ORIGINAL PAPER

Modeling the three-dimensional structures of an unbound single-chain variable fragment (scFv) and its hypothetical complex with a *Corynespora cassiicola* toxin, cassiicolin

Adeel Malik · Ahmad Firoz · Vivekanand Jha · Elumalai Sunderasan · Shandar Ahmad

Abstract Rubber trees infected with a host-specific cassii-colin toxin often experience considerable leaf fall, which in turn results in loss of crop productivity. It was recently revealed that cassiicolin-specific single-chain variable fragments (scFv) can successfully reduce the toxic effects of cassiicolin. However, the detailed mechanism of antibody action remains poorly understood. The primary sequence of the newly sequenced cassiicolin-specific scFv was highly homologous to several members of single-chain antibodies in the 14B7 family. In this study, with the aid of homology modeling, the three-dimensional structure of cassiicolin-specific scFv was elucidated, and was found to exhibit a characteristic immunoglobulin fold that mainly consists of β sheets. Additionally, molecular docking between the modeled scFv antibody and the available three-dimensional crystal structure of cassiicolin toxin was also performed. The predicted structural complex and the change in accessible surface area between the toxin and the scFv antibody upon complexation reveal the potential role of certain complementarity determining region (CDR) amino acid residues in the formation of the complex. These computational results suggest that mutagenesis experiments that are aimed at validating the model and improving the binding affinity of cassiicolin-specific scFv antibodies for the toxin should be performed.

Keywords Cassiicolin-specific scFv antibody ·
Homology modeling · Protein–protein docking ·
Antibody–antigen interaction

A. Malik (✉) · A. Firoz · V. Jha
Biomedical Informatics Center, PGIMER,
Chandigarh 160012, India
e-mail: adeel@netasa.org

A. Firoz
e-mail: firoz@bmi.icmr.org.in

V. Jha
e-mail: vjha@pginephro.org

V. Jha
Department of Nephrology, PGIMER,
Chandigarh, India

E. Sunderasan
Biotechnology Unit, Malaysian Rubber Board,
RRIM Research Station,
Sungai Buloh, 47000 Selangor D.E., Malaysia
e-mail: sunderasan@lgm.gov.my

S. Ahmad
National Institute of Biomedical Innovation,
Saito-Asagi, Ibaraki Osaka, Japan
e-mail: shandar@nibio.go.jp

Introduction

Cassiicolin, a host-selective toxin (HST) [1], is produced by the pathogenic fungus *Corynespora cassiicola* (Berk. & Curt.) Wei. It infects a diverse range of crops, including economically important rubber [2]. Infection of rubber trees with *C. cassiicola*, characterized by the development of necrotic lesions and browning of the leaves, causing massive defoliation and thus crop losses [3], is known as Corynespora leaf fall disease (CLFD) [1]. In 1998, according to the IRRDB Disease Survey, CLFD was identified as the primary pathogenesis affecting the rubber tree population [4], mainly in Asia. It has also been reported that the toxin secreted by the fungus *C. cassiicola* that is responsible for CLFD is a 27-residue O-glycoprotein, cassiicolin [5, 6].

Recombinant single-chain variable fragment (scFv) antibodies have been successfully employed in plants for various different reasons [7, 8]. Recombinant antibody engineering is an attractive contemporary concept for the design of high-affinity, protein-based targeting reagents [9, 10], and it is utilized for more than 25% of all the proteins undergoing clinical trials [11, 12]. Antibodies (immunoglobulins) exhibit a characteristic structure consisting of two identical heavy and light chains connected together by disulfide and noncovalent bonds. Intact antibodies offer high target binding specificity, but slow tissue penetration, long circulating half-lives and often undesirable effector functions limit their application in rapid tumor targeting [11]. Additionally, antibodies have been complexed with many molecules, including toxins, enzymes and viruses for prodrug therapy, cancer treatment and gene delivery. Consequently, recombinant antibody technology has enabled knowledgeable manipulations in the construction of complex antibody library repertoires for the selection of high-affinity reagents against refractory targets [12]. In the context of this work, many groups have successfully engineered fab or scFv molecules into dimers or multimers for the production of favorably sized high-avidity reagents for in vivo imaging and therapeutics [13–16]. The co-authors of this work demonstrated that cassiicolin toxicity could be efficiently reduced with the aid of recombinant antibody technology; scFv specific to cassiicolin could bind and inhibit or decrease the toxic effects of cassiicolin in a *Hevea* leaf bioassay [3]. Therein, the authors verified the toxin-deactivating properties of the cassiicolin-specific antibody expressed by the scFv clones obtained from the phage library with a high specificity for cassiicolin [3]. The translated cassiicolin-specific scFv antibody sequence comprises of 268 amino acid residues. However, the three-dimensional (3D) structure of this antibody has not yet been determined, imposing significant limitations on a complete understanding of the antibody recognition mechanism of cassiicolin. In the absence of an experimentally determined structure of the unbound antibody or its complex with cassiicolin, we found that suitable template structures with significant homology to the sequence under consideration are available in the protein data bank (PDB) [17]. This enabled us to develop a modeled 3D structure of the cassiicolin-specific scFv. Experimental determinations of the 3D structures of antibodies complexed with protein antigens have been performed for at least two decades [18–27]. It has been reported that antibodies to protein antigens target a discontinuous epitope on the antigen [19]. It is also known that all six complementarity-determining regions (CDRs) of the antibody may interact with the antigen [19, 27–29], in addition to some framework residues [19]. However, in the absence of an experimentally determined 3D crystal structure, a comparative method of 3D structure

prediction is utilized. This method predicts the 3D structure of the target protein on the basis that the available 3D coordinates of the template structure exhibit reasonable sequence similarity. Homology modeling of scFvs has been recently reported for *anti*-HepG2 single-chain immunoglobulin, *anti*-CMV scFv antibody and scFv-GFP fusion product [30–33].

In the present study, a three-dimensional structural model of cassiicolin-specific scFv antibody was built on the basis of the template crystal structure of the anthrax-neutralizing single-chain antibody 14B7 (PDB code: 3ESU). The predicted scFv antibody was consistent with the experimentally observed immunoglobulin-like fold, which mainly consists of β sheets [34, 35]. Since the antigen structure is already known, this allowed us to dock the two structures and try to generate a hypothetical model of the complex. Based on our computational results, the potential interacting residues of the scFv–cassiicolin complex, as well as the involvement of various cation–π and hydrophobic interactions were identified, which were consistent with the experimental findings.

Materials and methods

Cassiicolin-specific scFv antibody sequence

The complete sequence of cassiicolin-specific scFv was recently reported by Sunderasan et al. [3], and has the GenBank accession number EU414027. The deduced sequence of the cassiicolin-specific scFv antibody consists of 268 amino acid residues and a calculated molecular mass of 29,110.79 Da. The scFv sequence contains a 7 amino acid (GGSSRSS) linker, a 6 amino acid long histidine (His) tag and a 12 residue long influenza hemaglutinin (HA) tag at its C-terminal end.

Template selection and homology modeling

SWISS–MODEL Workspace [36] was used to identify a suitable template for determining the 3D structure of scFv by homology modeling. Five different three-dimensional structural models for the cassiicolin-specific scFv were then generated using Modeller9v2 [37] on the basis of the 3D structures of the templates. These models were optimized by adding hydrogen atoms to each of them at pH 7.0 and then using CHARMM all-atom forcefield minimization with the implicit solvent model set to none. Energy was minimized for at least 1000 steps until the gradient converged to 0.5 kcal mol^{-1} using the steepest descent protocol available in Discovery Studio version 2.0 (Accelrys Software, Inc.) [38] to remove any steric clashes and to stabilize the models. Ramachandran and Whatcheck

analysis was performed for subsequent optimization of all five models using the SAVES [39] web server, and the model exhibiting the best Ramachandran score and Whatcheck RMS Z score summary was retained for further studies.

Multiple sequence alignment and secondary structure prediction

Multiple sequence alignment was carried out to check the positions of the CDR regions in the templates and the query. Since scFv antibodies are known to have β structures, in order to confirm whether our query sequence also possessed this type of secondary structure its secondary structure was predicted. Multiple sequence alignment of the cassiicolin-specific scFv antibody and its selected homologs was carried out using CLUSTALW [40]. Automatic alignments were critically analyzed and compared to each other. Secondary structural elements of the aligned proteins were predicted by the Jpred 3 program [41].

Molecular docking

For docking, the 3D structure of cassiicolin from the PDB (PDB code 2HGO) was used and processed and cleaned in Discovery Studio version 2.0 by removing all the nonprotein parts. Docking of the cassiicolin-specific scFv (model 4) to the cassiicolin was performed using the automated initial stage docking algorithm implemented by ZDOCK [42]. ZDOCK employs a fast Fourier transform to search every single possible binding mode for the proteins, evaluating based on shape complementarity, desolvation energy, and electrostatics.

The output of the ZDOCK was passed to the ClusPro [43] server with the clustering radius set at 5 Å, and 30 cassiicolin–scFv complexes were retrieved. Again, energy minimization was performed on all the 30 complexes returned by ClusPro [43] using the steepest descent protocol available in Discovery Studio version 2.0 [38]. The energy minimization criterion was kept the same as the one used to optimize the homology model of the cassiicolin-specific scFv antibody (i.e., CHARMM all-atom forcefield minimization with the implicit solvent model set to none was used). All 30 cassiicolin–scFv complexes were analyzed for interacting residues at a cut-off distance of 6 Å using an in-house program written in PERL.

Computational alanine scanning and accessible surface area (ASA) or solvent accessibility analysis of the docked complex

In order to analyze the roles of the interacting residues in the stability of the scFv–cassiicolin complex, computa-

tional alanine scanning mutagenesis was carried out using the ROBETTA alanine scanning software [44]. The computational alanine scanning program employs a simple free-energy function consisting of a linear combination of a Lennard–Jones potential, an orientation-dependent hydrogen-bonding potential, statistical terms approximating the backbone-dependent amino acid type and rotamer probabilities, and estimates of the unfolded reference state energies to calculate the effects of alanine mutations on the binding free energy of the protein–protein complex [44].

$$\Delta\Delta G_{bind} = \left[\left(\Delta G_{complex}^W - \Delta G_{protein\ A}^W - \Delta G_{protein\ B}^W \right) \right] \\ - \left[\left(\Delta G_{complex}^M - \Delta G_{protein\ A}^M - \Delta G_{protein\ B}^M \right) \right] \quad (1)$$

where

$\Delta\Delta G_{bind}$	is the predicted change in binding free energy upon alanine mutation
$\Delta G_{complex}^W$	is the predicted change in protein stability of the wild-type complex
$\Delta G_{protein\ A}^W$	is the predicted change in protein stability of protein A in the wild-type complex
$\Delta G_{protein\ B}^W$	is the predicted change in protein stability of protein B in the wild-type complex
$\Delta G_{complex}^M$	is the predicted change in protein stability of the mutant complex
$\Delta G_{protein\ A}^M$	is the predicted change in protein stability of protein A in the mutant complex
$\Delta G_{protein\ B}^M$	is the predicted change in protein stability of protein B in the mutant complex.

The input for the computational alanine scanning consisted of a three-dimensional modeled structure of the scFv–cassiicolin complex, and it reported a list of potential amino acids that are predicted to drastically destabilize the interface when mutated to alanine, similar to the results of experimental alanine-scanning mutagenesis. These energetically important residues or "hotspots" were defined by Clackson and Wells as an experimentally observed change in binding free energy upon alanine substitution of more than 1 kcal mol^{-1} in their work on the binding of human hormone to its receptor [45].

Further analysis of the selected complexes was performed by calculating the change in accessible surface area (ASA). The total ASAs were calculated for the unbound structures and for the complexes using the DSSP [46] program. The changes in ASA for the complexes were calculated by subtracting the total ASA of

the complex from the sum of the ASAs of the unbound structures [47]:

$$\Delta ASA = ASA_{complex} - [ASA_{scf} + ASA_{cf}]. \qquad (2)$$

Here, ΔASA denotes the interface area (to be computed), $ASA_{complex}$ refers to the total ASA of the docked complex, ASA_{scf} is the ASA of free cassiicolin-specific scFv, and ASA_{cf} is the ASA of free cassiicolin in the modeled structures.

Results and discussion

Homology modeling

The 3D structure of the cassiicolin-specific scFv antibody was built by homology modeling based on the recently determined template crystal structure of the anthrax-neutralizing single-chain antibody 14B7 (PDB code: 3ESU) [48]. Templates were identified with the help of SWISS-MODEL Workspace [36] and the 3D coordinates of neutralizing single-chain antibody 14B7 (PDB code: 3ESU, chain F) was selected as the final template due to its high-resolution structure (1.30 Å) and ~61% of sequence identity with the target cassiicolin-specific scFv. In order to validate whether our selection of template for homology modeling was correct, we also performed a protein structure comparison between our modeled and template structures using the secondary structure matching (SSM) [49] analysis tool. Structure alignment results also confirmed that the murine monoclonal antibody (PDB code 3ESU) was the best template for model building. SSM offers various different types of measuring scores to aid structural comparison: (i) Q score, which represents the quality function of $C\alpha$ alignment; (ii) P score, which represents minus the logarithm of the P value (the higher P score, i.e., the lower P value, the more statistically significant the match; for an alignment to be statistically significant, the P score must be greater than 3); (iii) Z score, which measures the statistical significance of a match in terms of Gaussian statistics (the higher Z score, the higher the statistical significance of the match). Figure 1 summarizes all of these scores as well as the RMSDs for our modeled cassiicolin-specific scFv antibody and its homologs in the PDB [17]. Figure 1 shows that the Q score between the modeled query and 3ESU is 0.70. This score only reaches 1 for similar structures. Similarly, very high values of the P score (17.4) and Z score (13) indicated that 3ESU was a high-quality template for homology modeling.

Since there was no alignment at the C-terminal region, 15 residues in addition to the histidine (His) and hemaglutinin (HA) tags were removed before building the model

and five model structures were generated in this way. After minimizing the energy and stabilizing the models by removing steric clashes, Ramachandran and Whatcheck analysis was performed on all five models using the SAVES [39] web server. Although a few residues (ALA13 and THR50) were found in the disallowed regions of the Ramachandran plot, and they possibly represent incorrect side-chain orientations, we did not try to correct them, as they lie outside the interface and have no effect on the recognition process. Among all of these models, model 4 exhibited the best Ramachandran score and Whatcheck RMS Z score summary and thus was selected for further analysis (Tables 1 and 2). The selected model has 88.6% of its residues in the allowed region, 7.5% in the additionally allowed region and 3% in the generously allowed region. Only 1% of its residues were in the disallowed regions (Fig. 2).

The modeled structure of cassiicolin-specific scFv and its template murine monoclonal antibody exhibit the characteristic immunoglobulin fold indicating a large binding surface formed by the six CDRs [48]. Both of these scFv antibodies have practically the same folds and consist chiefly of β sheets. The presence of β sheets within the V domains of antibodies is a well-known fact [34, 35].

Multiple sequence alignment and secondary structure prediction

The amino acid sequences of cassiicolin-specific scFv antibody and its closest homologs with the PDB [17] codes 1MOEA, 3ESUF, 3ESVF, 2GJJA and 1H8SA were aligned with the CLUSTALW [40] algorithm. Given their amino acid sequences, their secondary structures were also predicted and analyzed by the Jpred 3 program [41]. The secondary structure of the cassiicolin-specific scFv antibody showed a close similarity to the whole structures of the prototype scFv antibodies, which consisted mainly of β strands [34]. Furthermore, in spite of several amino acid variations in the primary structures of cassiicolin-specific scFv and its homologs, their secondary structures turned out to be identical. Based on the alignment, it can be seen that all the light chain CDRs are of equal length, but the CDRs of heavy chains show some differences (Fig. 3). Despite these differences in the lengths of CDRs, these regions are rich in hydrophobic residues that are important for hydrophobic interactions and are found to predominant in protein–protein interactions [50, 51].

Molecular docking

The docking of cassiicolin onto the scFv antibody was performed with the help of ZDOCK [42] and the top 2000

Structure Alignment Results

explanation of output

Query: **model.6.pdb**, chain A ▾ : 235 residues.

Examined 56960 entries (133869 chains),
Matches 1-20 of 28.

Back to query   >>   last page

resort results

| ## | Scoring | | | Rmsd | N$_{algn}$ | N$_g$ | %seq | Query | | | N$_{res}$ | × | Target (PDB entry) |
	Q	P	Z					%$_{sse}$	Match	%$_{sse}$			Title
1	0.71	4.4	8.0	1.50	217	5	66	74	3etb:F	88	227	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY M18 COMPLEXED WITH ANTHRAX PROTECTIVE ANTIGEN DOMAIN 4
2	0.70	4.0	7.9	1.50	217	5	66	74	3etb:H	93	228	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY M18 COMPLEXED WITH ANTHRAX PROTECTIVE ANTIGEN DOMAIN 4
3	0.70	4.4	8.1	1.47	216	5	66	74	3etb:G	93	228	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY M18 COMPLEXED WITH ANTHRAX PROTECTIVE ANTIGEN DOMAIN 4
4	0.70	17.4	13.0	1.58	216	5	65	89	3esu:F	85	222	☐	CRYSTAL STRUCTURE OF ANTHRAX-NEUTRALIZING SINGLE-CHAIN ANTIBODY 14B7
5	0.69	17.7	13.2	1.53	216	5	65	89	3esv:F	81	228	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY M18
6	0.69	17.5	13.0	1.60	217	5	65	89	3et9:F	85	226	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY 1H
7	0.68	17.1	13.0	1.63	217	5	65	89	3esv:G	81	227	☐	CRYSTAL STRUCTURE OF THE ENGINEERED NEUTRALIZING ANTIBODY M18
8	0.61	16.2	12.8	1.74	212	4	64	95	2gjj:B	82	234	☐	CRYSTAL STRUCTURE OF A SINGLE CHAIN ANTIBODY SCA21 AGAINST HER2/ERBB2
9	0.61	13.4	12.0	1.82	215	5	64	89	2gjj:A	81	236	☐	CRYSTAL STRUCTURE OF A SINGLE CHAIN ANTIBODY SCA21 AGAINST HER2/ERBB2
10	0.58	12.5	11.5	2.01	214	5	55	95	1jp5:B	86	232	☐	CRYSTAL STRUCTURE OF THE SINGLE-CHAIN FV FRAGMENT 1696 IN COMPLEX WITH THE EPITOPE PEPTIDE CORRESPONDING TO N-TERMINUS OF HIV-1 PROTEASE
11	0.58	11.3	10.8	1.87	203	5	67	89	1h8s:B	85	218	☐	THREE-DIMENSIONAL STRUCTURE OF ANTI-AMPICILLIN SINGLE CHAIN FV FRAGMENT COMPLEXED WITH THE HAPTEN.
12	0.57	11.9	11.2	2.01	213	4	55	95	1jp5:A	86	232	☐	CRYSTAL STRUCTURE OF THE SINGLE-CHAIN FV FRAGMENT 1696 IN COMPLEX WITH THE EPITOPE PEPTIDE CORRESPONDING TO N-TERMINUS OF HIV-1 PROTEASE
13	0.57	11.0	10.9	2.01	213	5	55	89	1svz:B	81	232	☐	CRYSTAL STRUCTURE OF THE SINGLE-CHAIN FV FRAGMENT 1696 IN COMPLEX WITH THE EPITOPE PEPTIDE CORRESPONDING TO N-TERMINUS OF HIV-2 PROTEASE
14	0.57	11.1	11.3	2.06	214	4	55	95	1svz:A	82	232	☐	CRYSTAL STRUCTURE OF THE SINGLE-CHAIN FV FRAGMENT 1696 IN COMPLEX WITH THE EPITOPE PEPTIDE CORRESPONDING TO N-TERMINUS OF HIV-2 PROTEASE

Fig. 1 Comparison of the Q, P and Z scores of different templates for the query protein cassiicolin-specific antibody. The best template, which has a PDB code of 3ESU, based on these scores is enclosed in a *red box*

putative scFv antibody–cassiicolin binding configurations were retained in the final results of ZDOCK [42]. These 2000 scFv antibody–cassiicolin binding configurations from ZDOCK [42] were passed to the ClusPro [43] server, which is an automated rigid-body docking and discrimination method that quickly filters docked conformations by selecting those with favorable desolvation and electrostatic properties, clusters the retained structures using a hierarchical pair-wise RMSD algorithm, and selects the centers of the most populated clusters as predictions for the unknown complex [43]. In this way, 30 docked complexes were generated, and each of these 30 complexes were again optimized via CHARMM forcefield minimization using the steepest descent algorithm in Discovery Studio 2. Out of these 30 complexes, we selected complex 6 because it showed interactions at 5 out of the 6 CDR regions. On the

other hand, the potential energy of complex 30 was slightly lower than that of complex 6, but it exhibited interactions at only 4 out of the 6 CDR regions. Both of these complexes showed similar changes in ASA upon binding to the modeled VH and VL regions of cassiicolin-specific scFv (model 4). Table 3 shows a comparison between the energies and the potential CDRs that may participate in the complex formation. Another reason for selecting complex 6 was that it was ranked much higher than complex 30 by the ClusPro server [43]. A comparison of various interaction energies (interaction energy, electrostatic interaction energy and VDW interaction energy) for complex 6 is shown in Fig. 4. The CDRs of both the light and heavy chains of the scFv antibody face cassiicolin in the antibody–antigen complex. Within the Fv domains, the residues in scFv that were involved in contacts occur

Table 1 Whatcheck RMS Z-score summary. The potential energies (kcal mol^{-1}) of the minimized models are given in parentheses

	Model 1 (−12667.20)	Model 2 (−12921.88)	Model 3 (−12609.37)	Model 4 (−2635.03)	Model 5 (−2372.59)
Bond lengths	1.40	1.44	1.37	1.37	1.42
Bond angles	1.38	1.40	1.38	1.39	1.40
Omega angle restraints	0.88	0.81	0.78	0.87	0.88
Side-chain planarity	0.33	0.46	0.32	0.37	0.37
Improper dihedral distribution	0.93	1.07	0.98	0.97	1.01
Inside/outside distribution	1.01	1.03	0.97	1.00	1.04
Average	0.99	1.03	0.97	1.00	1.02

Table 2 Ramachandran analysis of the five models

	Model 1	Model 2	Model 3	Model 4	Model 5
Residues in most favored regions (%)	86.1	88.6	87.6	88.6	88.6
Residues in additional allowed regions (%)	9.5	7.5	8.5	7.5	8.0
Residues in generously allowed regions (%)	3.0	2.5	2.0	3.0	2.0
Residues in disallowed regions (%)	1.5	1.5	2.0	1.0	1.5

mainly at the VL–VH interface. The interacting residues identified on the CDR regions of the cassiicolin-specific scFv were as follows: SER30, TYR31, ASP49, TRP90, SER91, SER92, ASN93, LYS173, and TYR216. The residues were mainly involved in buried hydrophobic interactions that stabilized the interface between the VH and VL domains. Additionally, scFv ARG219 seemed to be one of the interacting residues (Fig. 5a, b). From the cassiicolin side, PHE8, PHE12, GLY18, ASN19 and SER20 may interact with the scFv antibody. A number of possible interactions between scFv and toxin residues can be predicted in the modeled complex. In particular, some hydrophobic interactions can be predicted between scFv SER30-VL (CDR1-VL) and toxin PHE12 and ALA22,

scFv TYR31-VL (CDR1-VL) and toxin PHE12 and TRP21, and scFv TRP90-VL (CDR3-VL) and toxin SER20. Furthermore, in the modeled scFv–toxin complex, toxin SER20 and GLY18 appear to interact with two serine residues of the light chain scFv antibody, namely SER91-VL (CDR3-VL) and scFv SER92-VL (CDR3-VL). Similarly, hydrophobic interactions can also be predicted for scFv TYR216-VH (CDR3-VH) and toxin PHE8 and GLY9 amino acids. Moreover, the presence of a cation–π interaction can also be predicted in the modeled complex between the scFv ARG219 and toxin PHE8. The presence of these hydrophobic residues and their interactions is consistent with experimentally derived antibody structures and their complexes with antigens [48, 52–54].
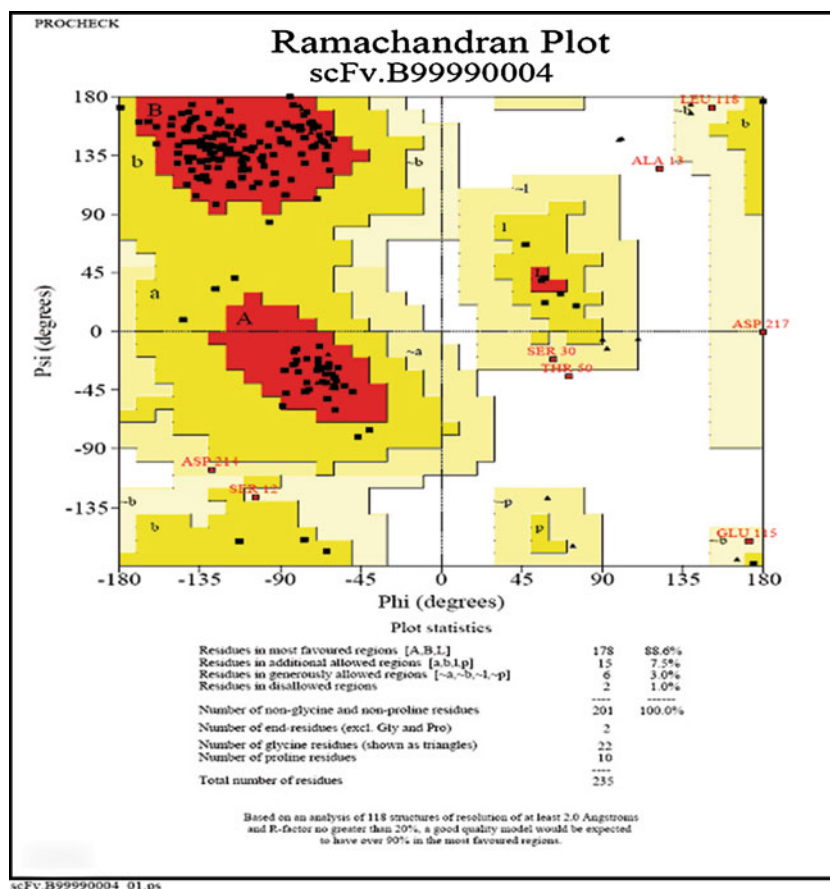
Fig. 2 Ramachandran analysis of the modeled cassiicolin-specific scFv antibody

Fig. 3 Multiple sequence alignment of cassiicolin-specific scFv and its homologs. The CDR regions of the homologs are enclosed in *black boxes*, while the CDRs of our query scFv are enclosed within a *blue box*

Computational alanine scanning and accessible surface area (ASA) or solvent accessibility analysis of the docked complex

Computational mutagenesis alanine scanning [44] analysis of the proposed complex indicated positive binding free energies ($\Delta\Delta G_{bind}$) upon alanine mutation for CDR-L1 TYR31 and ARG219 of the light and heavy chains, respectively (Table 4). Similar positive binding free energies

Table 3 Comparison of different docked complexes in terms of their potential energies and the involvement of various CDR regions. Each complex is numbered according to its ClusPro rank

Name	PE (kcal/mol)	CDRs that show contacts
Complex 6	−13329.58	VL1, VL2, VL3, VH2, VH3
Complex 19	−13227.53	VL1, VL2, VH1, VH3
Complex 21	−13206.90	VL1, VL2, VH3
Complex 22	−13272.90	VL2, VH1, VH3
Complex 24	−13301.21	VH1, VH2
Complex 29	−13308.86	VL1, VL2, VH3
Complex 30	−13338.29	VL1, VL2 VL3, VH3

were obtained for the cassiicolin residues PHE8 and PHE12. Heavy chain ARG219 is well conserved in the template structure 3ESU represented by ARG208, and is a part of the heavy chain CDR region (H3). These results show the significance of these residues in stabilizing the interfaces. CDR-L1 TYR31 could be one of the hotspot residues predicted to be primarily responsible for the binding energy of the complex, contributing >1 kcal mol^{-1} of free energy. Similarly, the substitution of SER30 and SER92 may have a neutral effect on the stability of the complex, indicating that these two amino acids could potentially participate in the binding with the toxin.

To further identify most of the stabilizing/hotspot residues, we also carried out an analysis of the interface area of each residue. Although all ten interacting residues exhibited changes in ASA, only four of them were calculated by computational alanine scanning to either destabilize the complex (hotspot) or to cause partial binding energy loss (neutral) upon mutations in the scFv–toxin complex rather than to destabilize the complex. This could be due to the limitations of the computational alanine scanning tools available to us (e.g., inter-residue coupling, indirect effects and structural flexibility are not considered).

Fig. 4 Comparison of the interaction energies for docked complexes before energy minimization and after energy minimization. Here IE-AEM is the interaction energy after energy minimization of the complex, IE-BEM is the interaction energy before energy minimization, EIE-AEM is the electrostatic interaction energy after energy minimization, EIE-BEM is the electrostatic interaction energy before energy minimization, VDW-IE-AEM is the VDW interaction energy after energy minimization, and VDW-IE-BEM is the VDW interaction energy before energy minimization
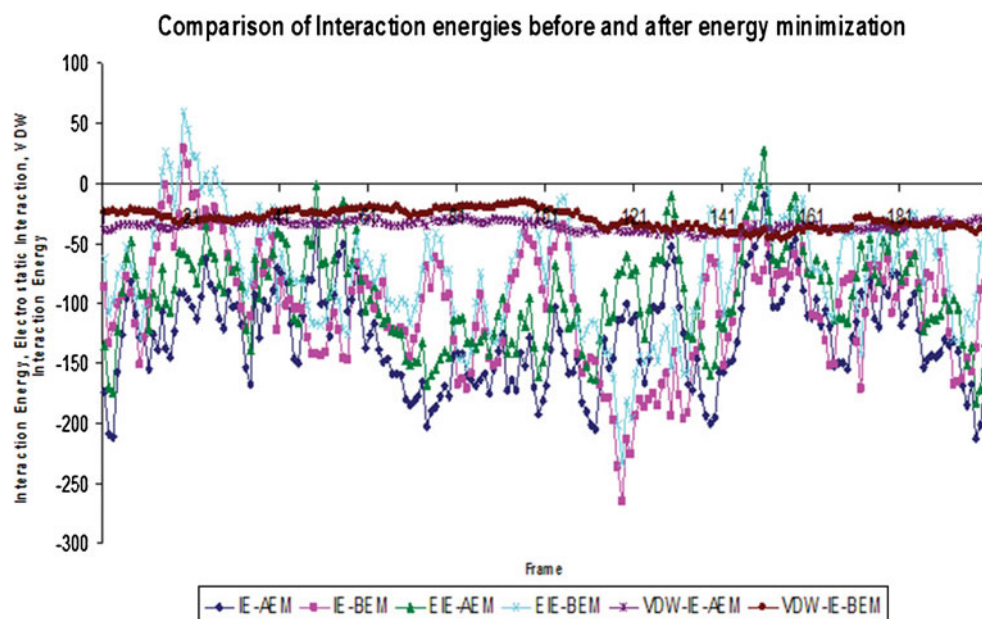
However, a broad picture of the hotspot candidates does emerge. It would be interesting to obtain a more firmly bound complex by introducing mutations that enhance the change in ASA.

Additional analysis of the selected complexes was performed by calculating the change in accessible surface area (ASA). For the proposed cassiicolin–scFv complex (i.e., complex 6), the total change in ASA observed was
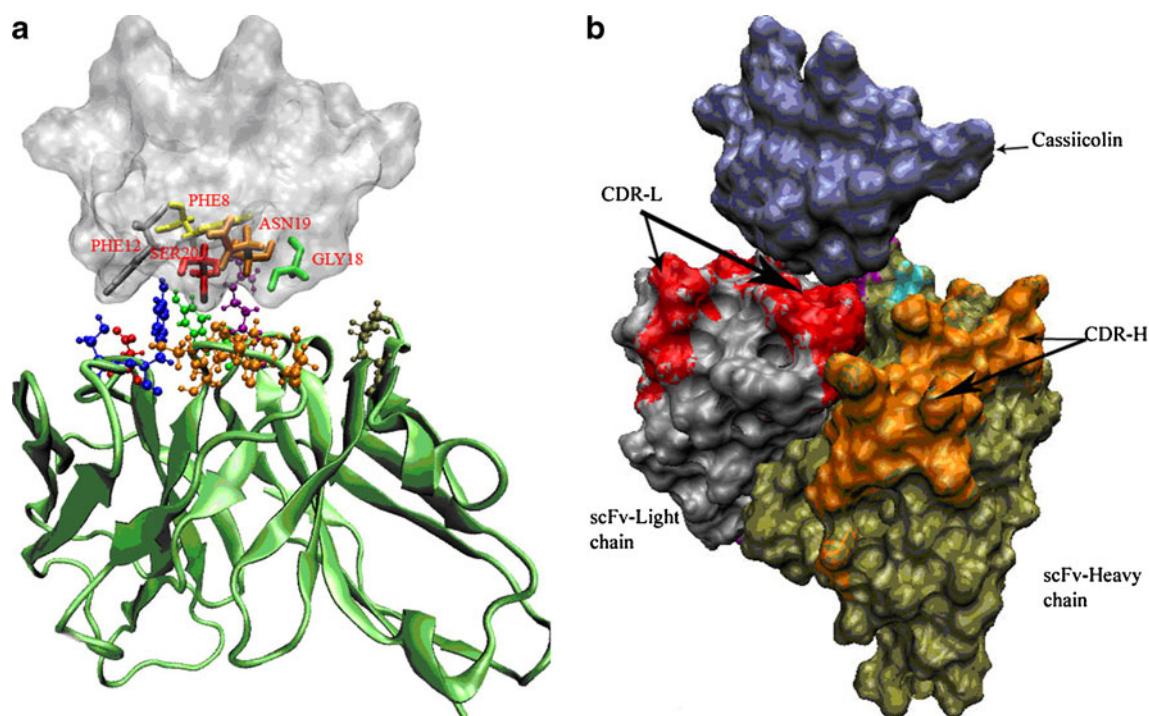

Fig. 5 **a** Cassiicolin–scFv complex. In this figure, the scFv antibody is represented in *green* and cassiicolin toxin in *gray*. Here, the interacting residues (SER30, TYR31, ASP49, TRP90, SER91, SER92, ASN93, LYS173, TYR216 AND ARG219) from scFv are shown as CPK, while the interacting residues from cassiicolin toxin are shown as licorice and are also labeled. **b** Cassiicolin–scFv complex: light chain CDRs are shown in *red* and heavy chain CDRs in *orange*

Table 4 The results of computational alanine scanning on the modeled scFv–cassiicolin complex. Here, *Residue (PDB)* represents the residue number in the original PDB coordinate file, *Residue (complex)* represents the continuous residue numbering of both chains, starting with residue number 1, *Protein type* indicates whether the binding partner is a receptor or a ligand, *Chain* represents the chain name in the docked complex, and $\Delta\Delta G$ *(bind)* represents the predicted change in binding free energy upon alanine mutation. Positive values of $\Delta\Delta G$ (bind) indicate that replacement by alanine is predicted to destabilize the complex, while negative values may predict a stabilizing effect

Residue (PDB)	Residue (complex)	Protein type	Chain	$\Delta\Delta G$ (bind)
SER30	SER30	Receptor	A	−0.1
TYR31	TYR31	Receptor	A	1.36
SER92	SER92	Receptor	A	−0.01
ARG219	ARG219	Receptor	A	0.26
PHE8	PHE243	Ligand	B	0.31
PHE12	PHE247	Ligand	B	0.60
ASN19	ASN254	Ligand	B	0.02
SER20	SER255	Ligand	B	−0.02

~460 Å^2, but the change in ASA for light chain CDRs is greater than that for heavy chain CDRs, indicating a greater contribution of the former to the binding. To further establish the reliability of the modeled complex, we computed the accessible surface area to compare the change in the ASA of the docked complex with the available experimental data in the benchmark database [55]. The benchmark database contains 12 experimentally determined antigen–antibody complexes, but there was no complex in the benchmark database that was sufficiently similar to our modeled complex to enable a comparison. Only one antigen–antibody (PDB code: 1I9R) complex exhibited a significant sequence similarity to our scFv antibody. Also, the sizes of the antigens were very different in these cases. While cassiicolin is 27 residues long, the length of the antigen in 1I9R is 146 residues. No other antigen–antibody complex in the benchmark database exhibited any similarity with our scFv antibody. In a similar docking analysis, Konstantakaki et al. [47] reported a total change in ASA of >750 Å^2 (as compared to ~460 Å^2 in our study) for their proposed complex, which is comparable to that for an intermediate size binding interface for an antigen–antibody complex. The size of the antigen in their study was 370 residues.

The change in ASA for each of the CDRs upon complex formation can be ordered as follows: L1>L3>H2>L2>H3>H1. All CDRs seem to make contact with cassiicolin, but very few contacts were observed for the second CDR of the light chain (L2). This is in agreement with the general notion that L2 is often not required for binding, while L3 and H3 are almost always involved [56, 57]. Amino acid residues with large changes in ASA in the cassiicolin–scFv complex for the light chain are CDR-L1 TYR31 (~93 Å^2), CDR2-L2 ASP49 (~20 Å^2), CDR3-L3 TRP90 (~10 Å^2), SER91 (~28 Å^2), SER92 (~16 Å^2), ASN93 (~20 Å^2), and for the heavy chain

are CDR2-H2 ASP170 (~10 Å^2), LYS177 (~12 Å^2) and CDR3-H3 GLU203 (~14 Å^2). Additionally, a large change in ASA was observed for ARG219 (~39 Å^2). Similarly, large changes in ASA were observed for PHE8 (~37 Å^2), PHE12 (~52 Å^2), GLY18 (~32 Å^2), ASN19 (~20 Å^2) and SER20 (~63 Å^2) for cassiicolin, and these seem to face towards the CDRs of the scFv antibody.

In the scFv–cassiicolin complex, ARG219 seems to be one of the few residues that may contribute to the stability of the complex, and it exhibits a significant change in ASA. It would be interesting to explore how this residue is involved in stabilizing the scFv–toxin complex by performing mutagenesis experiments.

Conclusions

In the present work, we predict the structural characterization of the molecular complex formed between a single-chain variable fragment (scFv) antibody specific to *C. cassiicola* toxin, cassiicolin, which induces necrotic lesions in rubber tree (*Hevea brasiliensis*). The interacting antigen and antibody surfaces complement each other structurally, as in other known antigen–antibody interactions, as well as chemically, with a few polar residues from the scFv antibody interacting with the polar residues from the toxin. Many of the surface amino acids of the antibody at the interface are aromatic, and therefore present large areas of hydrophobic surface to the toxin. Additionally, residues like ARG219 may contribute to cation–π interactions with the toxin. Finally, in the absence of the crystal structure of the complex, we hope that the three-dimensional modeled complex described here will prove valuable as a structural scaffold for designing further mutational studies and binding experiments.

References

1. Barthe P, Pujade-Renaud V, Breton F, Gargani D, Thai R, Roumestand C, de Lamotte F (2007) Structural analysis of cassiicolin, a host-selective protein toxin from *Corynespora cassiicola*. J Mol Biol 367:89–101

2. Silva WPK, Multani DS, Deverall BJ, Lyon BR (1995) RFLP and RAPD analyses in the identification and differentiation of isolates of the leaf spot fungus *Corynespora cassiicola*. Aust J Bot 43:609–618

3. Sunderasan E, Kadir RA, Pujade-Renaud V, De Lamotte F, Yeang HY, Nathan S (2009) Single-chain variable fragments antibody specific to *Corynespora cassiicola* toxin, cassiicolin, reduces necrotic lesion formation in *Hevea brasiliensis*. J Gen Plant Pathol 75:19–26

4. Hashim I (1998) Disease survey. IRRDB, Kuala Lumpur

5. Bréton F, Sanier C, d'Auzac J (2000) Role of cassiicolin, a host-selective toxin, in pathogenicity of *Corynespora cassiicola*, causal agent of a leaf fall disease of *Hevea*. J Rubber Res 3:115–128

6. De Lamotte F, Duviau M-P, Sanier C, Thai R, Poncet J, Bieysse D, Breton F, Pujade-Renaud V (2007) Purification and characterization of cassiicolin, the toxin produced by *Corynespora cassiicola*, causal agent of the leaf fall disease of rubber tree. J Chromatogr B 849:357–362

7. Van der Logt CPE, Sidebottom CS, Davies PJ (1998) Antibody production in plants. In: Shewry PR, Napier JA, Davies PJ (eds) Engineering crop plants for industrial end uses. Portland, London, pp 17–33

8. Malembic S, Saillard C, Bové JM, Garnier M (2002) Effect of polyclonal, monoclonal, and recombinant (single-chain variable fragment) antibodies on in vitro morphology, growth, and metabolism of the phytopathogenic mollicute *Spiroplasma citri*. Appl Environ Microbiol 68:2113–2119

9. Todorovska A, Roovers RC, Dolezal O, Kortt AA, Hoogenboom HR, Hudson PJ (2001) Design and application of diabodies, triabodies and tetrabodies for cancer targeting. J Immunol Methods 248:47–66

10. Tomlinson I, Holliger P (2000) Methods for generating multivalent and bispecific antibody fragments. Methods Enzymol 326:461–479

11. Carter P (2001) Improving the efficacy of antibody-based cancer therapies. Nat Rev Cancer 1:118–129

12. Hudson PJ, Souriau C (2001) Recombinant antibodies for cancer diagnosis and therapy. Expert Opin Biol Ther 1:845–855

13. Casey JL, Napier MP, King DJ, Pedley RB, Chaplin LC, Weir N, Skelton L, Green AJ, Hope-Stone LD, Yarranton GT, Begent RH (2002) Tumour targeting of humanised cross-linked divalent-Fab′ antibody fragments: a clinical phase I/II study. Br J Cancer 86:1401–1410

14. Willuda J, Kubetzko S, Waibel R, Schubiger PA, Zangemeister-Wittke U, Plückthun A (2001) Tumor targeting of mono-, di-, and tetravalent anti-p185(HER-2) miniantibodies multimerized by self-associating peptides. J Biol Chem 276:14385–14392

15. Plückthun A, Pack P (1997) New protein engineering approaches to multivalent and bispecific antibody fragments. Immunotechnology 3:83–105

16. Kortt AA, Dolezal O, Power BE, Hudson PJ (2001) Dimeric and trimeric antibodies: high avidity scFvs for cancer targeting. Biomol Eng 18:95–108

17. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) The protein data bank. Acta Crystallogr D 58:899–907

18. Amit AG, Mariuzza RA, Phillips SE, Poljak RJ (1986) Three-dimensional structure of an antigen–antibody complex at 2.8 Å resolution. Science 233:747–753

19. Padlan EA, Silverton EW, Sheriff S, Cohen GH, Smith-Gill SJ, Davies DR (1989) Structure of an antibody–antigen complex: crystal structure of the HyHEL-10 Fab–lysozyme complex. Proc Natl Acad Sci USA 86:5938–5942

20. Braden BC, Souchon H, Eiselé JL, Bentley GA, Bhat TN, Navaza J, Poljak RJ (1994) Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. J Mol Biol 243:767–781

21. Davies DR, Cohen GH (1996) Interactions of protein antigens with antibodies. Proc Natl Acad Sci USA 93:7–12

22. Li Y, Li H, Yang F, Smith-Gill SJ, Mariuzza RA (2003) X-ray snapshots of the maturation of an antibody response to a protein antigen. Nat Struct Biol 10:482–488

23. Braden BC, Fields BA, Poljak RJ (1995) Conservation of water molecules in an antibody–antigen interaction. J Mol Recognit 8:317–325

24. Colman PM, Laver WG, Varghese JN, Baker AT, Tulloch PA, Air GM, Webster RG (1987) Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. Nature 26:358–363

25. Zhou T, Xu L, Dey B, Hessell AJ, Van Ryk D, Xiang SH, Yang X, Zhang MY, Zwick MB, Arthos J, Burton DR, Dimitrov DS, Sodroski J, Wyatt R, Nabel GJ, Kwong PD (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. Nature 445:732–737

26. Lok SM, Kostyuchenko V, Nybakken GE, Holdaway HA, Battisti AJ, Sukupolvi-Petty S, Sedlak D, Fremont DH, Chipman PR, Roehrig JT, Diamond MS, Kuhn RJ, Rossmann MG (2008) Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. Nat Struct Mol Biol 15:312–317

27. Mylvaganam SE, Paterson Y, Getzoff ED (1998) Structural basis for the binding of an anti-cytochrome c antibody to its antigen: crystal structures of FabE8-cytochrome c complex to 1.8 A resolution and FabE8 to 2.26 A resolution. J Mol Biol 281:301–322

28. Huang M, Syed R, Stura EA, Stone MJ, Stefanko RS, Ruf W, Edgington TS, Wilson IA (1998) The mechanism of an inhibitory antibody on TF-initiated blood coagulation revealed by the crystal structures of human tissue factor, Fab 5G9 and TF.G9 complex. J Mol Biol 275:873–894

29. Prasad L, Waygood EB, Lee JS, Delbaere LT (1998) The 2.5 Å resolution structure of the jel42 Fab fragment/HPr complex. J Mol Biol 280:829–845

30. Sivasubramanian A, Maynard JA, Gray JJ (2008) Modeling the structure of mAb 14B7 bound to the anthrax protective antigen. Proteins 70:218–230

31. Ni M, Yu B, Huang Y, Tang Z, Lei P, Shen X, Xin W, Zhu H, Shen G (2008) Homology modelling and bivalent single-chain Fv construction of anti-HepG2 single-chain immunoglobulin Fv fragments from a phage display library. J Biosci 33:691–697

32. Heng CK, Othman RY (2009) Bioinformatics in molecular immunology laboratories demonstrated: modeling an anti-CMV scFv antibody. Bioinformation 1:118–120

33. Hink MA, Griep RA, Borst JW, van Hoek A, Eppink MH, Schots A, Visser AJ (2000) Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. J Biol Chem 275:17556–17560

34. Sutton BJ (1988) Antigen recognition by B cells: antibody–antigen interactions at the atomic level. Immunol Suppl 1:31–34

35. Chothia C, Novotný J, Bruccoleri R, Karplus M (1985) Domain association in immunoglobulin molecules. The packing of variable domains. J Mol Biol 186:651–663

36. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201

37. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815

38. Accelrys Software, Inc. (2007) Discovery Studio Version 2.0. http://www.accelrys.com/products/discovery-studio/

39. NIH MBI Laboratory for Structural Genomics and Proteomics (2010) SAVES (Stuctural Analysis and Verification Server). http://nihserver.mbi.ucla.edu/SAVES/

40. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

41. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36:W197–W201

42. Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52:80–87

43. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20:45–50

44. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. Sci STKE pl2

45. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267:383–386

46. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

47. Konstantakaki M, Tzartos SJ, Poulas K, Eliopoulos E (2007) Molecular modeling of the complex between Torpedo acetylcholine receptor and anti-MIR Fab198. Biochem Biophys Res Commun 356:569–575

48. Leysath CE, Monzingo AF, Maynard JA, Barnett J, Georgiou G, Iverson BL, Robertus JD (2009) Crystal structure of the engineered neutralizing antibody M18 complexed to domain 4 of the anthrax protective antigen. J Mol Biol 387:680–693

49. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D 60:2256–2268

50. Berzofsky JA (1985) Intrinsic and extrinsic factors in protein antigenic structure. Science 229:932–940

51. Chothia C, Janin J (1975) Principles of protein–protein recognition. Nature 256:705–708

52. Webster DM, Henry AH, Rees A (1994) Antibody–antigen interactions. Curr Opin Struct Biol 4:123–129

53. Burmester J, Spinelli S, Pugliese L, Krebber A, Honegger A, Jung S, Schimmele B, Cambillau C, Plückthun A (2001) Selection, characterization and X-ray structure of anti-ampicillin single-chain Fv fragments from phage-displayed murine antibody libraries. J Mol Biol 309:671–685

54. Carmichael JA, Power BE, Garrett TP, Yazaki PJ, Shively JE, Raubischek AA, Wu AM, Hudson PJ (2003) The crystal structure of an anti-CEA scFv diabody assembled from T84.66 scFvs in V (L)-to-V(H) orientation: implications for diabody flexibility. J Mol Biol 326:341–351

55. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z (2003) Protein–protein docking benchmark, version 3.0. Proteins 73:705–709

56. Muller YA, Chen Y, Christinger HW, Li B, Cunningham BC, Lowman HB, de Vos AM (1998) VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 A resolution and mutational analysis of the interface. Structure 6:1153–1167

57. Fleury D, Wharton SA, Skehel JJ, Knossow M, Bizebard T (1998) Antigen distortion allows influenza virus to escape neutralization. Nat Struct Biol 5:119–123