



Venom allergen-like protein 28 in *Clonorchis sinensis*: four epitopes on its surface and the potential role of Cys124 for its conformational stability

Myoung-Ro Lee¹ · Won Gi Yoo^{1,2} · Yu Jung Kim¹ · Eun Ju Chung¹ · Shin-Hyeong Cho¹ · Jung-Won Ju¹ 

Received: 25 September 2017 / Accepted: 23 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Venom allergen-like (VAL) proteins are important to host-parasite interactions. We previously demonstrated that a *Clonorchis sinensis* VAL (CsVAL) protein-derived synthetic peptide suppresses allergic and inflammatory responses. However, little is known regarding the physicochemical and antigenic properties of CsVAL proteins. Here, we identified a novel 194 amino acid VAL protein, named *C. sinensis* VAL 28 (CsVAL28), and characterized its functional motifs and structural details as a new member of the CAP superfamily. Unlike members of the *Schistosoma mansoni* VAL (SmVAL) family, CsVAL28 has a single CAP1 motif and six highly conserved disulfide bond-forming cysteines. Tertiary models of wild-type CsVAL28 and mutants were built using SmVAL4 as template via homology modeling. Normal mode analysis predicted that disulfide bond breaking by mutation of cysteine 124 to serine would greatly affect protein mobility. Four major immunoreactive linear epitopes were identified in the surface-exposed region or its vicinity via epitope mapping, using sera from clonorchiasis patients and healthy controls. Our findings provide in-depth knowledge on the structure-function properties of VAL proteins and may help determine highly antigenic regions for developing new diagnostic approaches.

Keywords *Clonorchis sinensis* · Antigenicity · Epitope mapping · Disulfide bond · Structure · Normal mode analysis

Introduction

Clonorchis sinensis is a foodborne zoonotic trematode that infects humans, causing clonorchiasis through the

consumption of raw freshwater fish infected with metacercariae. It is estimated that more than 1.3 million people in Korea and more than 12.5 million people in China have been infected with *C. sinensis* (Hong and

Myoung-Ro Lee and Won Gi Yoo are equal contributors

Section editor: Sabine Specht

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00436-018-5941-4>) contains supplementary material, which is available to authorized users.

✉ Jung-Won Ju
jupapa@korea.kr

Myoung-Ro Lee
blackcg@korea.kr

Won Gi Yoo
wongi.yoo@gmail.com

Yu Jung Kim
hoiyui@korea.kr

Eun Ju Chung
kakkakk@korea.kr

Shin-Hyeong Cho
jo4u@cdc.go.kr

¹ Division of Vectors and Parasitic Diseases, Center for Laboratory control of Infectious Diseases, Korea Centers for Disease Control and Prevention, Chungbuk 28159, Republic of Korea

² Present address: Department of Medical Environmental Biology, Chung-Ang University College of Medicine, Seoul 06974, Republic of Korea

Fang 2012). Dwelling in bile ducts, *C. sinensis* induces an inflammatory reaction and causes epithelial hyperplasia, periductal fibrosis, dyspepsia, and sometimes cholangiocarcinoma (Bouvard et al. 2009; Choi et al. 2006). *C. sinensis* has been reported as a biocarcinogen to humans by IARC (International Agency for Research on Cancer) (Bouvard et al. 2009).

Definitive diagnosis of clonorchiasis is still based on the discovery of eggs in stool samples, using strategies such as the Kato-Katz method and the formalin-ether sedimentation technique (Hong et al. 2003). However, stool examination is insufficient because only a few *C. sinensis* eggs can be found in the stool of patients with early-stage or light infection (Choi et al. 2006). Moreover, this method is not suitable for large-scale examination because of the time- and labor-intensive tasks involved. Enzyme-linked immunosorbent assay for the detection of antibodies against *C. sinensis* antigens represents an alternative diagnostic approach. Several recombinant proteins have been reported that might facilitate the use of this assay, such as legumain (Ju et al. 2009), 14-3-3 protein, propionyl-CoA carboxylase (Lee et al. 2012a), and α -N-acetylgalactosaminidase (Lee et al. 2012b). Nonetheless, further expansion of the antigen candidate repertoire is necessary, since the use of multiple antigens allows more accurate diagnosis and effective vaccination against *C. sinensis* (Rappuoli and Aderem 2011).

Venom allergen-like (VAL) proteins belong to the CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins) superfamily, which is also termed the SCP/TAPS (sperm coating protein/Tpx-1/Ag5/Pr-1/Sc7) superfamily (Gibbs et al. 2008). These proteins play important roles in host-parasite interactions. In *Schistosoma mansoni*, the excretory-secretory VAL 9 protein modulates expression of host extracellular matrix-remodeling genes, which is necessary for processes such as intra-mammalian egg translocation, miracidia infection, and intra-molluscan sporocyst development/migration (Yoshino et al. 2014). Moreover, several VAL proteins have been recognized as good antigen candidates for the early diagnosis of clonorchiasis, since they are up-regulated during the infection of the mammalian host, as shown by a previous study investigating the gene expression patterns at each larval stage (Parker-Manuel et al. 2011). In the nematode *Heligmosomoides polygyrus*, VAL proteins were identified as its dominant excretory-secretory products (Hewitson et al. 2011).

In the current study, we identified *C. sinensis* venom allergen-like protein 28 (CsVAL28) as a new member of the VAL family and investigated the features of CAP motifs and the role of disulfide bonds in maintaining structural stability. Epitope mapping was performed to identify major immunoreactive epitopes.

Materials and methods

Sera

Serum samples were collected from 15 patients with clonorchiasis and 5 healthy individuals. Patient sera were ranked according to the number of eggs per gram of stool. The study protocol involving serum specimens was approved by the Institutional Review Board of KCDC (Permit No. 2015-02-EXP-05-3C-A). All adult participants provided informed consent. All personal identifiers and patient information were delinked from the serum specimens.

Analyses of sequence and phylogeny

The physicochemical properties of the predicted protein, including molecular weight and isoelectric point (pI), were calculated by ProtParam (Wilkins et al. 1999). The canonical and non-classical signal peptide were predicted using SignalP v4.1 (Petersen et al. 2011) and SecretomeP (Bendtsen et al. 2005), respectively. Multiple sequence alignment was performed with MAFFT v7.294 using the L-INS-i (Katoh and Standley 2013), and its results were displayed by Jalview (Waterhouse et al. 2009). The evolutionary history was inferred by the maximum likelihood (ML) method using the Molecular Evolutionary Genetic Analysis (MEGA) v6.06 (Kumar et al. 2016). All positions containing gaps and missing data were eliminated. To confirm sequences belonging to the CAP subfamily, CAP motifs were searched for in all sequences, using an in-house Python script.

Homology models and features of wild-type and mutant CsVAL28

LOMETS (LOcal MEta-Threading-Server) (Wu and Zhang 2007) was used for searching for a suitable template, *S. mansoni* VAL 4 (SmVAL4, PDB ID: 4P27_A). The disulfide-bonding cysteines were explored using the “Disulfide of Design 2” (Craig and Dombkowski 2013). Each cysteine was individually substituted with serine based on the target and template alignment, using a text editor (Guo and Feng 2001). The 3D models of wild-type (WT) CsVAL28 and mutants (C75S, C124S, C128S, C144S, C149S, and C160S) were generated using the “Target-Template Alignment” mode of Swiss-Model (Biasini et al. 2014). COACH (CONsensus ApproaCH) (Yang et al. 2013) was used to predict Zn²⁺-binding sites based on the identification of analogs with similar sites. The electrostatic potentials on the surface of the WT CsVAL28 model were calculated with the Adaptive Poisson–Boltzmann Solver and PDB2PQR (Unni et al. 2011). Models were investigated and visualized using UCSF CHIMERA (Huang et al. 2014).

Structural quality assessment

Structural stereochemical qualities of the WT CsVAL28 3D model were evaluated using Ramachandran plots (Lovell et al. 2003) and ERRAT (Colovos and Yeates 1993) from PROCHECK (Laskowski et al. 1996) and ModFOLD6 (McGuffin et al. 2013). In mutants, quality evaluations were not performed because those were manipulated with the purpose of breaking the disulfide bonds. The root mean square deviation (RMSD) of all 3D models was calculated using the ResQ (Yang et al. 2016).

Mobility via normal mode analysis

Normal mode analysis (NMA), which provides insights into large-scale mobility, was performed for WT CsVAL28 and mutants using the iMODS (Internal coordinates normal MODE analysis Server) (Lopez-Blanco et al. 2014) with all parameters set to the default values. The 3D models were rendered in different colors based on B-factor values (as inferred by NMA) by turning the arrow field toggle on.

Cloning of the CsVAL28-coding sequence

The gene encoding the CsVAL28 protein was identified from an adult worm expressed sequence tag (EST) (Kim et al. 2014; Yoo et al. 2011). Polymerase chain reaction (PCR) was performed to identify transcripts containing the open reading frame of CsVAL28 from adult *C. sinensis*. The forward and reverse primers for CsVAL28 amplification were 5'-TTT GAC GGA TCC ATG ATC AGG GAA GGG TGA G-3' and 5'-GGG CCC AAG CTT TTA CAA ACA AAT GCA TGG C-3', respectively, and contained *BamH I* and *Hind III* restriction sites (underlined bases), respectively.

The gene coding for CsVAL28 was identified from an adult worm EST library by PCR, using the primers mentioned above. The amplified PCR product was purified, digested with endonucleases, and run on a 1% agarose gel. The bands corresponding to the cDNA fragments were excised and the fragments were ligated into the bacterial expression vector pRSETA (Novagen, Madison, WI, USA) to code for CsVAL28 fused with a 6×His tag in its N-terminal. Successful cloning was confirmed by DNA sequencing and the vector was transformed into *Escherichia coli* BL21 (DE3) (Invitrogen, Carlsbad, CA, USA). The cDNA sequence of

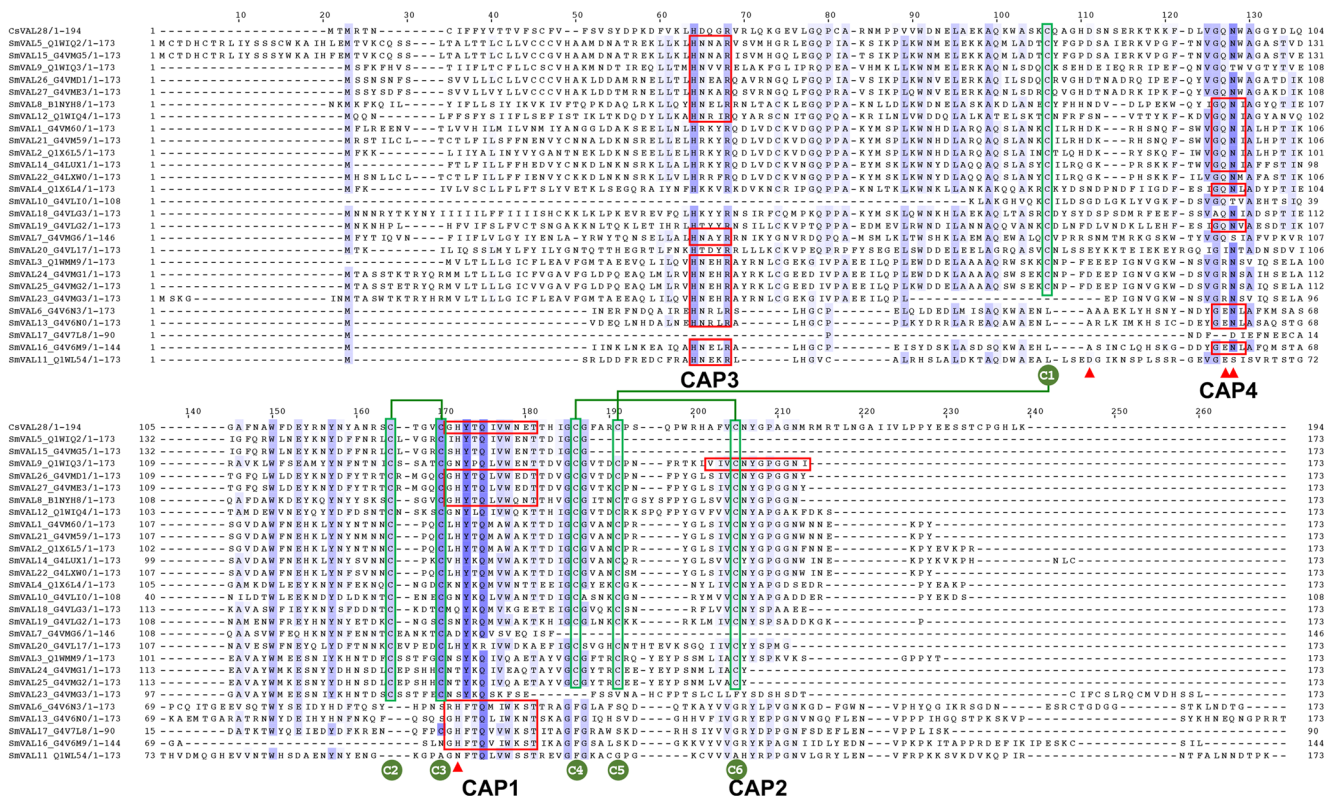


Fig. 1 Homology analysis of CsVAL28 and various to SmVAL members. Residues that are 75% conserved among the VAL proteins are shaded in *light blue*, and those that are 100% conserved are highlighted in *dark blue*. Residues that are necessary for the coordination of the Zn^{2+} in CsVAL28 are marked with *triangles*.

Accession IDs were retrieved from the UniProtKB/Swiss-Prot database (UniProt 2015). CAP motifs were boxed in *red* and their names were added below. *Green-circled C's* represent cysteine residues and the three disulfide bonds are as follows: C1–C5, C2–C3, and C4–C6

CsVAL28 was deposited in the GenBank under accession ID MH172541.

Peptide microarray

To map the epitopes of CsVAL28, peptides were spotted onto microarrays and used to screen the reactivity against sera from patients with clonorchiasis or healthy subjects. Each peptide comprised a 15-mer and its sequence was offset from the one of the previous peptide by 4 amino acids. A custom peptide array was manufactured by JPT Peptide Technologies (Berlin, Germany), wherein all 15-mer peptides were displayed in 3 identical sub-arrays on each glass slide. The slides were incubated with blocking buffer (Superblock TBS T20, Thermo Fisher Scientific Inc. Rockford, IL, USA), washed with Tris-buffered saline (TBS), and then incubated with sera (1:100 dilution) from patients with clonorchiasis or healthy subjects at 4 °C overnight. After washing with TBS, the slides were incubated with 0.1 µg/mL Cy5-labeled mouse anti-human IgG (#209-175-082; Jackson ImmunoResearch Products, Bar Harbor, ME, USA), and consecutively washed with TBS and deionized water (5 times in each for 5 min). The slides were dried by centrifugation at 500×g for 1 min. The fluorescence scan was performed with an Axon GenePix Scanner 4300A (Molecular Devices, Sunnyvale, CA, USA). Data were analyzed using the GenePix Pro 6.0 software (Molecular Devices) and the mean signal intensity for each spot was extracted between 0 and 65,535 arbitrary units (a.u.).

Signals were normalized by subtracting signals obtained from internal control peptides. Signal relative intensities were calculated by dividing the mean intensity of each peptide in the clonorchiasis group by the mean intensity of the same peptide in the healthy group, and the final data were visualized using Heatmap Illustrator (HemI) v1.0.3 (Deng et al. 2014).

Results and discussion

Sequence features

An isolated cDNA clone was found to code for a novel VAL protein, named CsVAL28. The 3'-incomplete coding sequence of CsVAL28 was 582 bp in length and coded for a putative protein of 194 amino acids with a calculated molecular mass 22.09 kDa and a theoretical pI of 8.67. CsVAL28 is possibly secreted through non-classical route. Typical VAL proteins contain a CAP domain (Pfam ID: PF00188). However, similar to SmVAL4 (Kelleher et al. 2014), CsVAL28 does not contain a prototypical CAP domain since its amino acid (aa) sequence does not align with the aa1–8 region of a Hidden Markov Model (HMM) profile of the domain.

CsVAL28 as a new CAP superfamily member

To determine the phylogenetic relationship of CsVAL28 with the SmVAL subfamily, 27 proteins were collected from the

Fig. 2 Phylogenetic relationship between CsVAL28 and the SmVAL family. **a** Phylogenetic tree constructed by performing sequence alignment between CsVAL28 (closed circle) and SmVAL family members to infer evolutionary relationships. Bootstrap values (1000 replicates) are shown next to the branches. The scale bar represents amino acid substitutions. UniProtKB/Swiss-Prot accession IDs are given in parentheses. **b** CAP motif organization in CsVAL28 and members of the SmVAL family. CAP1, yellow box; CAP2, blue box; CAP3, orange box; CAP4, green box

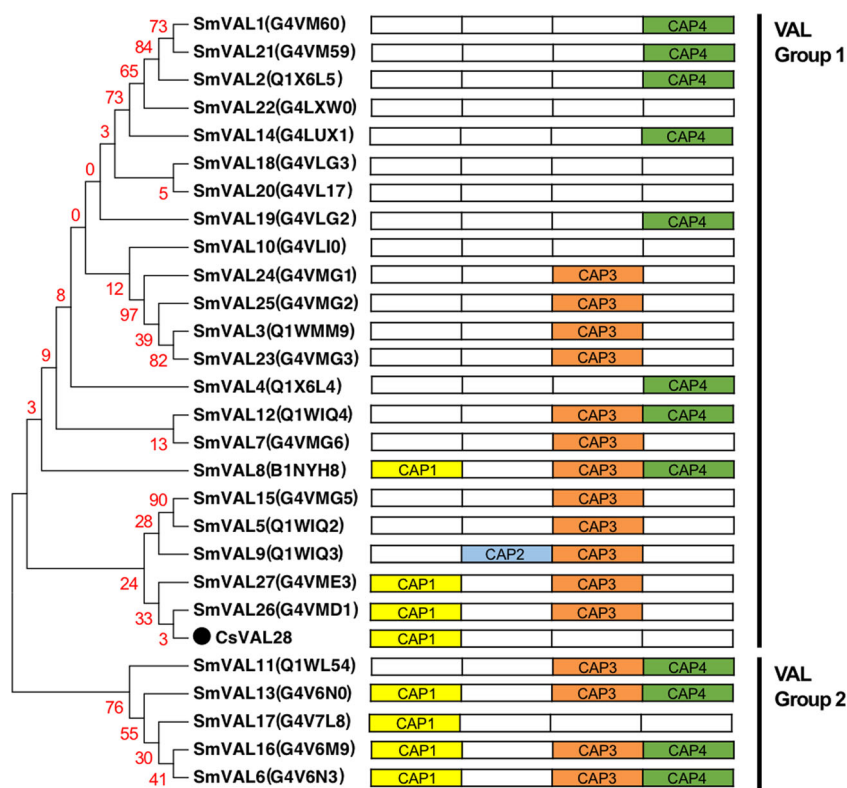
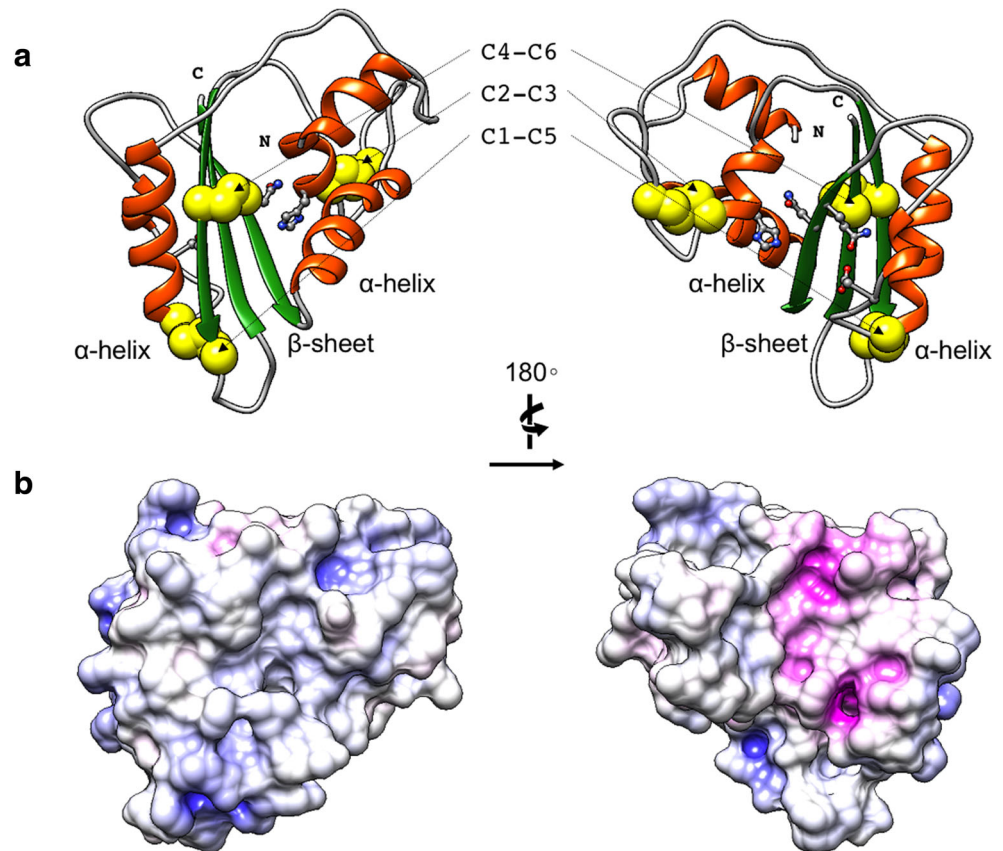


Fig. 3 Ribbon and surface representations of the CsVAL28 3D model. **a** Overall structure; α -helices and β -strands are depicted as *orange* and *green ribbon* diagrams, respectively, and coiled-coil as a *gray line*. Residues that bind Zn^{2+} are shown in *ball-and-stick*, and disulfide bonds are indicated by C1–C5, C2–C3, and C4–C6. **b** Electrostatic potential map of the surface of the CsVAL28 model. Positively and negatively charged regions are colored *blue* and *pink*, respectively. The two faces of CsVAL28 are presented by a vertical rotation of 180°



UniProtKB/Swiss-Prot database (UniProt 2015). Multiple sequence alignment revealed the presence of six highly conserved cysteine residues (C1 to C6), which potentially form three disulfide bonds (C1–C5, C2–C3, and C4–C6) in CsVAL28 (Fig. 1). Chalmers et al. reported that VAL proteins in platyhelminths can be divided into two distinct groups (Chalmers and Hoffmann 2012), with the presence of an

extended first loop region and all six disulfide-bond-forming cysteines in Platyhelminth VAL Group 1 differentiating it from VAL Group 2. On the other hand, according to the study of Cantacessi et al. (2012), the occurrence and conservation of cysteine residues at specific sequence positions should be used as the key criteria for the classification of trematode SCP/TAPS proteins. According to this system, CsVAL28 is very close to the Trematode single-domain SCP/TAPS Group 2, which contains a CAP domain and six highly conserved cysteine residues.

Members of the VAL subfamily are characterized by a pattern of four CAP motifs that have been evolutionarily conserved across diverse species (Gibbs et al. 2008): CAP1, [GDER][HR][FYWH][TVS][QA][LIVM][LIVMA]Wxx[ST-N]; CAP2, [LIVMFYH][LIVMFY]xC[NQRHS]Yx[PARH]x[GL]N[LIVMFYWDN]; CAP3, HNxxR; CAP4, G[EQ]N[ILV] (x = any amino acid). Our analysis grouped CsVAL28 with SmVAL26 (182 aa) and SmVAL27 (182 aa), with both of which it shared 34.8% identity, even though SmVAL26 and SmVAL27 have an additional CAP3 motif (Fig. 2). SmVAL17 (90 aa), albeit harboring an exclusive CAP1 motif, showed low identity (11.9%) with CsVAL28 and was distinct from the VAL Group 1, as it had no conserved cysteines. Therefore, CsVAL28 was characterized as a new CAP superfamily member that belongs to the Platyhelminth VAL Group 1. In the Pfam database, classification of CAP proteins is based on the distribution of CAP domains and the

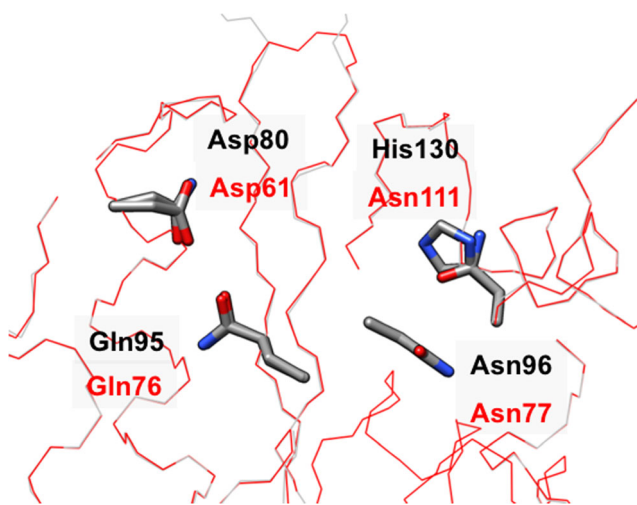


Fig. 4 Zn^{2+} -binding residues of CsVAL28. The superposed residues that form the binding site are shown in *stick* mode and labeled *black* for *C. sinensis* and *red* for *S. mansoni* (PDB ID: 4P27_A). Side-chain oxygen and hydrogen atoms are colored *red* and *white*, respectively

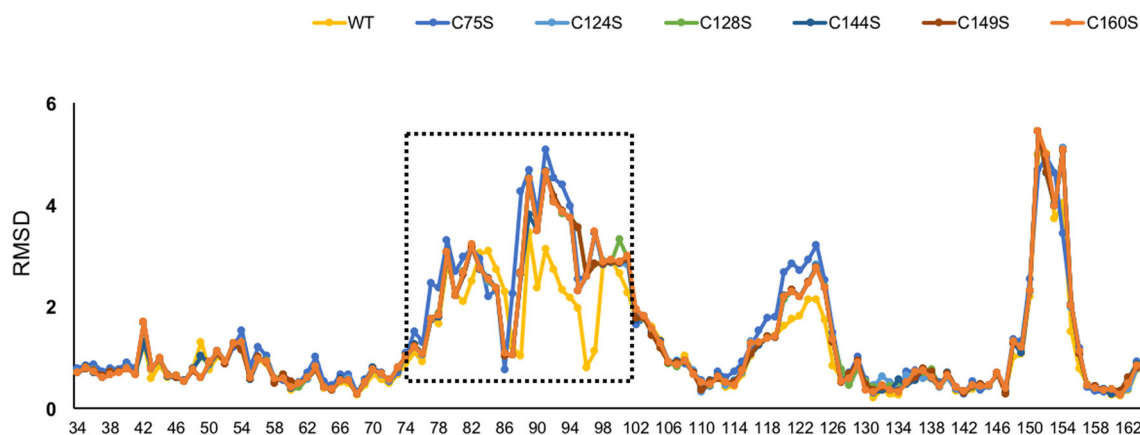


Fig. 5 Residual flexibility of wild-type (WT) CsVAL28 and its mutant variants. The regions around Lys74 and Asp102 with high RMSD values are marked by dashed boxes

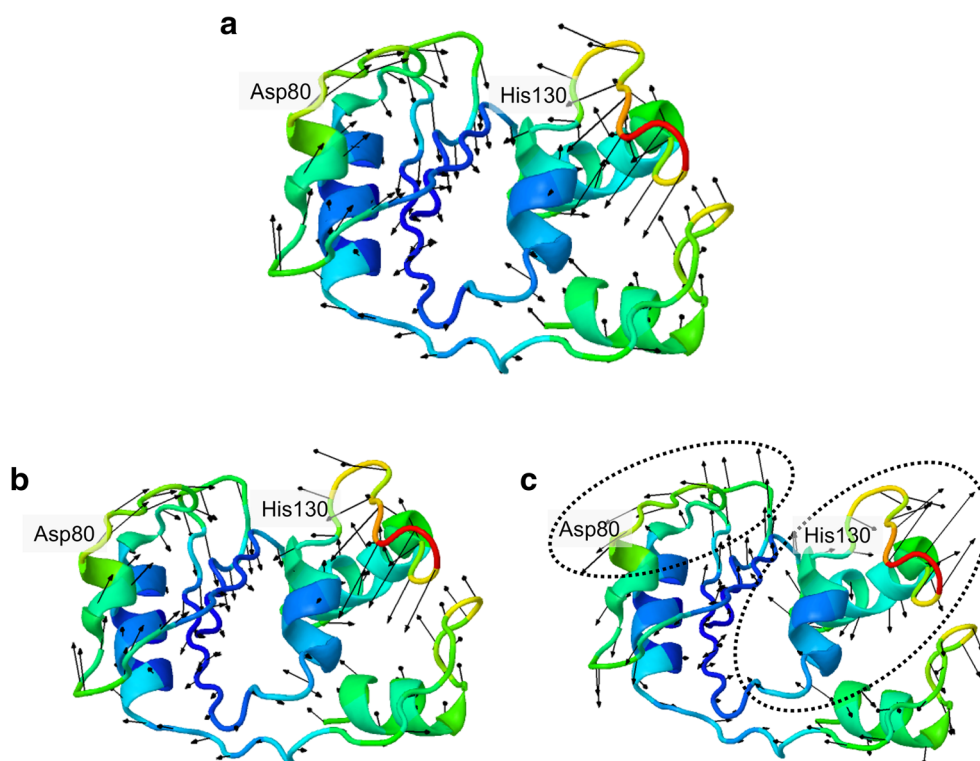
conservation of CAP motifs. In the case of the pathogenic nematode *Haemonchus contortus*, all CAP proteins had a conserved CAP4 motif but variable remaining CAP motifs (Mohandas et al. 2015). In trematodes, the frequency of CAP motifs is high in the case of CAP3 and CAP4, moderate for CAP1, and low for CAP2. These findings are in line with the results of Philippsen et al., who showed that a selective expansion of the VAL gene family occurred in different *Schistosoma* species (Philippsen et al. 2015).

Homology modeling and validation

SmVAL4 (PDB ID: 4P27_A) (Kelleher et al. 2014) was selected as the best template by the LOMETS, as seen in Table

S1 (Wu and Zhang 2007). The identity and coverage between the target and template sequences were 36 and 66% (from His34 to Pro164), respectively. Since the identity score was in the 25–50% range, the model can be used for studying mutants (Franca 2015) (see subsection “Disulfide bonds and their role in maintaining structural stability”). The 3D model of CsVAL28 was predicted based on the template, using the Swiss-Model server (Biasini et al. 2014). The quality of the model was verified by constructing a Ramachandran plot (Lovell et al. 2003), which showed that 87.4% of residues lay in the favorable region, 9.9% in the additionally allowed region, only 2.7% (Asp47, Lys55, and His123) in the generously allowed region, and none in the disallowed region (Fig. S1A). The overall quality score was calculated by ERRAT

Fig. 6 Comparison of point mutations with respect to their effect on the mobility of CsVAL28. The direction is given by the arrow, while the length of the line represents the magnitude of mobility in 3D models of wild-type (WT) CsVAL28 (a), the C75S mutant (b), and the C124S mutant (c). Increasing magnitudes of B-factor values are color-coded from blue to red. The Zn^{2+} -binding residues, Asp80 and His130, are displayed. Dashed circles in the C124S mutant model indicate movement in opposite direction compared to WT



(Colovos and Yeates 1993) at 79.68, suggesting a good 3D model (Fig. S1B). The global model quality score of 0.49, as calculated by ModFOLD6 (McGuffin et al. 2013), ensured that the model was properly folded, and the 3D cartoon of the model was colored according to the per-residue error considering the B-factor values (Fig. S1C). No further refinement of the CsVAL28 3D model was performed to keep the three disulfide bonds intact.

Overall structural features

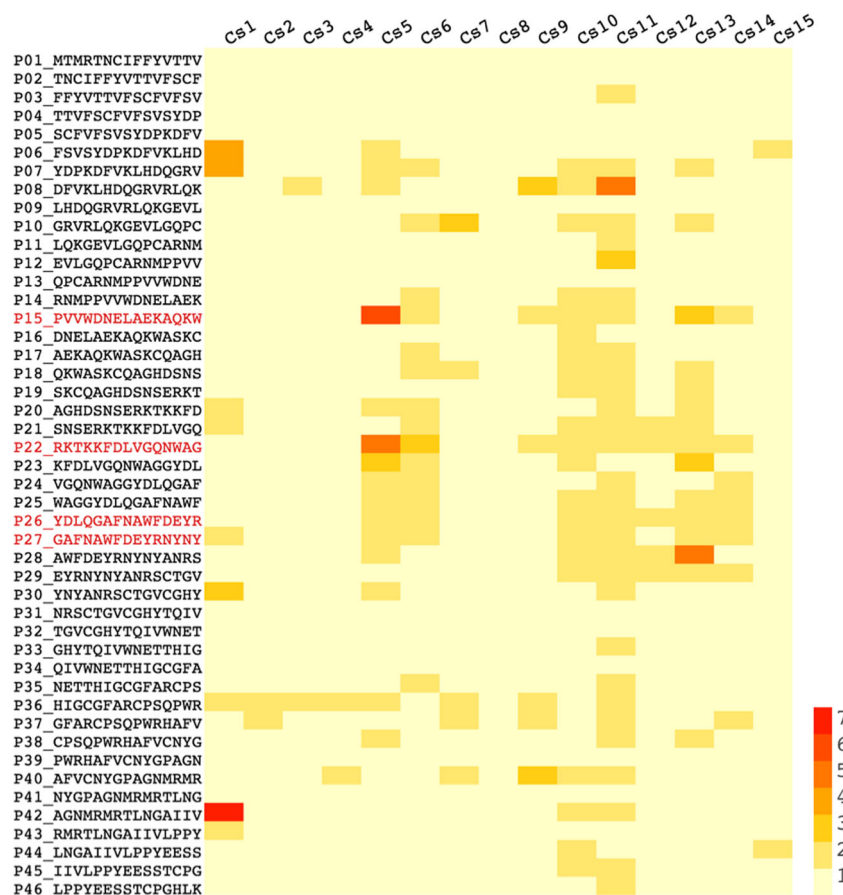
Structure examination revealed a core helix-sheet-helix (α - β - α) sandwich fold, which is a characteristic of members of the CAP superfamily, including VAL proteins (Gibbs et al. 2008) (Fig. 3a). Electronegative regions were found in the center of the Zn^{2+} -binding site (Fig. 3b). Superposition of CsVAL28 with SmVAL4 structures showed that the structural coordinates of the Zn^{2+} -binding site were highly conserved, except for the position of His130 (Fig. 4). However, His130 has also been predicted as a conserved residue for Zn^{2+} binding in CRISPs (cysteine-rich secretory proteins) by COACH, based on the identification of analogs with similar binding sites (Yang et al. 2013).

Disulfide bonds and their role in maintaining structural stability

Several highly conserved cysteines are found in different groups of VAL proteins and participate in intrachain and inter-chain disulfide bonds that are critical for maintaining protein structure stability (Cantacessi et al. 2012; Chalmers and Hoffmann 2012). We examined the importance of the three putative disulfide bonds (Cys75-Cys149, Cys124-Cys128, and Cys144-Cys160) for the structure of CsVAL28 by using the “Disulfide function of Design 2” (Craig and Dombkowski 2013). Specifically, each cysteine was individually substituted with serine and we evaluated which point mutation affects the protein stability, as previously suggested (Guo and Feng 2001). Then, 3D models of WT CsVAL28 and mutants (C75S, C124S, C128S, C144S, C149S, and C160S) were generated and compared. When comparing the WT with mutant models, RMSD values were remarkably different in the aa74–102 region (Fig. 5). Among mutants, differences of residual and overall RMSD were highest in C124S and C75S, respectively (Table S2).

All 3D models were subjected to NMA using iMODS (Lopez-Blanco et al. 2014) to determine which mutation has a greater impact on protein mobility. NMA allows the accurate

Fig. 7 Screening of antigenic peptides via epitope mapping. In the array, peptides with an intensity fold ratio (clonorchiasis sera to control sera) ranging 1 to 7 were colored from *light yellow* to *red*. Among the entire CsVAL28 sequence, four peptides (marked in *red*) were found to be highly antigenic: $_{57}$ PVVWDNELA EKAQKW $_{71-85}$ RKTKKFDLV GQNWAG $_{99-101}$ YDLQGAFN AWFDEYR $_{115}$, and $_{105}$ GAFN AWFDEYRN $_{119}$



characterization of the direction and extent of the inherent motions of the structure much faster than molecular dynamics (MD) simulations and iMODS has been widely used for assessing how mutations affect structural dynamics (Awan et al. 2017; Prabhakar et al. 2016). The results for WT, C124S (the mutation with the highest residual RMSD), and C75S (the mutation with the highest overall RMSD) are displayed in Fig. 6, whereas the results for the rest of the mutations can be found in Fig. S2. In the WT model, the residual mobility of Asp80 and His130 seems to be oriented toward each other, around the Zn^{2+} -binding site (Fig. 6a). In the C124S model, the mobility of the two residues shifted in the opposite direction, which seems to open the box (Fig. 6c). The results of the remaining mutants were similar with those of WT (Figs. 6b and S2). Therefore, Cys124 out of the six highly conserved cysteines was predicted to have a significant impact on protein mobility. Future, molecular-level investigation of the effects of disrupting this disulfide bond by substitution of the conserved cysteines should enrich our understanding of the structure-function properties of VAL proteins.

Identification of a highly antigenic region via epitope mapping

VAL proteins are considered potentially good antigens since they are completely exposed to the host's immune system as excretory-secretory products (Chalmers et al. 2008). The identification of linear epitopes bound by antibodies is crucial for scrutinizing the antigenicity and immunogenicity of native antigens and especially for developing synthetic multi-chimeric antigens that could be utilized in immunodiagnostic tests (Noya et al. 2003). Thus, epitope mapping of CsVAL28 was performed using a peptide microarray that consisted of 46 linear peptide 15-mers overlapping an 11-mer of CsVAL28. We incubated 15 sera from patients with clonorchiasis and 5 healthy sera with the peptide microarray (Fig. 7). The relative intensities of the signals from the clonorchiasis group compared to the healthy group ranged from 0.6 to 6.8. Among them, the following four peptides had intensities higher by a factor of ≥ 1.2 compared to sera from five or more clonorchiasis patients: peptide no. 15 (P15, $_{57}PVVWDNELAEKAQKW_{71}$), peptide no. 22 (P22, $_{85}RKTKKFDLVGQNWAG_{99}$), peptide no. 26 (P26, $_{101}YDLQGAFNAWFDEYR_{115}$), and peptide

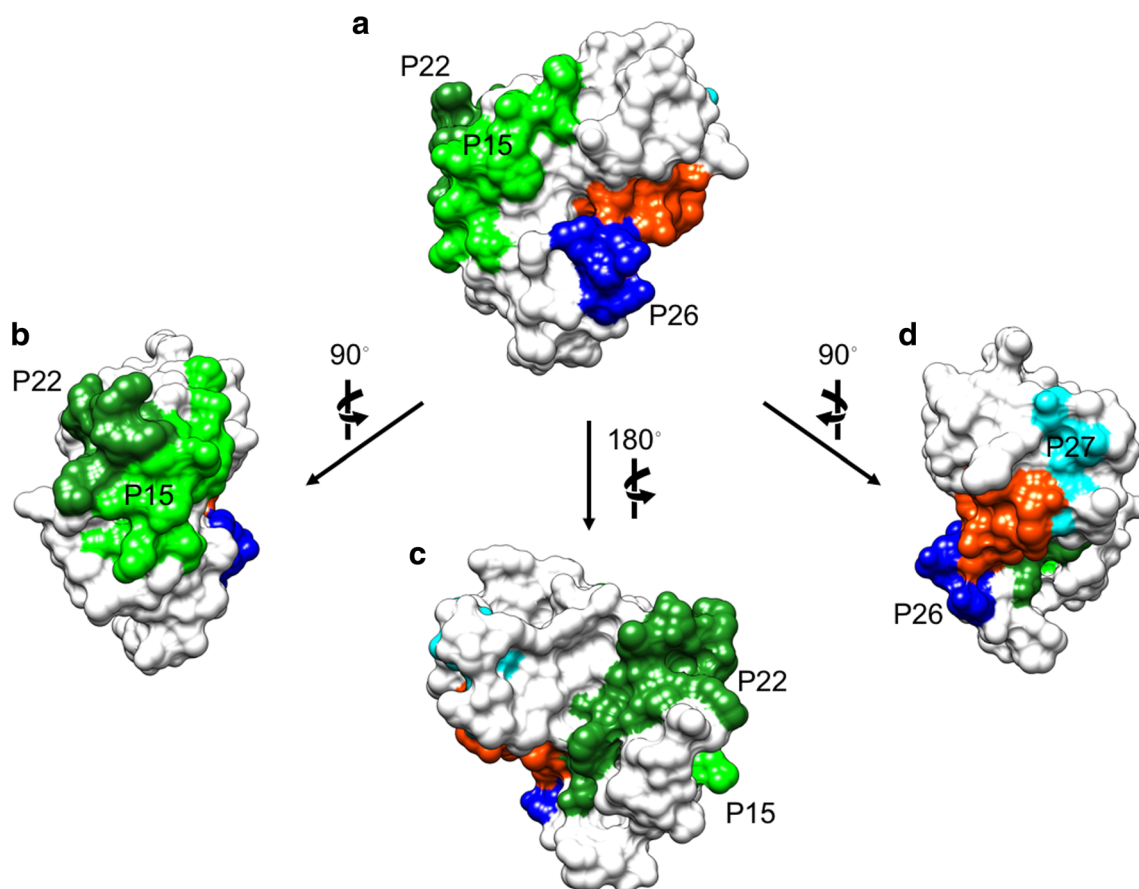


Fig. 8 Surface mapping of linear B-cell epitopes. **a** Surfaces corresponding to the peptides P15 (green), P22 (forest green), P26 (blue), and P27 (cyan) are shown on the 3D structure of WT CsVAL28 while the remaining part of the model is displayed in white. The short sequence

“GAFNAWFDEYR” (orange) overlapped between P26 and P27. Each peptide can be viewed from three different vertical rotations, displayed in **b**, **c**, and **d**

no. 27 (P27, ₁₀₅GAFNAWFDEYRNYNY₁₁₉). These were recognized as linear epitopes, since they exhibited the strongest signals.

It is important to identify conformational B-cell epitopes (also known as discontinuous epitopes), which play an actual role in forming antigen-antibody complex and are believed to amount to more than 90% of B-cell epitopes (Sun et al. 2011). Interestingly, four epitopes were exposed on or near the protein surface and were vulnerable to antibodies (Fig. 8). Furthermore, P26 and P27 are parts of the putative linear B-cell epitope “₉₄GQNWAGGYDLQGAFNAWFDEYRNYNYANRS₁₂₃”, which was demonstrated in one of our previous studies to display an anti-allergic effect when applied on RBL-2H3 cells by inhibiting the activity of the protein kinase C- α and the extracellular signal-regulated kinase 1/2 (Jeong et al. 2014). Thus, both P26 and P27 epitopes may contribute significantly to the immunogenicity of CsVAL28.

Conclusions

We identified CsVAL28, a novel VAL protein expressed by *C. sinensis*, and characterized its functional motifs and structural details as a new member of the CAP superfamily. Comparative sequence alignment with members of the *S. mansoni* VAL family showed that CsVAL28 has only one CAP1 motif and six highly conserved cysteines. We also demonstrated that mutation of cysteine to serine at position 124 is predicted to affect structure mobility. Four highly immunoreactive epitopes were observed in the internal region via epitope mapping on the peptide microarray. Further studies are still needed to confirm the biological functions of CsVAL28, which are associated with adaptation to a definitive host and, by extension, the development of an early diagnostic technique.

Acknowledgments This work was supported by a grant from the Korea National Institute of Health, Korea Centers for Disease Control and Prevention, Ministry of Health and Welfare, Korea (KCDC 4800-4847-311).

References

- Awan FM, Obaid A, Ikram A, Janjua HA (2017) Mutation-structure-function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): a comprehensive informatics approach. *Int J Mol Sci* 18(1):139. <https://doi.org/10.3390/ijms18010139>
- Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005) Non-classical protein secretion in bacteria. *BMC Microbiol* 5:58. <https://doi.org/10.1186/1471-2180-5-58>
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42(Web Server issue):W252–W258. <https://doi.org/10.1093/nar/gku340>
- Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L (2009) A review of human carcinogens—part B: biological agents. *Lancet Oncol* 10(4):321–322
- Cantacessi C, Hofmann A, Young ND, Broder U, Hall RS, Loukas A, Gasser RB (2012) Insights into SCP/TAPS proteins of liver flukes based on large-scale Bioinformatic analyses of sequence datasets. *PLoS One* 7(2):e31164
- Chalmers IW, Hoffmann KF (2012) Platyhelminth venom allergen-like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum. *Parasitology* 139(10):1231–1245
- Chalmers IW, McArdle AJ, Coulson RM, Wagner MA, Schmid R, Hirai H, Hoffmann KF (2008) Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* 9(1):89
- Choi D, Lim JH, Lee KT, Lee JK, Choi SH, Heo JS, Jang K-T, Lee NY, Kim S, Hong S-T (2006) Cholangiocarcinoma and *Clonorchis sinensis* infection: a case-control study in Korea. *J Hepatol* 44(6):1066–1073
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2(9):1511–1519. <https://doi.org/10.1002/pro.5560020916>
- Craig DB, Dombkowski AA (2013) Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* 14:346. <https://doi.org/10.1186/1471-2105-14-346>
- Deng W, Wang Y, Liu Z, Cheng H, Xue Y (2014) HemI: a toolkit for illustrating heatmaps. *PLoS One* 9(11):e111988. <https://doi.org/10.1371/journal.pone.0111988>
- Franca TC (2015) Homology modeling: an important tool for the drug discovery. *J Biomol Struct Dyn* 33(8):1780–1793. <https://doi.org/10.1080/07391102.2014.971429>
- Gibbs GM, Roelants K, O'Bryan MK (2008) The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, Cancer, and immune defense. *Endocr Rev* 29(7):865–897
- Guo ZY, Feng YM (2001) Effects of cysteine to serine substitutions in the two inter-chain disulfide bonds of insulin. *Biol Chem* 382(3):443–448. <https://doi.org/10.1515/BC.2001.054>
- Hewitson JP, Hargus Y, Murray J, van Agtmaal M, Filbey KJ, Grainger JR, Bridgett S, Blaxter ML, Ashton PD, Ashford DA, Curwen RS, Wilson RA, Dowle AA, Maizels RM (2011) Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins. *J Proteome* 74(9):1573–1594
- Hong S-T, Fang Y (2012) *Clonorchis sinensis* and clonorchiasis, an update. *Parasitol Int* 61(1):17–24
- Hong ST, Choi MH, Kim CH, Chung BS, Ji Z (2003) The Kato-Katz method is reliable for diagnosis of *Clonorchis sinensis* infection. *Diagn Microbiol Infect Dis* 47(1):345–347
- Huang CC, Meng EC, Morris JH, Pettersen EF, Ferrin TE (2014) Enhancing UCSF Chimera through web services. *Nucleic Acids Res* 42(Web Server issue):W478–W484. <https://doi.org/10.1093/nar/gku377>
- Jeong Y-I, Kim Y-J, Ju JW, Hong S-H, Lee M-R, Cho S-H, Lee W-J, Lee S-E (2014) Identification of anti-allergic effect of *Clonorchis sinensis*-derived protein venom allergen-like proteins (CsVAL). *Biochem Biophys Res Commun* 445(3):549–555
- Ju JW, Joo HN, Lee MR, Cho SH, Cheun HI, Kim JY, Lee YH, Lee KJ, Sohn WM, Kim DM, Kim IC, Park BC, Kim TS (2009) Identification of a serodiagnostic antigen, legumain, by immunoproteomic analysis of excretory-secretory products of

- Clonorchis sinensis* adult worms. Proteomics 9(11):3066–3078. <https://doi.org/10.1002/pmic.200700613>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. <https://doi.org/10.1093/molbev/mst010>
- Kelleher A, Darwiche R, Rezende WC, Farias LP, Leite LCC, Schneider R, Asojo OA (2014) *Schistosoma mansoni* venom allergen-like protein 4 (SmVAL4) is a novel lipid-binding SCP/TAPS protein that lacks the prototypical CAP motifs. Acta Crystallogr D-Biol Crystallogr 70:2186–2196. <https://doi.org/10.1107/S1399004714013315>
- Kim DW, Yoo WG, Lee S, Lee MR, Kim YJ, Cho SH, Lee WJ, Ju JW (2014) ClonESTdb: a comprehensive database for *Clonorchis sinensis* EST sequences. BMC Res Notes 7:388. <https://doi.org/10.1186/1756-0500-7-388>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477–486
- Lee MR, Kim YJ, Kim DW, Yoo WG, Cho SH, Hwang KY, Ju JW, Lee WJ (2012a) The identification of antigenic proteins: 14-3-3 protein and propionyl-CoA carboxylase in *Clonorchis sinensis*. Mol Biochem Parasitol 182(1–2):1–6. <https://doi.org/10.1016/j.molbiopara.2011.11.006>
- Lee MR, Yoo WG, Kim YJ, Kim DW, Cho SH, Hwang KY, Ju JW, Lee WJ (2012b) Molecular characterization of an alpha-N-acetylgalactosaminidase from *Clonorchis sinensis*. Parasitol Res 111(5):2149–2156. <https://doi.org/10.1007/s00436-012-3063-y>
- Lopez-Blanco JR, Aliaga JI, Quintana-Orti ES, Chacon P (2014) iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res 42(Web Server issue):W271–W276. <https://doi.org/10.1093/nar/gku339>
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C α geometry: phi, psi and C β deviation. Proteins 50(3):437–450. <https://doi.org/10.1002/prot.10286>
- McGuffin LJ, Buenavista MT, Roche DB (2013) The ModFOLD4 server for the quality assessment of 3D protein models. Nucleic Acids Res 41(Web Server issue):W368–W372. <https://doi.org/10.1093/nar/gkt294>
- Mohandas N, Young ND, Jabbar A, Korhonen PK, Koehler AV, Amani P, Hall RS, Sternberg PW, Jex AR, Hofmann A, Gasser RB (2015) The barber's pole worm CAP protein superfamily—a basis for fundamental discovery and biotechnology advances. Biotechnol Adv 33(8):1744–1754. <https://doi.org/10.1016/j.biotechadv.2015.07.003>
- Noya O, Patarroyo ME, Guzman F, Alarcon de Noya B (2003) Immunodiagnosis of parasitic diseases with synthetic peptides. Curr Protein Pept Sci 4(4):299–308
- Parker-Manuel SJ, Ivens AC, Dillon GP, Wilson RA (2011) Gene expression patterns in larval *Schistosoma mansoni* associated with infection of the mammalian host. PLoS Negl Trop Dis 5(8):e1274. <https://doi.org/10.1371/journal.pntd.0001274>
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786. <https://doi.org/10.1038/nmeth.1701>
- Philippens GS, Wilson RA, DeMarco R (2015) Accelerated evolution of schistosome genes coding for proteins located at the host-parasite interface. Genome Biol Evol 7(2):431–443. <https://doi.org/10.1093/gbe/evu287>
- Prabhakar PK, Srivastava A, Rao KK, Balaji PV (2016) Monomerization alters the dynamics of the lid region in campylobacter jejuni CstII: an MD simulation study. J Biomol Struct Dyn 34(4):778–791. <https://doi.org/10.1080/07391102.2015.1054430>
- Rappuoli R, Aderem A (2011) A 2020 vision for vaccines against HIV, tuberculosis and malaria. Nature 473(7348):463–469. <https://doi.org/10.1038/nature10124>
- Sun J, Xu T, Wang S, Li G, Wu D, Cao Z (2011) Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens. Immunome Res 7(3):1–11
- UniProt C (2015) UniProt: a hub for protein information. Nucleic Acids Res 43(Database issue):D204–D212. <https://doi.org/10.1093/nar/gku989>
- Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, Nielsen JE, Baker NA (2011) Web servers and services for electrostatics calculations with APBS and PDB2PQR. J Comput Chem 32(7):1488–1491. <https://doi.org/10.1002/jcc.21720>
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. Methods Mol Biol 112:531–552
- Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res 35(10):3375–3382. <https://doi.org/10.1093/nar/gkm251>
- Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29(20):2588–2595. <https://doi.org/10.1093/bioinformatics/btt447>
- Yang J, Wang Y, Zhang Y (2016) ResQ: an approach to unified estimation of B-factor and residue-specific error in protein structure prediction. J Mol Biol 428(4):693–701. <https://doi.org/10.1016/j.jmb.2015.09.024>
- Yoo WG, Kim DW, Ju JW, Cho PY, Kim TI, Cho SH, Choi SH, Park HS, Kim TS, Hong SJ (2011) Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*. PLoS Negl Trop Dis 5(6):e1208. <https://doi.org/10.1371/journal.pntd.0001208>
- Yoshino TP, Brown M, Wu X-J, Jackson CJ, Ocádiz-Ruiz R, Chalmers IW, Kolb M, Hokke CH, Hoffmann KF (2014) Excreted/secreted *Schistosoma mansoni* venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression. Int J Parasitol 44(8):551–563

Table S1

Top 10 models selected as putative templates using LOMETS server.

Rank	PDB ID	Name	Coverage	Z-score ^a	Identity (%)	Program
1	1XTA_A	Cysteine-rich venom protein 1	0.86	13.26	0.32	MUSTER (Wu and Zhang, 2008)
2	3NT8_A	Na-ASP-1	0.87	134.70	0.23	PRC (Madera, 2008)
3	4P27_A	Venom allergen-like protein 4	0.80	36.18	0.31	SP3 (Soding, 2005)
4	4P27_A	Venom allergen-like protein 4	0.80	31.83	0.31	SPARKS-X (Zhou and Zhou, 2004)
5	4P27_A	Venom allergen-like protein 4	0.79	85.00	0.32	FFAS03 (Jaroszewski et al., 2005)
6	4P27_A	Venom allergen-like protein 4	0.80	23.62	0.32	cdPPAS (Yan et al., 2013)
7	3NT8_A	Na-ASP-1	0.93	37.44	0.32	HHSEARCH (Soding, 2005)
8	4P27_A	Venom allergen-like protein 4	0.80	106.00	0.32	FFAS03 (Jaroszewski et al., 2005)
9	1XTA_A	Cysteine-rich venom protein 1	0.85	30.09	0.31	Neff-PPAS (Yan et al., 2013)
10	4P27_A	Venom allergen-like protein 4	0.79	33.22	0.32	wdPPAS (Yan et al., 2013)

^a Z-score is defined as the difference of the raw score from mean divided by the standard deviation.**References:**

- Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., Godzik, A., 2005. FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res* 33, W284-288.
- Madera, M., 2008. Profile Comparer: a program for scoring and aligning profile hidden Markov models. *Bioinformatics* 24, 2630-2631.
- Soding, J., 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960.
- Wu, S., Zhang, Y., 2008. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* 72, 547-556.
- Yan, R., Xu, D., Yang, J., Walker, S., Zhang, Y., 2013. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. *Sci Rep* 3, 2619.
- Zhou, H., Zhou, Y., 2004. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. *Proteins* 55, 1005-1013.

Table S2

Comparison of residual and overall RMSD values between wild-type CsVAL28 and the mutants.

Wild-type				Mutant			Difference	
Position		Residue	Overall	Position	Residue	Overall	Residue	Overall
C1	75	1.09	1.53	C75S	1.49	1.90	-0.40	-0.37
C2	124	2.13	1.53	C124S	2.81	1.80	-0.68	-0.27
C3	128	0.55	1.53	C128S	0.45	1.81	0.10	-0.28
C4	144	0.40	1.53	C144S	0.46	1.78	-0.06	-0.25
C5	149	1.05	1.53	C149S	1.20	1.80	-0.15	-0.27
C6	160	0.25	1.53	C160S	0.38	1.81	-0.13	-0.28

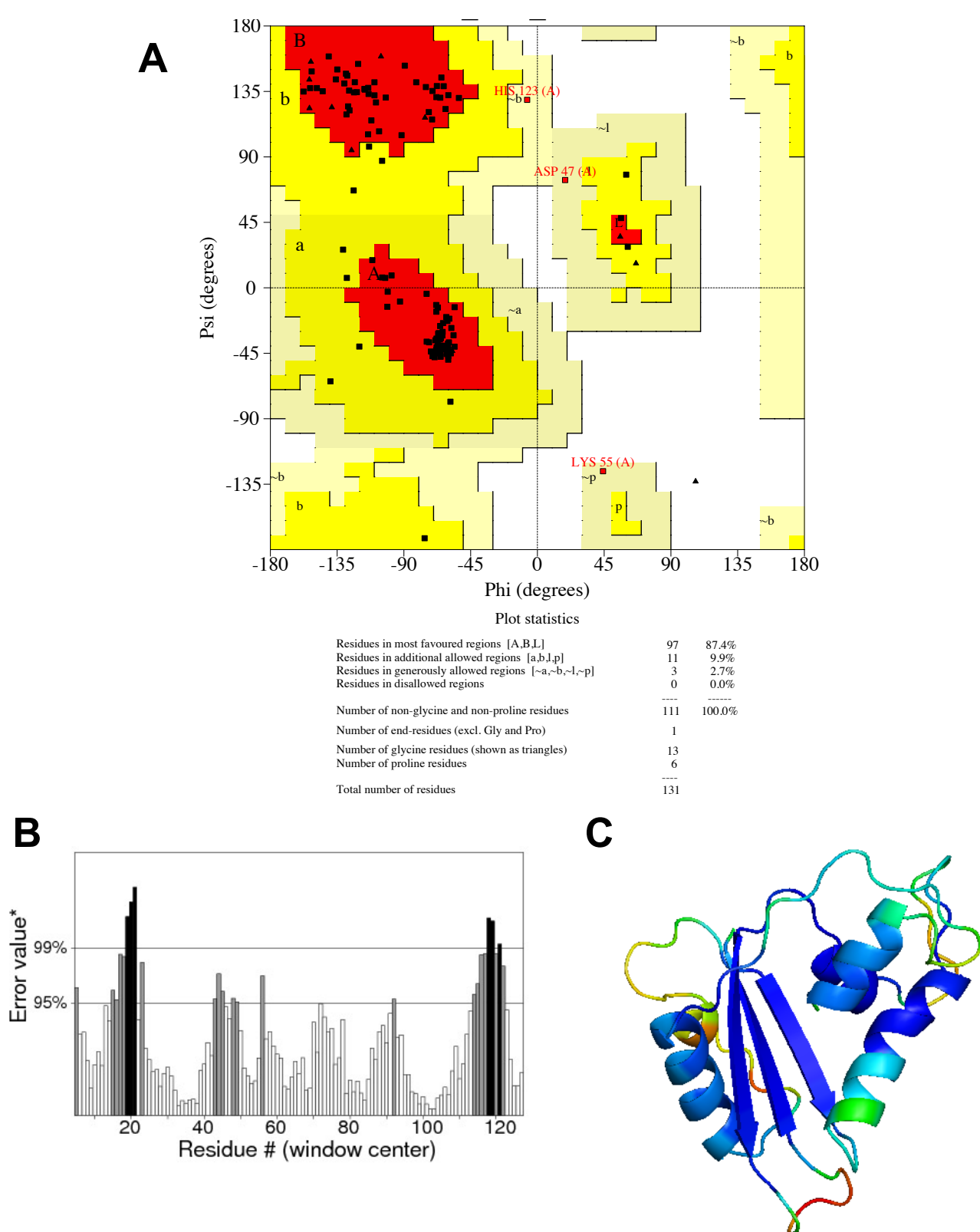
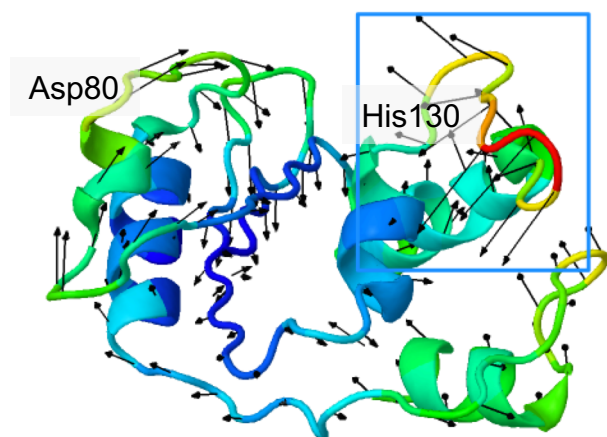
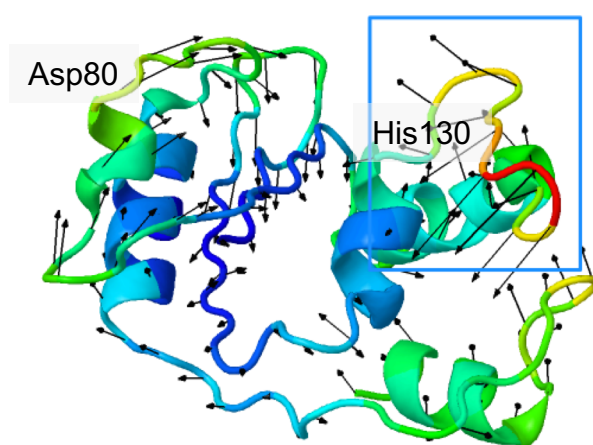


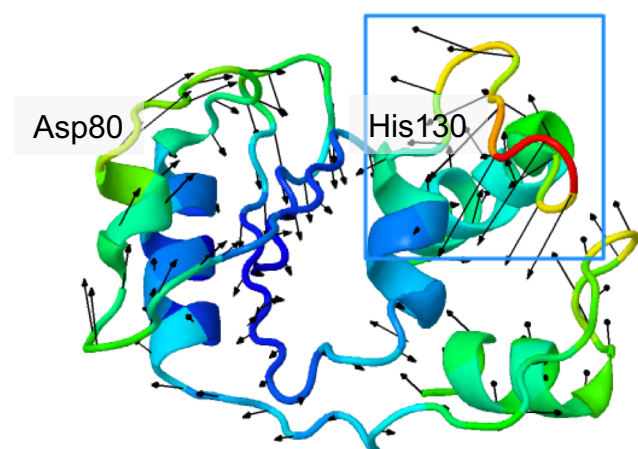
Figure S1. Validation of the CsVAL28 3D model. (A) The Ramachandran plot shows the residue in most favored regions (87.4%), additional allowed regions (9.9%), generously allowed regions (2.7%) and disallowed regions (0.0%). Red (A, B, L), yellow (a, b, l, p) and light yellow (~a, ~b, ~l, ~p) indicate the most favored regions, allowed regions and generously allowed regions. White shows disallowed regions. All non-glycine and non-proline residues are shown as closed black squares while glycines (non-end) are shown as closed black triangles. Generously allowed or disallowed residues are colored in red; (B) in the ERRAT plot, the overall quality factor is 79.68%; (C) 3D cartoon of the model was colored by the per-residue error according to the B-factor values using ModFOLD6. Decreasing magnitudes, as B-factor values, are color coded from red to blue.

A

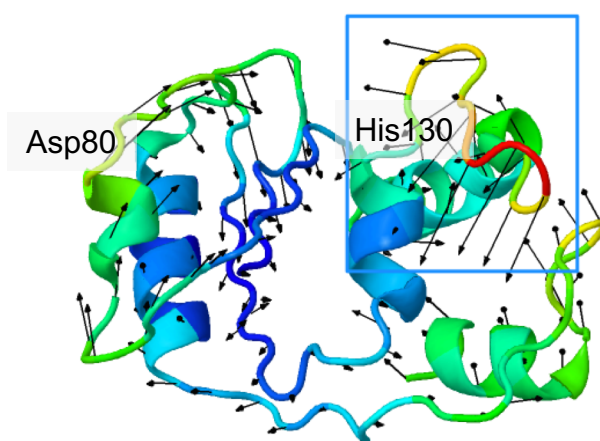
C128S model

B

C144S model

C

C149S model

D

C160S model

Figure S2. Comparison of mobility of CsVAL28 according to point mutation. The arrows represent directionality and magnitude of mobility in 3D models of C128S mutant (A), C144S mutant (B), C149S mutant (C) and C160S mutant (D). Increasing magnitudes, as B-factor values, are color coded from blue to red. Asp80 and His130 of Zn^{2+} -binding residues were displayed.