

RESEARCH

Open Access



# A voting mechanism-based linear epitope prediction system for the host-specific *Iridoviridae* family

Tao-Chuan Shih<sup>1†</sup>, Li-Ping Ho<sup>2†</sup>, Jen-Leih Wu<sup>3,4</sup>, Hsin-Yiu Chou<sup>2,5\*</sup> and Tun-Wen Pai<sup>1,6\*</sup>

From The 12th International Conference on Computational Systems Biology (ISB 2018) Guiyang, China. 18-21 August 2018

## Abstract

**Background:** The *Iridoviridae* family is categorized into five genera and clustered into two subfamilies: *Alphairidovirinae* includes *Lymphocystivirus*, *Ranavirus* (GIV), and *Megalocystivirus* (TGIV), which infect vertebrate hosts and *Betairidovirinae* includes *Iridovirus* and *Chloriridovirus*, which infect invertebrate hosts. Clustered *Iridoviridae* subfamilies possess host-specific characteristics, which can be considered as exclusive features for *in-silico* prediction of effective epitopes for vaccine development. A voting mechanism-based linear epitope (LE) prediction system was applied to identify and endorse LE candidates with a minimum length requirement for each clustered subfamily

**Results:** The experimental results showed that four conserved epitopes among the *Iridoviridae* family, one exclusive epitope for invertebrate subfamily and two exclusive epitopes for vertebrate family were predicted. These predicted LE candidates were further validated by ELISA assays for evaluating the strength of antigenicity and cross antigenicity. The conserved LEs for *Iridoviridae* family reflected high antigenicity responses for the two subfamilies, while exclusive LEs reflected high antigenicity responses only for the host-specific subfamily

**Conclusions:** Host-specific characteristics are important features and constraints for effective epitope prediction. Our proposed voting mechanism based system provides a novel approach for *in silico* LE prediction prior to vaccine development, and it is especially powerful for analyzing antigen sequences with exclusive features between two clustered groups.

**Keywords:** *Iridoviridae*, Linear epitope, Propensity scales, Host specificity, Group feature

## Background

Members of the family *Iridoviridae* are large icosahedral viruses that contain a single molecule of double-stranded DNA with genome sizes ranging from 104 to 220 Kbp [1]. The virus structure contains three domains, including an outer proteinaceous capsid, an intermediate lipid membrane, and a central core. Some viruses possess an outer envelope, and the outer capsid is composed of major capsid protein (MCP), which appears to be highly conserved

among the family and possesses surface binding sites interacting with the surfaces of host's cells [2]. The family *Iridoviridae* is categorized into five genera and clustered into two subfamilies (Table 1). The first *Betairidovirinae* subfamily, also referred as invertebrate iridescent viruses, contains both *Iridovirus* and *Chloriridovirus* genera, which infects invertebrate hosts such as insects and crustaceans. The second subfamily of *Alphairidovirinae*, also called vertebrate iridoviruses, includes *Megalocystivirus* [3] and *Lymphocystivirus* [3, 4] which infects fish only and *Ranavirus* [3] genus that infects fish, amphibian, and reptiles [5, 6]. In fact, vertebrate iridoviruses have caused high mortality of farmed fish and have led to huge economical lost [7]. For example, grouper is an important cultured species, which has suffered from vertebrate iridovirus infection in

\* Correspondence: [hychou@mail.ntou.edu.tw](mailto:hychou@mail.ntou.edu.tw); [twp@csie.ntut.edu.tw](mailto:twp@csie.ntut.edu.tw)

<sup>†</sup>Tao-Chuan Shih and Li-Ping Ho contributed equally to this work.

<sup>2</sup>Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan

<sup>1</sup>Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan

Full list of author information is available at the end of the article



recent years. The mortality of infected groupers was up to 60% for length 2–4 in. [8]. Due to high-density farming, the virus can be horizontally transferred once one fish is infected by vertebrate iridoviruses to cause severe damage. Surviving fish might retain the pathogens and continuously spread the virus to other health fish within the same pool [8, 9]. The prevention of vertebrate iridovirus infection has become an important task in fish farming. However, despite several injectable and oral commercial vaccines, the prevention results of specific vaccines have not been satisfactory for high-density farming. Therefore, a more effective immunization strategy and comprehensive vaccine development for different vertebrate iridoviruses have become important for challenging agriculture environments.

In general, immunobiologists have developed an integrated method for vaccine development based on analyzing protein sequences and structures of target viruses [10]. However, only few protein structures have been currently resolved for the *Iridoviridae* family, none of which are associated with the outer capsid proteins. Therefore, the prediction of MCP structures will be performed for surface analysis to facilitate the verification of the predicted LEs. B-cells play an important role in the immune system, and receptors on the cell membrane allow B-cells to bind to specific antigens [11]. Antigen proteins are usually too large to bind as a whole to any receptor. Hence, partial antigen segments located on surface areas called epitopes are bound by specific antibodies. Epitopes are categorized into two different types: linear epitope (LE) and conformational epitope (CE). An LE is a continuous sequence of amino acids contacting the receptors directly. By contrast, a CE is composed of discontinuous primary peptides, which are close neighbors on the three-dimensional surface. Generally speaking, more antibodies recognize CEs than LEs [12]. This is mainly due to native conformations being maintained to retain protein function [13]. However, the CE prediction requires antigenic structures to be resolved in advance for conformational region analysis. Due to this requirement, the LE prediction is a popular approach when the corresponding antigenic structures do not exist. In order to predict effective LEs as vaccine candidates, the predicted peptides should effectively elicit antibodies from specific hosts that recognize the antigens and provide protection against infection [14]. In addition, the selected peptides for vaccine design should be ideally conserved across different stages of the pathogen and possess binding affinity for the major populations of specific hosts [13]. In this study, we developed a voting mechanism-based computational system for linear epitope prediction, using the family *Iridoviridae* as an example for biological validation. The proposed system selected five existing epitope prediction tools, including LBTOPE [15], BepiPred [16], BCPREDS [17],

ABCPred [18], and LEPS [19]. Corresponding predicted epitopes for each tool were obtained, aligned, and selected through a voting procedure. The LBTOPE retrieved experimentally validated B-cell epitopes as well as non B-cell epitopes from the Immune Epitope Database (IEDB), and the system discriminated all candidate segments using an SVM classifier on various features, such as binary profile, dipeptide composition, and amino acid pair (AAP). Jespersen and researchers developed the BepiPred2.0 based on the random forest algorithm, and the system trained the epitope dataset annotated from antibody-antigen protein structures. BepiPred predicted epitopes derived from solved 3D structures and a large collection of linear epitopes downloaded from the IEDB database. EL-Manzalawy and researchers proposed the BCPREDS, which is based on SVM approaches, and utilized five different kernel methods and five-fold cross-validation of homology-reduce datasets that contained 701 linear B-cell epitopes and non-epitopes. The predicted epitopes were verified by predicted structures and structural alignment. In terms of the artificial neural network approach, ABCPred utilized a recurrent neural network (RNN) mode to train and test using different input window lengths and hidden units. The final network yielded an overall prediction accuracy of 65.93% through a five-fold cross-validation. The last LEPS prediction system was based on physicochemical propensity feature analysis and SVM classification. Peptides with globally or locally high physicochemical propensities were first identified as primitive linear epitope candidates. Subsequently, candidates were further validated through SVM classifiers according to collected known epitopes. After the target antigenic sequences were individually evaluated by these five LE predictors, each residue or a continuous segment within the target antigenic sequences was predicted as epitope or non-epitope residues. Therefore, a multiple sequence alignment should be performed to discover the conserved or exclusive LE candidates according to the clustered group features. A voting rule approach that incorporates surface structure evaluation is proposed in this study to reconfirm identified conserved and/or exclusive LE candidates. All designed prediction procedures and experimental materials are introduced and discussed in the following sections.

## Materials and methods

### Data collection

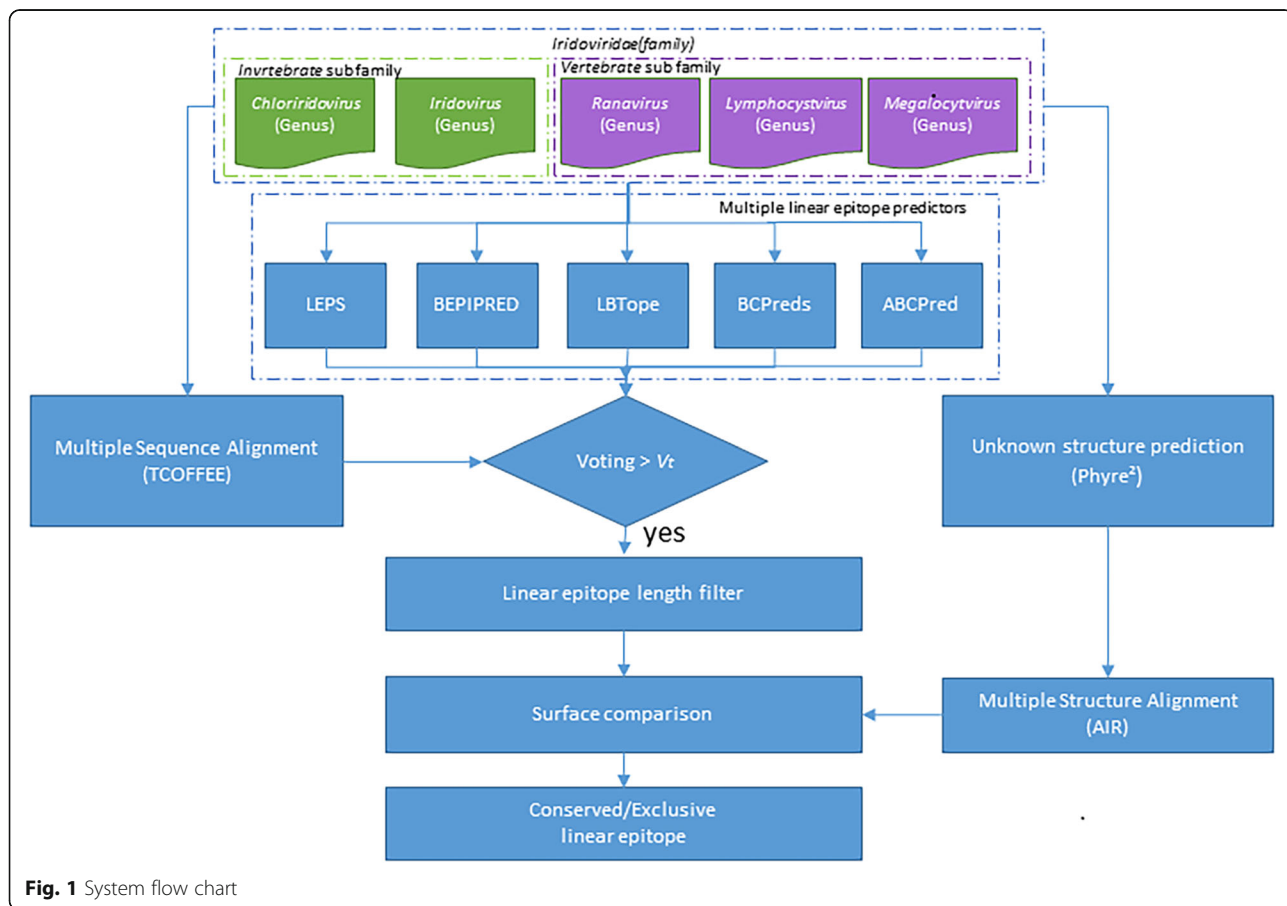
Five genera have been phylogenetically classified in the *Iridoviridae* family based on biological properties of the viruses, and representative species from each genus were selected according to the previously published results [5]. Uncertain taxonomic statuses of tentative species to the five established iridovirid genera were deliberately not considered in this study. Both *Lymphocystivirus* and

*Chloriridovirus* genera contain one recognized virus species, LCDV-1 and IIV-3, and these two viral species were solely selected as the representative sequences. The *Iridovirus* genus is consisted of two clearly classified species including IIV-6 and IIV-1 species. However, there is no MCP sequence could be found for IIV-1, so the IIV-6 species was selected as the representative species for *Iridovirus* genus. As for the *Megalocystivirus* and *Ranavirus* genera, each genus contains several recognized species by ICTV Subcommittee (International Committee on Taxonomy of Viruses). Since TGIV (*Megalocystivirus*) and GIV (*Ranavirus*) were host-specific antigens for groupers, these two virus species were selected as representative antigen sequences for the following biological experiment validation. All corresponding major capsid protein sequences were retrieved from NCBI [20] and Uniport [21]. To observe their corresponding virtual structures for surface verification, we applied Phyre2 [22] to predict the three-dimensional model for each capsid protein. The genus, representative species, and specific infected hosts are shown in 1.

from five different representative genera in the *Iridoviridae* family were input as query data. The system delivered five representative MCP sequences to T-Coffee [23] for performing multiple sequence alignment, and the Phyre2 was applied for structure prediction of each sequence. As the host specificity of each species in the *Iridoviridae* family were previously known, two groups, including vertebrate iridovirus subfamily (*Megalocystivirus*, *Ranavirus*, and *Lymphocystivirus*) and invertebrate iridovirus subfamily (*Chloriridovirus* and *Iridovirus*), were clustered according to host specificity. Five existing epitope prediction tools, including LBTOPE, BepiPred, BCPREDS, ABCPred, and LEPS, were utilized to predict LEs for each representative antigenic sequence. As a result, several conserved and/or exclusive structural segments between these two clustered subfamilies were determined, of which unaligned protein sequences and conformational variances may lead to host specificity. Nevertheless, the conserved epitope segments within the same clustered antigen group may play important roles for antibody binding with the same specific host species groups. To present multi-function of the proposed system, the five genera were applied as two different trials for LE prediction, including (1) the whole *Iridoviridae* family as a single group for conserved LEs

**System flowchart**

The designed voting mechanism-based LE prediction system is shown in Fig. 1. Representative MCP sequences



**Fig. 1** System flow chart

and (2) two clustered groups (vertebrate and invertebrate iridovirus subfamilies) for detecting conserved and exclusive LEs. Each antigenic sequence in the grouped clusters was analyzed using the five linear epitope prediction tools. These published web tools have been designed to identify each antigenic residue as an epitope or non-epitope residue. The results were encoded by a binary format, one for prediction as an epitope candidate residue and zero for a non-epitope residue. Hence, each residue in an antigen sequence possessed a total score from the five different prediction tools. According to various feature aspects of the different prediction tools, an accumulated higher score represented a higher confidence to suggest the residue as an epitope residue. To evaluate the conserved and/or exclusive characteristics among the whole *Iridoviridae* family, the final epitope scores were accumulated based on aligned residue positions from multiple sequence alignments. Though the sequence residue contents of an aligned position were different for distinct subfamilies due to evolution, the physicochemical and geometrical characteristics of aligned positions were mostly preserved. Hence, regarding the multiple sequence aligned results, if the accumulated epitope score of a residue from all prediction tools was higher than one half of the maximum score from all predicted residues, then the residue was considered as a predicted epitope candidate. The next step was to concatenate the continuous individual epitope candidates in neighboring relationships for an epitope with a minimum length requirement. Finally, both conserved and exclusive LEs for both vertebrate iridovirus and invertebrate iridovirus subfamilies were carefully enumerated and reconfirmed.

## Results

### LE prediction

The results of LE prediction were illustrated by two different trials: (1) import all five *Iridoviridae* family members as a single group to identify conserved epitopes and (2) import two host-specific groups of invertebrate iridovirus (IIV) subfamily and vertebrate iridovirus (VIV) subfamily to discover conserved and exclusive linear epitopes. Based on the designed system, all target pathogen sequences could be imported and assigned as a single or two groups for the antigenic feature analysis. A consensus voting system for LE prediction based on five linear

epitope tools was performed on the target sequences individually. To make the final decision, a default threshold setting of half of the maximum voting counts for an epitope residue candidate was applied according to multiple sequence alignment results. In addition, continuous peptide segment with lengths longer than 7 consecutive epitope candidate residues were considered as the final predicted LEs. The voting count distributions of three trials for the aligned sequences are shown in Fig. 2, and the dynamic threshold settings are depicted with orange lines. The final predicted LEs with lengths shorter than 7 or longer than 25 amino acids were discarded, as B cell epitopes are rarely outside of these ranges [24]. All candidates of conserved and/or exclusive LEs for the *Iridoviridae* family, VIV subfamily, and IIV subfamily are shown in Table 2 and Table 3 respectively.

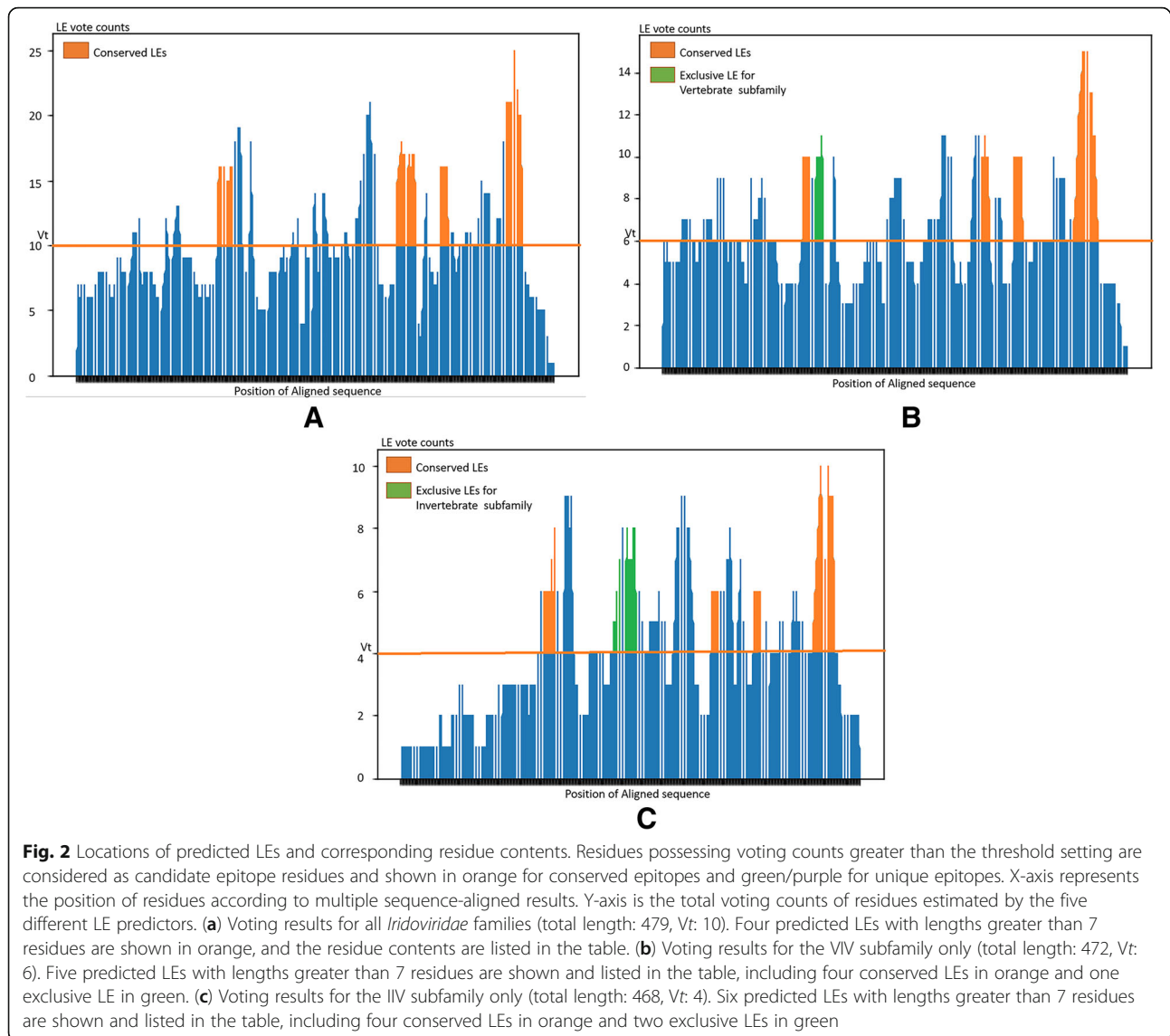
For the first trial of five *Iridoviridae* family members as a single group, there were four predicted LEs satisfying the criteria. Predicted peptides for each iridovirus sequence and corresponding aligned residue positions are listed in Fig. 2(a) and Table 2. In the second trial clustering five *Iridoviridae* family members into two groups, four conserved LEs were predicted as the results of the first trial (identical conserved LEs in both Table 2 and Table 3), one exclusive LE were predicted for the VIV subfamily (Fig. 2(b) and Table 3), and two exclusive LEs were predicted for the IIV subfamily (Fig. 2(c) and Table 3). All these detected LEs possessed high antigenicity characteristics, which were supported by five linear epitope predictors.

### Conserved and diverse structural segments of predicted LEs

To explore the structural variations of predicted LEs among different iridoviruses, the structure alignment of detected epitope candidates was performed by the Alignment Incorporate Refinement (AIR) system (unpublished system). AIR can measure the special similarity and molecular evolutionary relationship among protein structures by calculating core residue-based root-mean-square-deviation (RMSD) variations in atomic distances. In addition, the physicochemical variations could be easily viewed through the aligned structural conformation. The aligned results of the designed clustered experiments and the corresponding measurements by AIR are shown in both Table 2 and Table

**Table 1** Selected genera of the *Iridoviridae* family and infected host species

Genus	Selected species	Hosts
<i>Lymphocystivirus</i>	<i>Lymphocystis</i> disease virus 1 (LCDV-1)	Fish species (vertebrate)
<i>Megalocystivirus</i>	Taiwan grouper <i>Iridovirus</i> (TGIV)	Fish species (vertebrate)
<i>Ranavirus</i>	Grouper <i>Iridovirus</i> (GIV)	Amphibians, reptiles, and fish species (vertebrate)
<i>Iridovirus</i>	Invertebrate iridescent virus-6 (IIV-6)	Insects, crustaceans (invertebrate)
<i>Chloriridovirus</i>	Invertebrate iridescent virus-3 (IIV-3)	Mosquitoes (invertebrate)



3, and an example of structure aligned result is shown in Fig. 3 with visualized RMSD variations. RMSD variation was used to measure the geometrical similarity of aligned three-dimensional structures regarding the spatial distance of corresponding residues. In this study, RMSD measurements were applied for searching conserved and diverse structural segments between invertebrate *Iridoviridae* and vertebrate *Iridoviridae* subfamilies. The higher RMSDs represented lower structural similarity of the predicted LE segments. We assessed one of the predicted epitopes of “TAGNGGNTSGYKDAQK” (from *Lymphocystivirus*) as an example since the segments possessed a relative higher epitope score. The segment was predicted as a conserved linear epitope for the *Iridoviridae* family. Five aligned predicted LE segments possess an average RMSD of 1.122 (Å), and these LE segments can be selected and viewed through a 3D view page which utilizes Jmol open-sources (a Java

viewer for 3D chemical structures). Multiple predicted LE segment alignment and their corresponding structural aligned results from predicted protein structures can be viewed and compared through the AIR system simultaneously. Surface properties of predicted LEs and structural diversity among various viruses could be evaluated and confirmed for the next step of vaccine design and development.

#### ELISA assays for host-specific antigenicity and cross antigenicity tests

To validate immune responses evoked by the predicted LEs from both VIV and IIV subfamilies, ELISA assays were performed to evaluate the strength of antigenicity and cross antigenicity. According to the predicted results shown in Table 3, we selected exclusive LEs of TIDMTQPVDS for *Lymphocystivirus*<sub>157–166</sub>, TSDMTNPTPA for *Megalocystivirus*<sub>157–166</sub>, and RSDLVGGITN for *Ranavirus*<sub>152–161</sub> from

**Table 2** Predicted conserved LEs and corresponding RMSDs of virtual structures by taking the whole *Iridoviridae* family as a single group

Iridoviridae groups	Predictive LEs of representative peptide	Residue Location	RMSDs of structural alignment (Å)
<i>Iridoviridae</i> family (Conserved LEs)	KAVGYDNMIGN	<i>Iridovirus</i> :145~155	0.533
	KRNGYDNMIGN	<i>Chloriridovirus</i> :46~56	
	KRIGYDNMIGN	<i>Lymphocystivirus</i> :146~156	
	KRIGYDNMIGN	<i>Megalocystivirus</i> :146~156	
	KQSGYNKMIGM	<i>Ranavirus</i> :141~151	
	SNYTTASPVITSTS	<i>Iridovirus</i> :323~336	1.313
	SNYGTSSPVVSGTS	<i>Chloriridovirus</i> :222~233	
	SNYTSSSPVIFDGG	<i>Lymphocystivirus</i> :315~327	
	SNYTCVTPVNGPGN	<i>Megalocystivirus</i> :320~332	
	SNYTAASPVVNNK	<i>Ranavirus</i> :311~323	
	RLNHMGSD	<i>Iridovirus</i> :360~367	0.9
	RLGTMGSD	<i>Chloriridovirus</i> :258~265	
	RLNEMGSE	<i>Lymphocystivirus</i> :352~359	
	RLANMGVE	<i>Megalocystivirus</i> :356~363	
	RLHQMGVD	<i>Ranavirus</i> :346~353	
	GAAGTGPAAGSGQNFQPT	<i>Iridovirus</i> :424~442	1.122
	ASTGAGDGAGANYNQS	<i>Chloriridovirus</i> :324~339	
	TAGGNGGNTSGYKDAQK	<i>Lymphocystivirus</i> :415~434	
	AAAGGGNNNSGYNEPQR	<i>Megalocystivirus</i> :419~438	
	TAAGGGNGTGYTVAQK	<i>Ranavirus</i> :409~428	

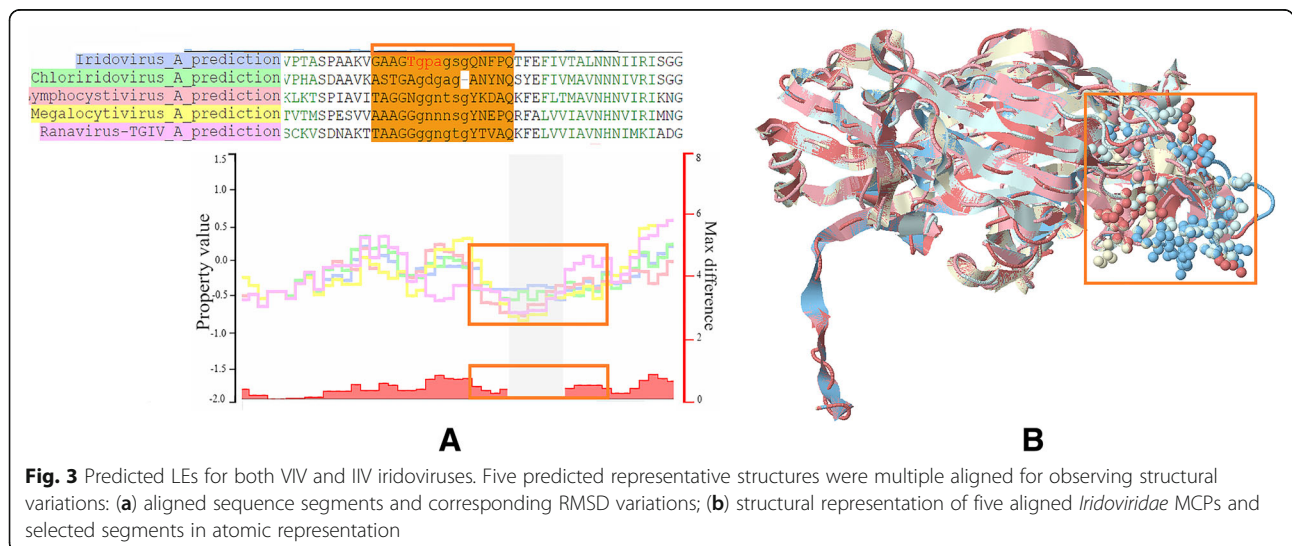
the VIV subfamily for antigenicity tests, and for the IIV subfamily, we selected the four predicted exclusive LEs of VALPTAALPYNE for *Iridovirus*<sub>193~203</sub>, VSLPTAALPYNE for *Chloriridovirus*<sub>92~102</sub>, VASQTVVPVVG for *Iridovirus*<sub>226~236</sub>, and TTGNPYQTIDV for *Chloriridovirus*<sub>125~135</sub> for cross antigenicity tests. All selected LE segments were synthesized by GeneScript Company (USA) for biological validation. After synthesis procedures, the synthesized epitopes were used as antigen in the ELISA assay respectively. Rabbit anti-rTGIVmcp and anti-rGIVmcp antibodies were individually prepared by the local bio-supplier. The secondary antibody Goat anti-rabbit IgG(h + l) conjugated with horseradish peroxidase (HRP) was purchased from Bethyl Laboratories (TX, USA). Briefly, 10 µg of synthesized epitopes were coated on a 96-well microplate with coating buffer (pH 9.6, Sigma-Aldrich, USA). All procedures were performed in accordance to the manufacturer's protocol [25]. After hybridization, substrate TMB (3,3',5,5'-tetramethylbenzidine, Amersco, USA) was added for HRP detection and read at 630 nm on an ELISA reader. The ELISA results were analyzed by ANOVA (Analysis of Variance) in the GraphPad Prism 5.0 biological statistical software (GraphPad Software, Inc.).

The corresponding immune responses are shown in Fig. 4. Each subfigure contains 4 vertical bars related

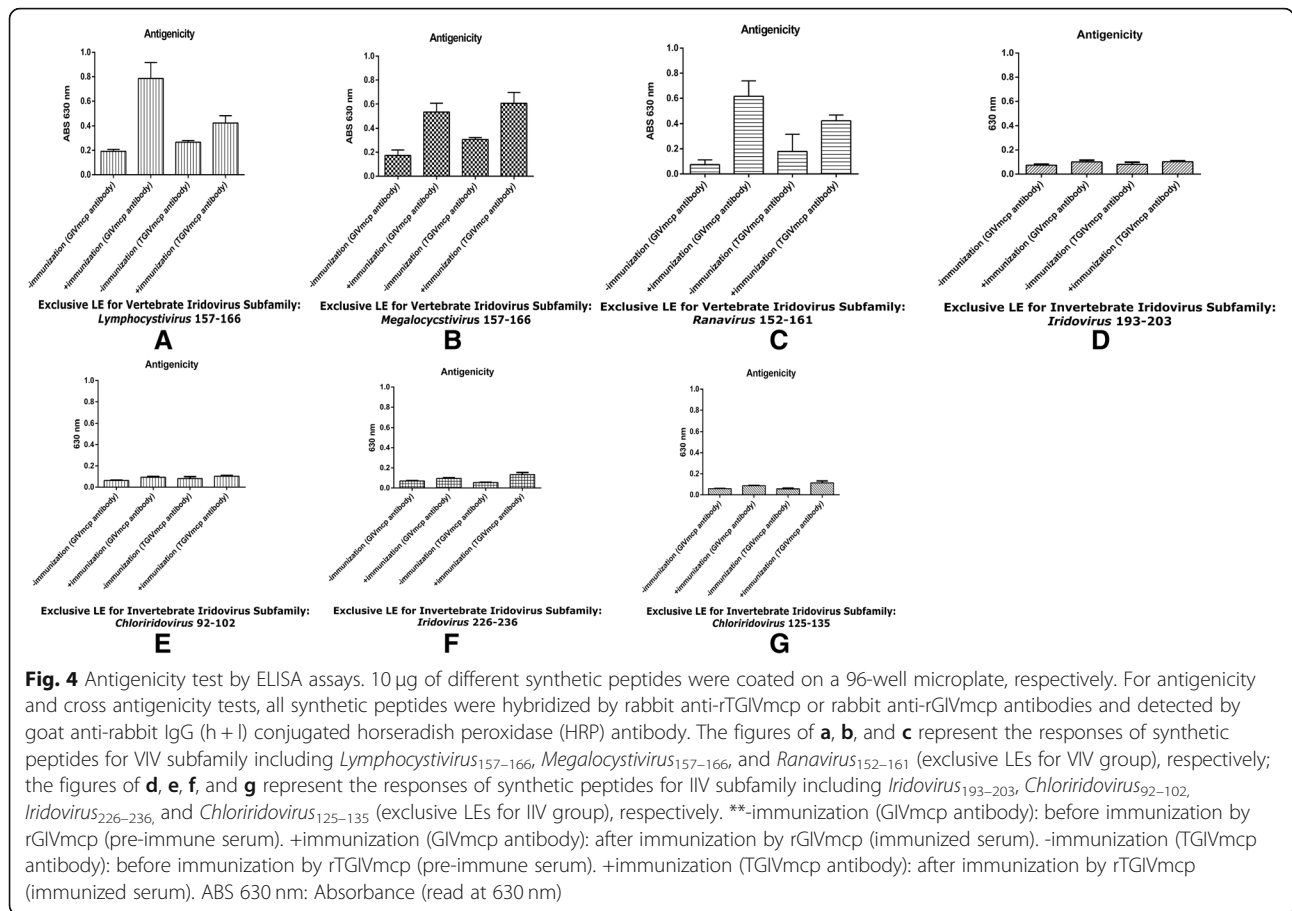
to the strength of immune responses, and higher bars imply higher antigenicity and higher immune responses induced by Rabbit anti-GIVmcp and Rabbit anti-TGIVmcp antibodies. The first and the third bars in each subfigure are for the pre-immune serum before rGIVmcp and rTGIVmcp immunization, while the second and the fourth bars represent the immune responses induced by Rabbit anti-GIVmcp antibody and Rabbit anti-TGIVmcp antibody. After observing the results of antigenicity test by ELISA assays, the exclusive LE segments obtained from the VIV subfamily (*Lymphocystivirus*<sub>157~166</sub>, *Megalocystivirus*<sub>157~166</sub>, and *Ranavirus*<sub>152~161</sub>) indeed induced significant immune responses (Fig. 4 a, b and c), while the exclusive LE segments obtained from the IIV subfamily of VALPTAALPYNE for *Iridovirus*<sub>193~203</sub>, VSLPTAALPYNE for *Chloriridovirus*<sub>92~102</sub>, VASQTVVPVVG for *Iridovirus*<sub>226~236</sub>, and TTGNPYQTIDV for *Chloriridovirus*<sub>125~135</sub> reflect little cross host group reaction after immunization (Fig. 4 d, e, f, and g). In other words, both predicted exclusive LEs from IIV subfamily evoked little response to Rabbit anti-GIVmcp antibody and Rabbit anti-TGIVmcp antibody after ELISA assay validation. These results demonstrate important evidences of host-specific features for facilitating LE prediction and for effective vaccine development.

**Table 3** Predicted conserved and exclusive LEs for two clustered groups: vertebrate and invertebrate iridovirus subfamilies. (\*N/A represents the predicted segments possessing missing residues among the five virtual structures)

<i>Iridoviridae</i> groups	Predictive LEs of representative peptide	Residue Location	RMSDs of structural alignment(Å)
<i>Iridoviridae</i> family (Conserved LEs)	KAVGYDNMIGN	(IIV)	<i>Iridovirus</i> :145~155
	KRNGYDNMIGN		<i>Chloriridovirus</i> :46~56
	KRIGYDNMIGN	(VIV)	<i>Lymphocystivirus</i> :146~156
	KRIGYDNMIGN		<i>Megalocystivirus</i> :146~156
	KQSGYNKMIGM		<i>Ranavirus</i> :141~151
	SNYTTASPVITSTT	(IIV)	<i>Iridovirus</i> :323~336
	SNYGTSSPWVGTS		<i>Chloriridovirus</i> :222~233
	SNYTSSSPVIFDGG	(VIV)	<i>Lymphocystivirus</i> :315~327
	SNYTCVTPVNGPGN		<i>Megalocystivirus</i> :320~332
	SNYTAASPVVNNK		<i>Ranavirus</i> :311~323
	RLNHMGSD	(IIV)	<i>Iridovirus</i> :360~367
	RLGTMGSD		<i>Chloriridovirus</i> :258~265
	RLNEMGSE	(VIV)	<i>Lymphocystivirus</i> :354~359
	RLANMGVE		<i>Megalocystivirus</i> :356~363
	RLHQMGVD		<i>Ranavirus</i> :346~353
	GAAGTGPAGSGQNFPQT	(IIV)	<i>Iridovirus</i> :424~442
	ASTGAGDGAGANYNQS		<i>Chloriridovirus</i> :324~339
	TAGGNGGNTSG	(VIV)	<i>Lymphocystivirus</i> :418~428
	AAAGGGNNNSG		<i>Megalocystivirus</i> : 422~432
	TAAGGGGNGTG		<i>Ranavirus</i> :412~422
VIV (Exclusive LEs)	TIDMTQPVDS		<i>Lymphocystivirus</i> :157~166
	TSDMTNPTPA		<i>Megalocystivirus</i> : 157~166
	RSDLVGGITN		<i>Ranavirus</i> :152~161
IIV (Exclusive LEs)	VALPTAALPYNE		<i>Iridovirus</i> :193~203
	VSLPTAALPYNE		<i>Chloriridovirus</i> :92~102
	VASQTVWPWG		<i>Iridovirus</i> :226~236
	TTGNPYQTIDV		<i>Chloriridovirus</i> :125~135



**Fig. 3** Predicted LEs for both VIV and IIV iridoviruses. Five predicted representative structures were multiple aligned for observing structural variations: (a) aligned sequence segments and corresponding RMSD variations; (b) structural representation of five aligned *Iridoviridae* MCPs and selected segments in atomic representation



### Discussion

A voting mechanism-based LE prediction system for host-specific antigens is proposed. For group feature detection, the antigen sequences could be clustered prior to importing the sequences into the proposed system. Taking the *Iridoviridae* family as an example, the family can be categorized into two different subfamilies, vertebrate iridovirus and invertebrate iridovirus subfamilies. In this study, we applied different combinations of clustered groups to predict the conserved and exclusive LEs. The imported antigen sequences of each subfamily were analyzed by a consensus voting mechanism respectively. Protein structure prediction for query antigens was performed to conjecture whether the predicted LEs are located on protein structural surface, and a multiple structure alignment analysis was also performed to reconfirm the conserved and exclusive characteristics. Using the multiple sequence aligned locations, the consensus voting module selected epitope candidate residues by accumulating votes provided by five different renowned epitope prediction tools. In addition to individually voted epitope residues, the minimum lengths of concatenated epitope residues were also required for further experimental design. All LE candidates for different

host specific groups were selected and cross-identified. In addition, to increase the successful results of vaccine design, we emphasized the surface structure characteristics of the predicted LE epitopes. Therefore, the antigens without resolved structures will be analyzed by Phyre<sup>2</sup> for creating corresponding virtual structures. The predicted epitopes would be checked for their surface conditions based on the predicted structures, and the predicted structures will be structurally aligned by any structure alignment system for reconfirming exclusiveness of predicted LEs. Based on the aligned results, the predicted conserved and/or exclusive LEs for different subfamilies can be structurally distinguished. In order to validate the predicted LE segments, the two frequent iridovirus infections (TGIV/GIV) in Taiwanese groupers were especially applied for experimental validation. A total of five predicted LEs were synthesized and investigated for antigenicity and cross-antigenicity tests. The exclusive LE candidates for vertebrate subfamily including *Lymphocystivirus*<sub>157-166</sub> (exclusive), *Megalocystivirus*<sub>157-166</sub> (exclusive), and *Ranavirus*<sub>152-161</sub> (exclusive) were selected as representative LEs for antigenicity tests against vertebrate hosts, while the LE candidates for invertebrate subfamily including *Chloriridovirus*<sub>92-102</sub> (exclusive), *Chloriridovirus*<sub>125-135</sub> (exclusive),



*Iridovirus*<sub>193–203</sub> (exclusive), and *Iridovirus*<sub>226–236</sub> (exclusive) were selected as exclusive LEs for invertebrate group for cross antigenicity tests against vertebrate hosts. According to the ELISA results, it showed the exclusive LE segments for VIV subfamily indicated high antigenicity after immunity test against vertebrate hosts and the exclusive LE segments for invertebrate *Iridoviridae* segments showed no or little antigenic responses. All these experimental results imply the predicted LEs for VIV subfamily possessing relatively high antigenicity responses either for conserved (data not shown) or exclusive peptides, since both antigens (GIVmcp or TGIVmcp) affect the same host of groupers. We further compared the predicted LE peptides in the Table 2 and Table 3, and found that all conserved LEs predicted in two-clustered groups (VIV and IIV) are the same as the predicted LEs by taking the whole *Iridoviridae* family as a single group. Though these predicted conserved LEs indeed possessed high antigenicity and confirmed by five well known LE predictors, they would not practically evoke the immune responses or cause clinical signs of disease in vertebrate hosts. Hence, these conserved LEs obtained from *in-silico* prediction may be ignored for the following vaccine development stage, though these peptides also possess relatively high antigenicity. Furthermore, we investigated immune responses of the predicted exclusive LEs from IIV against vertebrate hosts, and the results showed there was no significant response. This is mainly due to both *Megalocystivirus* and *Ranavirus* belong to vertebrate host specificity, and this proved our assumption of host specific characteristics. Regarding biological validation experiments, commercial ELISAs might not be able to reflect pertinent responses for the predicted and synthesized LE peptides. To deal with such issues, a lot of iridovirus vaccines were evaluated through inactivated viruses [26–29], virus like particles (VLPs) [30], recombinant coat proteins [31–33] and DNA vaccines [34, 35]. However, there still exist several limitations to overcome such as structures of synthesized protein segments and the length settings of final predicted segments. These factors might affect the immune specificity, and different virus species might result in different protective efficacy. Therefore, this study could provide a precision approach for peptide vaccine design in iridovirus disease control, which can accelerate the process of vaccine development, reduce costs and save time. In some previous studies, adding linkers improved bioreactivity, folding, and stability of fusion proteins, such as helical peptide linkers, [A (EAAAK)nA] m ( $n = 2-4$ ,  $m = 1$  or  $2$ ) [36], or a flexible linker, (GGGS)<sub>3</sub> [37]. Some linkers can be used as adjuvants to enhance immune specificity for peptide vaccines [38–41]. Therefore, addition of suitable linkers to the predicted peptides could enhance and promote immune responses in hosts. We will further apply adjuvant-based linkers to predicted LE segments

and enhance the specificity for immune responses in the near future. According to the results of the biological experiments and bioinformatics analysis, the developed voting mechanism-based linear epitope prediction system can successfully predict linear epitopes with significant antigenic specificity.

## Conclusions

We have developed voting mechanism-based linear epitope prediction for the host-specific *Iridoviridae* family. We used five renowned LE prediction systems and exclusive features, endorsed LE candidates with a minimum length requirement could be identified for each subfamily by various prediction systems. Our result suggests that there are some exclusive antigen features between the sequence of invertebrate Iridovirus and vertebrate Iridovirus which cause the host-specific features between these two groups. Furthermore, we apply surface comparison and ELISA to identify vertebrate host-specific LEs for both TGIV and GIV which reflect high antigenicity response for specific grouper species. According to this novel prediction system, the predicted LEs can facilitate immunologists in designing specific biological experiments for future vaccine development.

## Abbreviations

AIr: Alignment Incorporate Refinement; CE: Conformational epitope; GIV: Grouper Iridovirus; ICTV: International Committee on Taxonomy of Viruses; IIV: Invertebrate iridovirus; LE: Linear epitope; MCP: Major capsid protein; RMSD: residue-based root-mean-square-deviation; TGIV: Taiwan grouper Iridovirus; VIV: Vertebrate iridovirus

## Funding

Publication costs are funded by Ministry of Science and Technology, Taiwan (MOST 106–2313-B-019 -002 and MOST 107–2313-B-027-001 to Tun-Wen Pai, and MOST 107–2321-B-019-002 to Jen-Leih Wu).

## Availability of data and materials

Not applicable.

## About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 7, 2019: Selected papers from the 12th International Conference on Computational Systems Biology (ISB 2018)*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-7>.

## Authors' contributions

TCS and TWP conceived the algorithm and wrote the manuscript. LPH and HYC designed and performed biological experiments. JLW proofread and revised the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan. <sup>2</sup>Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan. <sup>3</sup>Department of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan. <sup>4</sup>Institute of Cellular and Organismic Biology, Academia Sinica, Taipei, Taiwan. <sup>5</sup>Department of Aquaculture, College of Life Science, National Taiwan Ocean University, Keelung, Taiwan. <sup>6</sup>Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei, Taiwan.

Published: 1 May 2019

### References

- Chinchar V, Hyatt A, Miyazaki T, Williams T. Family *Iridoviridae*: poor viral relations no longer. In: Lesser known large dsDNA viruses: Springer; 2009. p. 123–70.
- Nadugala MN, Jeewandara C, Malavige GN, Premaratne PH, Goonasekara CL. Natural antibody responses to the capsid protein in sera of dengue infected patients from Sri Lanka. PLoS One. 2017;12(6):e0178009.
- Murali S, Wu MF, Guo IC, Chen SC, Yang HW, Chang CY. Molecular characterization and pathogenicity of a grouper iridovirus (GIv) isolated from yellow grouper, *Epinephelus awoara* (Temminck & Schlegel). J Fish Dis. 2002;25(2):91–100.
- Do JW, Cha SJ, Kim JS, An EJ, Park MS, Kim JW, Kim YC, Park MA, Park JW. Sequence variation in the gene encoding the major capsid protein of Korean fish iridoviruses. Arch Virol. 2005;150(2):351–9.
- Chinchar V, Hyatt A. Iridoviruses: general features. 2008.
- Piégu B, Asgari S, Bideshi D, Federici BA, Bigot Y. Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily Megavirales. Mol Phylogenet Evol. 2015;84:44–52.
- Lin H-Y, Liou C-J, Cheng Y-H, Hsu H-C, Yiu J-C, Chiou PP, Lai Y-S. Development and application of a monoclonal antibody against grouper iridovirus (GIv) major capsid protein. J Virol Methods. 2014;205:31–7.
- Liu C-C, Chen C-M, Pai T-W, Chou H-Y, Hsu H-H. Exclusive genomic pathway analysis for groupers infected by different iridovirus. In: Complex, intelligent, and software intensive systems (CISIS), 2015 ninth international conference on: 2015: IEEE. p. 508–12.
- Liu C-C, Ho L-P, Yang C-H, Kao T-Y, Chou H-Y, Pai T-W. Comparison of grouper infection with two different iridoviruses using transcriptome sequencing and multiple reference species selection. Fish Shellfish Immunol. 2017;71:264–74.
- Zolla-Pazner S, Cardozo T. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. Nat Rev Immunol. 2010;10(7):527–35.
- Hilleman MR. Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. J Hum Virol. 2000;3(2):63–76.
- Forsström B, Axnäs BB, Rockberg J, Danielsson H, Bohlin A, Uhlen M. Dissecting antibodies with regards to linear and conformational epitopes. PLoS One. 2015;10(3):e0121673.
- Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J Biomed Inform. 2015;53:405–14.
- Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar A, Lund O, Morrison S. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. J Mol Recognit. 2007;20(2):75–82.
- Singh H, Ansari HR, Raghava GP. Improved method for linear B-cell epitope prediction using antigen's primary sequence. PLoS One. 2013;8(5):e62216.
- Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017;45(W1):W24–9.
- EL-Manzalawy Y, Dobbs D, Honavar V. Predicting linear B-cell epitopes using string kernels. J Mol Recognit. 2008;21(4):243–55.
- Saha S, Raghava G. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins. 2006;65(1):40–8.
- Wang H-W, Lin Y-C, Pai T-W, Chang H-T. Prediction of B-cell linear epitopes with a combination of support vector machine classification and amino acid propensity identification. Biomed Res Int. 2011. (<http://140.121.197.182/LEPS/>).
- Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012;40(Database issue):D130–5.
- UniProt C: UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(Database issue):D204–D212.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10(6):845–58.
- Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302(1):205–17.
- Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody: protein complexes. Mol Immunol. 2013;53(1–2):24–34.
- Kemeny D, Sarmay G. A practical guide to ELISA: D. M. Kemeny. Oxford: Pergamon Press; 1992.
- Nakajima K, Maeno Y, Honda A, Yokoyama K, Tooriyama T, Manabe S. Effectiveness of a vaccine against red sea bream iridoviral disease in a field trial test. Dis Aquat Org. 1999 Apr 15;36(1):73–5.
- Caipang CM, Hirono I, Aoki T. Immunogenicity, retention and protective effects of the protein derivatives of formalin-inactivated red seabream iridovirus (RSIV) vaccine in red seabream. Pagrus major. Fish Shellfish Immunol. 2006;20(4):597–609.
- Liu W, Xu J, Ma J, LaPatra SE, Meng Y, Fan Y, Zhou Y, Yang X, Zeng L. Immunological responses and protection in Chinese giant salamander *Andrias davidianus* immunized with inactivated iridovirus. Vet Microbiol. 2014;174(3–4):382–90.
- Ou-yang Z, Wang P, Huang X, Cai J, Huang Y, Wei S, Ji H, Wei J, Zhou Y, Qin Q. Immunogenicity and protective effects of inactivated Singapore grouper iridovirus (SGIV) vaccines in orange-spotted grouper. *Epinephelus coioides*. Dev Comp Immunol. 2012;38(2):254–61.
- Shimmoto H, Kawai K, Ikawa T, Oshima S. Protection of red sea bream *Pagrus major* against red sea bream iridovirus infection by vaccination with a recombinant viral protein. Microbiol Immunol. 2010;54(3):135–42.
- Liu HI, Chiou PP, Gong HY, Chou HY. Cloning of the major capsid protein (MCP) of grouper Iridovirus of Taiwan (TGIV) and preliminary evaluation of a recombinant MCP vaccine against TGIV. Int J Mol Sci. 2015;16(12):28647–56.
- Fu X, Li N, Lai Y, Liu L, Lin Q, Shi C, Huang Z, Wu S. Protective immunity against iridovirus disease in mandarin fish, induced by recombinant major capsid protein of infectious spleen and kidney necrosis virus. Fish Shellfish Immunol. 2012;33(4):880–5.
- Zhou Y, Fan Y, LaPatra SE, Ma J, Xu J, Meng Y, Jiang N, Zeng L. Protective immunity of a *Pichia pastoris* expressed recombinant iridovirus major capsid protein in the Chinese giant salamander, *Andrias davidianus*. Vaccine. 2015;33(42):5662–9.
- Zhang M, Hu YH, Xiao ZZ, Sun Y, Sun L. Construction and analysis of experimental DNA vaccines against megalocytivirus. Fish Shellfish Immunol. 2012;33(5):1192–8.
- Jung MH, Nikapitiya C, Jung SJ. DNA vaccine encoding myristoylated membrane protein (MMP) of rock bream iridovirus (RBIV) induces protective immunity in rock bream (*Oplegnathus fasciatus*). Vaccine. 2018;36(6):802–10.
- Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng. 2001;14(8):529–32.
- Shamriz S, Ofoghi H, Moazami N. Effect of linker length and residues on the structure and stability of a fusion protein with malaria vaccine application. Comput Biol Med. 2016;76:24–9.
- Klein JS, Jiang S, Galimidi RP, Keeffe JR, Bjorkman PJ. Design and characterization of structured protein linkers with differing flexibilities. Protein Eng Des Sel. 2014;27(10):325–30.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5(4):725.
- Hou J, Liu Y, Hsi J, Wang H, Tao R, Shao Y. Cholera toxin B subunit acts as a potent systemic adjuvant for HIV-1 DNA vaccination intramuscularly in mice. Hum Vaccin Immunother. 2014;10(5):1274–83.
- Wang C, Zhu W, Wang B-Z. Dual-linker gold nanoparticles as adjuvanting carriers for multivalent display of recombinant influenza hemagglutinin trimers and flagellin improve the immunological responses in vivo and in vitro. Int J Nanomedicine. 2017;12:4747.