD value of 1.96 Å. Therefore, the cluster centre structure was the most dominant structure of the simulation and thus selected as the most dominant conformation for *S.* Typhi TolC protein.

The hotspot approach was based on the idea that a pre-computed hotspot may act as the anchor of protein-protein interactions.[77] The hotspot served as an initial seed for, ie, scFv-TolC interface. In most cases, the designs incorporated the hotspot residues obtained from the hotspot library. As in 2a9m-7-0105, His1565 and Asp1615 were the pre-computed hotspot predicted to have high binding affinity against TolC protein extracellular loops. After incorporation into the scFv CDR backbone, His1565 and Asp1615 retained their optimal residue configuration and were predicted to make hydrogen bonds with the TolC protein at Thr1200 and Tyr738, respectively. In other cases, such as 2yc1-6-0007, the pre-computed hotspot lost their initial high affinity binding to the TolC protein. Even though the hotspots were integrated into the design, they were eventually reverted back to the wild type residue as the predicted binding energy contributed was below the threshold value. These observations suggested an area for improvement in computing the initial hotspot and the methods by which they can be incorporated into the scaffold without compromising their high affinity configuration. For targets with known binders, this seems to be a straightforward as hotspot residues from the known binder can be grafted onto a new scaffold. Fleishman and colleagues generated inverse rotamers for residues which made very favorable and geometrically constrained interactions with the target surface.[60] The de novo hotspot docking protocol could be improved to identify configurations which may form high affinity interactions with the target. Rather than a shotgun approach which uses all residue identities in the initial hotspot generation, identification of a small set of optimal configuration hotspots to be incorporated into a more diverse set of scaffold may be a more efficient strategy.

It was observed that there are some unfavorable interactions (eg, donor-donor interaction, negative-negative charge interaction, hydrogen acceptor-acceptor interaction) in the 5 selected scFv designs. However, with the presence of other strong favorable interactions (eg, hydrogen bonding, salt bridges, pi-anion, pi-alkyl, and pi-pi stacking), it was believed that these binders can outperform others. It is believed that further optimization to remove these unfavorable interactions could improve their binding affinities towards the *S.* Typhi TolC protein.

The introduction of flexible backbone design has allowed the creation of proteins with novel folds and sequences. Kuhlman and co-workers pioneered a broad-based flexible backbone remodeling protocol by iterating between fragment-based structure prediction and sequence design.[78] Their method produced a 93-residue protein with a novel α/β fold which has 1.2 Å of RMSD from its crystal structure. Alteration to enzyme specificity has also been achieved by placing new anchor residues.[79] The anchor residue was first placed at position that interacted with the transition state of the new substrate and then searched for backbone conformations capable of hosting these residues by de novo protein structure prediction. Subsequent crystal structure revealed excellent agreement (~1 Å) with the designed structure.[80] Despite these successes, provision for flexibility was seldom

included in protein design because searching the entire sequence and conformational space available may impose a too high computational cost. A semblance of flexibility was included in the design protocol via the BackrubDD mover and allowing CDR residues to repack and minimize. Additional flexibility during the hotspot insertion and the design process may also aid in better positioning of the hotspot against the target surface.

Much of the successfully designed structures which passed all the screenings emerged from a single scaffold, namely 3JUY. The low success rate with the other scaffolds might indicate that they were not in an optimal configuration to harbor hotspot residues. Generating a variety of starting conformations and using a diverse set of scaffolds could be beneficial. Studies showed that the de novo binding pair of Prb and Pdar was designed with a set of 37 structurally diverse thermostable proteins[50] while the hemagglutinin binder was searched from a set of 865 proteins.[60] The problem may also be relieved with the inclusion of conformational flexibility which may allow for more sequence diversity in the designed structure.[81]

Due to the many possible applications for protein design, several automated programs for designing proteins with the desired characteristics were developed. Automated approaches offer easy solution for design purposes and may be employed for the design of scFv for *S*. Typhi TolC protein as well. Notable design programs include Protein WISDOM,[82] RosettaDesign,[83] and OptCDR.[84] The Protein WISDOM online tool uses a structural model as an input. In general, the design process can be divided into 2 steps. The first step involved a rank-ordered list of sequences which can fold into a given template structure prior to validation step which involved calculating fold specificity and/or binding affinity.[82] Protein WISDOM has been employed to design peptide inhibitors for HIV-1 entry to CD4[85] and Bak inhibitors of Bcl-$x_L$ and Bcl-2.[86] Meanwhile, the RosettaDesign server searched for sequences which can pack well and satisfy the hydrogen bonding potential of polar atoms for any given target structure or complex.[83] Starting from a random sequence, fixed backbone design was performed, and low energy sequences were obtained using Monte Carlo optimization with simulated annealing. RosettaDesign has been employed successfully to design γD crystallin mutants that were aggregation resistant,[87] stabilize residues with high thermal fluctuations in Bcx protein,[88] develop thermostable CalB lipase enzyme,[89] and interface optimization of a cytochrome $cb_{562}$ variant for Zn directed protein self-assembly.[90] RosettaDesign has been implemented as part of the Rosetta suite and is also available as an online server. On the end of antibody design, OptCDR was developed to design the binding portion of antibody to any target epitope. OptCDR first selects suitable canonical CDR structures to bind the epitope, followed by sequence selection and simultaneous refinement of the sequence and structure of the CDR before finally generating a library of CDRs which is expected to bind the target epitope.[84] Till date, the success of OptCDR has yet to be demonstrated through experiment.

Despite the promising front from such automated design approaches, many works have demonstrated that human intuition and expertise were often needed to correct and improve solutions presented by the programs. Nonetheless, the use of design servers may be beneficial to identify initial binding conformations or search for alternative poses before a particular structure is advanced to more rigorous design or evaluation.

## 5 | CONCLUSIONS

This work applied MD simulation and hotspot residues grafting in order to design scFv specific against *S*. Typhi TolC protein. The dynamics of *S*. Typhi TolC protein including the C-terminal 43 residues which has not been previously characterized in simulation was first explored. Subsequently, the most dominant structure of TolC protein was identified. The scFv design was based on de novo design method using hotspot library. From the MD simulation, the TolC protein showed intrinsic flexibility in the extracellular loops and C-terminal regions, while it was relatively rigid and stable in the β-barrel and periplasmic helices. The most dominant structure from the MD simulation was obtained by calculating the central structure from clustering algorithm. Clustering was performed on trajectory of the last 15 ns of the MD simulation, and the single-linkage and GROMOS algorithms arrived at the structure at 13.12 ns as the central structure. Through de novo hotspot-based design, novel scFv designs against *S*. Typhi TolC protein were generated. After a series of screening and selection, we presented 5 scFv designs from different scaffolds with predicted high binding affinity, shape complementarity, and a network of favorable interactions for TolC protein. Analysis also showed that these scFvs have comparable packing and buried surface area in the interface with that of antibody-protein crystal structures. This works showed that the integration of MD simulation and hotspot residue grafting is able to produce desired scFv designs specific for *S*. Typhi TolC protein at computational level. These scFv designs could be useful in future development for possible application in *S*. Typhi TolC protein detection. The foreseeable future work would entail synthesizing and characterizing the scFv designs at in vitro level. The scFv designs can be synthesized by oligo assembly or mutagenesis of the original sequence. Subsequently, affinity maturation and binding specificity improvement could be undertaken via directed evolution or other mutational analysis for designed antibodies which showed good experimental binding against the TolC protein. Coupling iterations of computational design with mutagenesis may serve as a strategy for further binding characteristics improvement.

## ACKNOWLEDGEMENT

This work was funded by the Universiti Sains Malaysia RUC Grant (1001/PSKBP/8630004) and Higher Institutions Centre of Excellence (HICoE) Grant (311/CIPPM/4401005) from the Malaysia Ministry of Higher Education. S.W. Leong would also like to thank the Malaysia Ministry of Higher Education MyBrain Science scholarship.

## CONFLICT OF INTEREST

The authors declare that they have no competing interest.

## ORCID

*Yee Siew Choong* http://orcid.org/0000-0001-5067-2073

## REFERENCES


1. Khoury GA, Smadbeck J, Kieslich CA, Floudas CA. Protein folding and *de novo* protein design for biotechnological applications. *Trends Biotechnol*. 2014;32(2):99-109.
2. Nelson AL, Reichert JM. Development trends for therapeutic antibody fragments. *Nat Biotechnol*. 2009;27(4):331-337.
3. Ducancel F, Muller BH. Molecular engineering of antibodies for therapeutic and diagnostic purposes. *MAbs*. 2012;4(4):445-457.
4. Fasihi-Ramandi M, Amani J, Salmanian AH, Moazzeni SM, Ahmadi K. *In silico* designing, cloning, and heterologous expression of novel chimeric human B lymphocyte CD20 extra loop. *Tumour Biol*. 2016;37(9):12547-12553.
5. Kiyoshi M, Caaveiro JM, Miura E, et al. Affinity improvement of a therapeutic antibody by structure-based computational design: generation of electrostatic interactions in the transition state stabilizes the antibody-antigen complex. *PLoS One*. 2014;9(1):e87099
6. Krawczyk K, Dunbar J, Deane CM. Computational tools for aiding rational antibody design. *Methods Mol Biol*. 2016;1529:399-416.
7. Kuroda D, Shirai H, Jacobson MP, Nakamura H. Computer-aided antibody design. *Protein Eng Des Sel*. 2012;25(10):507-521.
8. Li T, Pantazes RJ, Maranas CD. OptMAVEn—a new framework for the de novo design of antibody variable region models targeting specific antigen epitopes. *PLoS One*. 2014;9(8):e105954
9. Rahbar MR, Rasooli I, Gargari SL, et al. A potential *in silico* antibody-antigen based diagnostic test for precise identification of *Acinetobacter baumannii*. *J Theor Biol*. 2012;294:29-39.
10. Shirai H, Prades C, Vita R, et al. Antibody informatics for drug discovery. *Biochim Biophys Acta*. 2014;1844(11):2002-2015.
11. Tiller KE, Tessier PM. Advances in antibody design. *Annu Rev Biomed Eng*. 2015;17(1):191-216.
12. Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ*. 2004;82(5):346-353.
13. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid fever. *Clin Infect Dis*. 2010;50(2):241-246.
14. Neil KP, Sodha SV, Lukwago L, et al. A large outbreak of typhoid fever associated with a high rate of intestinal perforation in Kasese district, Uganda, 2008-2009. *Clin Infect Dis*. 2012;54(8):1091-1099.
15. Singla N, Bansal N, Gupta V, Chander J. Outbreak of *Salmonella typhi* enteric fever in sub-urban area of North India: a public health perspective. *Asian Pac J Trop Med*. 2013;6(2):167-168.
16. Cherian J, Sampath S, Sunderamurthy B, Chavada V, Vasudevan K, Govindasamy A. An outbreak investigation of typhoid fever in Pondicherry, South India, 2013. *Int J Med Sci Public Health*. 2015;4(2):256-261.
17. Limpitikul W, Henpraserttae N, Saksawad R, Laoprasopwattana K. Typhoid outbreak in Songkhla, Thailand 2009–2011: clinical outcomes, susceptibility patterns, and reliability of serology tests. *PLoS One*. 2014;9(11):e111768
18. Lutterloh E, Likaka A, Sejvar J, et al. Multidrug-resistant typhoid fever with neurologic findings on the Malawi-Mozambique border. *Clin Infect Dis*. 2012;54(8):1100-1106.
19. WHO. Typhoid vaccines: WHO position paper. *Wkly Epidemiol Rec*. 2008;83:49-59.
20. Khan KH, Ganjewala D, Rao KVB. Recent advancement in typhoid research—a review. *Advanced Biotech*. 2008;7:35-41.
21. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid fever. *New Engl J Med*. 2002;347(22):1770-1782.
22. Chaicumpa W, Thin-Inta W, Khusmith S, et al. Detection with monoclonal antibody of *Salmonella typhi* antigen 9 in specimens from patients. *J Clin Microbiol*. 1988;26(9):1824-1830.
23. Quang NN, Tapchaisri P, Chongsa-nguan M, et al. Diagnosis of enteric fever caused by *Salmonella* spp. in Vietnam by a monoclonal antibody-based dot-blot ELISA. *Asian Pac. J. Allerg Immunol*. 1997;15:205-212.
24. Choong YS, Lim TS, Chew AL, Aziah I, Ismail A. Structural and functional studies of a 50 kDa antigenic protein from *Salmonella enterica* serovar Typhi. *J Mol Graph Model*. 2011;29(6):834-842.
25. Tikhonova EB, Zgurskaya HI. AcrA, AcrB and TolC of *Escherichia coli* form a stable intermembrane multidrug efflux complex. *J Biol Chem*. 2004;279(31):32116-32124.
26. Fleishman SJ, Corn JE, Strauch E-M, Whitehead TA, Karanicolas J, Baker D. Hotspot-centric *de novo* design of protein binders. *J Mol Biol*. 2011;413(5):1047-1062.
27. Biggin PC, Bond PJ. In: Kukol A, ed. *Molecular Modeling of Proteins*. Humana Press: Totowa; 2008.
28. Kandt C, Ash WL, Peter TD. Setting up and running molecular dynamics simulations of membrane proteins. *Methods*. 2007;41(4):475-488.
29. Raetz CR. Enzymology, genetics, and regulation of membrane phospholipid synthesis in *Escherichia coli*. *Microbiol Rev*. 1978;42(3):614-659.
30. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graphs*. 1996;14(1):33-38.
31. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-1718.
32. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys J*. 2006;90(4):L36-L38.
33. Mackerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*. 1998;102(18):3586-3616.
34. Klauda JB, Venable RM, Freites JA, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B*. 2010;114(23):7830-7843.
35. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):014101
36. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 1981;52(12):7182-7190.
37. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A*. 1985;31(3):1695-1697.
38. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys*. 1984;52(2):255-268.
39. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18(12):1463-1472.
40. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew Chem Int Ed*. 1999;38(1-2):236-240.
41. Ramaraj T, Angel T, Dratz EA, Jesaitis AJ, Mumey B. Antigen-antibody interface properties: composition, residue interactions and features of 53 non-redundant structures. *Biochim Biophys Acta*. 2012;1824(3):520-532.
42. Pierce BG, Hourai Y, Weng Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE*. 2011;6(9):e24657
43. Gray JJ, Moughon S, Wang C, et al. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol*. 2003;331(1):281-299.
44. Combs SA, Deluca SL, Deluca SH, et al. Small-molecule ligand docking into comparative models with Rosetta. *Nat Protoc*. 2013;8(7):1277-1298.
45. Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci*. 1997;6(8):1661-1681.
46. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol*. 1997;268(1):209-225.

47. Das R. Four small puzzles that Rosetta doesn't solve. *PLoS One*. 2011;6(5):e20044
48. Hu X, Wang H, Ke H, Kuhlman B. High-resolution design of a protein loop. *Proc Natl Acad Sci U S A*. 2007;104(45):17668-17673.
49. Siegel JB, Zanghellini A, Lovick HM, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*. 2010;329(5989):309-313.
50. Karanicolas J, Corn JE, Chen I, et al. A *de novo* protein binding pair by computational design and directed evolution. *Mol Cell*. 2010;42:250-260.
51. Thomas R, Patenaude SI, MacKenzie CR, et al. Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. *J Biol Chem*. 2002;277(3):2059-2064.
52. Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, Wittrup KD. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol*. 2004;343(3):685-701.
53. Honegger A, Spinelli S, Cambillau C, Plückthun A. A mutation designed to alter crystal packing permits structural analysis of a tight-binding fluorescein–scFv complex. *Prot Sci*. 2005;14(10):2537-2549.
54. Watanabe H, Tsumoto K, Taguchi S, et al. A human antibody fragment with high affinity for biodegradable polymer film. *Bioconjug Chem*. 2007;18(3):645-651.
55. Hwang WC, Lin Y, Santelli E, et al. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *J Biol Chem*. 2006;281(45):34610-34616.
56. Canul-Tec JC, Riaño-Umbarila L, Rudiño-Piñera E, Becerril B, Possani LD, Torres-Larios A. Structural basis of neutralization of the major toxic component from the scorpion *Centruroides noxius* Hoffmann by a human-derived single-chain antibody fragment. *J Biol Chem*. 2011;286(23):20892-20900.
57. Tanaka T, Williams RL, Rabbitts TH. Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *EMBO J*. 2007;26(13):3250-3259.
58. Clark KR, Walsh STR. Crystal structure of a 3B3 variant: a broadly neutralizing HIV-1 scFv antibody. *Protein Sci*. 2009;18(12):2429-2441.
59. Duhovny D, Nussinov R, Wolfson H. In: Guigó R, Gusfield D, eds. *Algorithms in Bioinformatics*. Springer-Verlag: Berlin; 2002.
60. Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011;332(6031):816-821.
61. Kunik V, Ashkenazi S, Ofran Y. Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. *Nucleic Acids Res*. 2012;40(W1):W521-W524.
62. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol*. 1993;234(4):946-950.
63. Accelrys Software Inc., Discovery studio modeling environment. 2013: San Diego.
64. Vaccaro L, Scott KA, Sansom MSP. Gating at both ends and breathing in the middle: conformational dynamics of TolC. *Biophys J*. 2008;95(12):5681-5691.
65. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*. 2000;405(6789):914-919.
66. Sundberg EJ, Mariuzza RA. In: Joel J, Shoshana JW, eds. *Advances in Protein Chemistry*. San Diego: Academic Press; 2002.
67. Chen S-WW, Van Regenmortel MHV, Pellequer J-L. Structure-activity relationships in peptide-antibody complexes: implications for epitope prediction and development of synthetic peptide vaccines. *Curr Med Chem*. 2009;16(8):953-964.
68. Johansson K. *Computational Protein Design, in The PhD School of Science*. Copenhagen: University of Copenhagen; 2012.
69. Shen J, Zhang W, Fang H, Perkins R, Tong W, Hong H. Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. *BMC Bioinformatics*. 2013;14(Suppl 14):S6
70. Kwon T, Harris AL, Rossi A, Bargiello TA. Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics. *J Gen Physiol*. 2011;138(5):475-493.
71. Andersen OS, Koeppe RE. Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct*. 2007;36(1):107-130.
72. Raunest M, Kandt C. Locked on one side only: ground state dynamics of the outer membrane efflux duct TolC. *Biochemistry*. 2012;51(8):1719-1729.
73. Wang B, Weng J, Fan K, Wang W. Elastic network model-based normal mode analysis reveals the conformational couplings in the tripartite AcrAB-TolC multidrug efflux complex. *Proteins*. 2011;79(10):2936-2945.
74. Azizian H, Bahrami H, Pasalar P, Amanlou M. Molecular modeling of *Helicobacter pylori* arginase and the inhibitor coordination interactions. *J Mol Graph Model*. 2010;28(7):626-635.
75. Arrigoni A, Bertini L, De Gioia L, Papaleo E. Inhibitors of the Cdc34 acidic loop: a computational investigation integrating molecular dynamics, virtual screening and docking approaches. *FEBS Open Bio*. 2014;4(1):473-484.
76. Sinko W, de Oliveira C, Williams S, et al. Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. *Chem Biol Drug Des*. 2011;77(6):412-420.
77. Procko E, Hedman R, Hamilton K, et al. Computational design of a protein-based enzyme inhibitor. *J Mol Biol*. 2013;425(18):3563-3575.
78. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003;302(5649):1364-1368.
79. Wandersman C, Delepelaire P. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci U S A*. 1990;87(12):4776-4780.
80. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. *Proc Natl Acad Sci U S A*. 2009;106(23):9215-9220.
81. Humphris EL, Kortemme T. Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure*. 2008;16(12):1777-1788.
82. Smadbeck J, Peterson MB, Khoury GA, Taylor MS, Floudas CA. Protein WISDOM: a workbench for in silico de novo design of biomolecules. *J Vis Exp*. 2013;e50476
83. Liu Y, Kuhlman B. RosettaDesign server for protein design. *Nucleic Acids Res*. 2006;34(Web Server):W235-W238.
84. Pantazes RJ, Maranas CD. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Eng Des Sel*. 2010;23(11):849-858.
85. Bellows ML, Taylor MS, Cole PA, et al. Discovery of entry inhibitors for HIV-1 via a new *de novo* protein design framework. *Biophys J*. 2010;99(10):3445-3453.
86. Sun J, Abdeljabbar DM, Clarke N, Bellows ML, Floudas CA, Link AJ. Reconstitution and engineering of apoptotic protein interactions on the bacterial cell surface. *J Mol Biol*. 2009;394(2):297-305.
87. Sahin E, Jordan JL, Spatara ML, et al. Computational design and biophysical characterization of aggregation-resistant point mutations for γD crystallin illustrate a balance of conformational stability and intrinsic aggregation propensity. *Biochemistry*. 2010;50:628-639.
88. Joo JC, Pack SP, Kim YH, Yoo YJ. Thermostabilization of *Bacillus circulans* xylanase: computational optimization of unstable residues based on thermal fluctuation analysis. *J Biotech*. 2011;151(1):56-65.

89. Kim HS, Le QAT, Kim YH. Development of thermostable lipase B from *Candida antarctica* (CalB) through *in silico* design employing B-factor and RosettaDesign. *Enzyme Microb Tech*. 2010;47(1-2):1-5.

90. Salgado EN, Radford RJ, Tezcan FA. Metal-directed protein self-assembly. *Acc Chem Res*. 2010;43(5):661-672.

91. Ismail A, Kader ZA, Ong KH. Dot enzyme immunosorbent assay for the serodiagnosis of typhoid fever. *Southeast Asian J Trop Med Public Health*. 1991;22(4):563-566.

92. Ismail A, Ong KH, Kader ZA. Demonstration of an antigenic protein specific for *Salmonella typhi*. *Biochem Biophys Res Commun*. 1991;181(1):301-305.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Leong SW, Lim TS, Ismail A, Choong YS. Integration of molecular dynamics simulation and hotspot residues grafting for de novo scFv design against *Salmonella* Typhi TolC protein. *J Mol Recognit*. 2017;e2695. https://doi.org/10.1002/jmr.2695