

Research Article

Model for Vaccine Design by Prediction of B-Epitopes of IEDB Given Perturbations in Peptide Sequence, In Vivo Process, Experimental Techniques, and Source or Host Organisms

Humberto González-Díaz,^{1,2} Lázaro G. Pérez-Montoto,³ and Florencio M. Ubeira³

¹ Department of Organic Chemistry II, University of the Basque Country UPV/EHU, 48940 Bilbao, Spain

² Ikerbasque, Basque Foundation for Science, 48011 Bilbao, Spain

³ Department of Microbiology and Parasitology, University of Santiago de Compostela (USC), 15782 Santiago de Compostela, Spain

Correspondence should be addressed to Humberto González-Díaz; humberto.gonzalezdiaz@ehu.es

Received 21 August 2013; Accepted 17 November 2013; Published 12 January 2014

Academic Editor: Darren R. Flower

Copyright © 2014 Humberto González-Díaz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Perturbation methods add variation terms to a known experimental solution of one problem to approach a solution for a related problem without known exact solution. One problem of this type in immunology is the prediction of the possible action of epitope of one peptide after a perturbation or variation in the structure of a known peptide and/or other boundary conditions (host organism, biological process, and experimental assay). However, to the best of our knowledge, there are no reports of general-purpose perturbation models to solve this problem. In a recent work, we introduced a new quantitative structure-property relationship theory for the study of perturbations in complex biomolecular systems. In this work, we developed the first model able to classify more than 200,000 cases of perturbations with accuracy, sensitivity, and specificity >90% both in training and validation series. The perturbations include structural changes in >50000 peptides determined in experimental assays with boundary conditions involving >500 source organisms, >50 host organisms, >10 biological process, and >30 experimental techniques. The model may be useful for the prediction of new epitopes or the optimization of known peptides towards computational vaccine design.

1. Introduction

National Institute of Allergy and Infectious Diseases (NIAID) supported the launch, in 2004, of the Immune Epitope Database (IEDB), <http://www.iedb.org/> [1–4]. The IEDB system withdrew information from approximately 99% of all papers published to date that describe immune epitopes. In doing so, IEDB system analyses over 22 million PubMed abstracts and subsequently curated ≈13 K references, including ≈7 K manuscripts about infectious diseases, ≈1 K about allergy topics, ≈4 K about autoimmunity, and 1 K about transplant/allo-antigen topics [5]. IEDB lists a huge amount of information about the molecular structure as well as the experimental conditions (c_{ij}) in which different i th molecules were determined to be immune epitopes or not. This explosion of information makes necessary both query/display functions for retrieval of known data from IEDB as well predictive tools for

new epitopes. Salimi et al. [5] reviewed advances in epitope analysis and predictive tools available in the IEDB. In fact, IEDB analysis resource (IEDB-AR: <http://tools.iedb.org/>) is a collection of tools for prediction of molecular targets of T- and B-cell immune responses (i.e., epitopes) [6, 7].

On the other hand, Quantitative Structure-Activity/Property Relationships (QSAR/QSPR) techniques are useful tool to predict new drugs, RNA, drug-protein complexes, and protein-protein complexes. In general, QSAR/QSPR-like methods transform molecular structures into numeric molecular descriptors (λ_i) in a first stage and later fit a model to predict the biological process. For example, DRAGON [8–10], CODESSA [11, 12], MOE [13], TOPS-MODE [14–17], TOMOCOMD [18, 19], and MARCH-INSIDE [20] are among the most used softwares to calculate molecular descriptors based on quantum mechanics (QM) and/or graph theory [21–27]. The software STATISTICA [28] and WEKA [29] are often

used to perform multivariate statistics and/or machine learning (ML) analysis in order to preprocess data and later fit the final QSAR/QSPR model using techniques like principal component analysis (PCA), linear discriminant analysis (LDA), support vector machine (SVM), or artificial neural networks (ANN) [28].

QSAR/QSPR models are also important in immunoinformatics to predict the propensity of different molecular structures to play different roles in immunological processes. They include skin vaccine adjuvants and sensitizers [30–38], drugs and their activity/toxicity protein targets in the immune system [39], and epitopes [40–49]. Moreover, Reche and Reinherz [50] implemented PEPVAC (promiscuous epitope-based vaccine), a web server for the formulation of multi-epitope vaccines that predict peptides binding to five distinct HLA class I supertypes (A2, A3, B7, A24, and B15). PEPVAC can also identify conserved MHC ligands, as well as those with a C-terminus resulting from proteasomal cleavage. The Dana-Farber Cancer Institute hosted the PEPVAC server at the site <http://immunax.dfci.harvard.edu/PEPVAC/>. To close with a last example, Lafuente and Reche [51] reviewed the available methods for predicting MHC-peptide binding and discussed their most relevant advantages and drawbacks.

In many complex QSPR-like problems in immunoinformatics, like in other areas, we know the exact experimental result (known solution) of the problem, but we are interested in the possible result obtained after a change (perturbation) on one or multiple values of the initial conditions of the experiment (new solution). For instance, we often know, for large collections of i th molecules (m_i), organic compounds, drugs, xenobiotics, and/or peptide sequences, the efficiency of the compound $\epsilon(c_{ij})$ as adjuvant, action as epitope, immunotoxicity, and/or the interaction (affinity, inhibition, etc.) with immunological targets. In addition, we often know for each molecule the exact conditions (c_{ij}) of assay for the initial experiment including structure of the molecule m_i (drug, adjuvant, and sequence of the peptide), source organism (so), host organism (ho), immunological process (ip), experimental technique (tq), concentration, temperature, time, solvents, and coadjuvants. This is the case of big data retrieved from very large databases like IEDB [1–4] and ChEMBL [52]. However, we do not know the possible result of the experiment if we change at least one of these conditions (perturbation). We refer to small changes or perturbations in both structure and condition for input or output variables. It means that we include changes in ho, so, ip, and tq, changes of the compound by one analogue compound with similar structure, changes in the sequence of the epitope (artificial by organic synthesis or natural mutations), and polarity of the solvent or coadjuvants. In these cases, we could use a perturbation theory model to solve the QSAR/QSPR problem. Perturbation theory includes methods that add “small” terms to a known solution of a problem in order to approach a solution to a related problem without known solution. Perturbation models have been widely used in all branches of science from QM to astronomy and life sciences including chaos or “butterfly effect,” Bohr’s atomic theory, Heisenberg’s mechanics, Zeeman’s and Stark’s effects, and other models

with applications in like protein spectroscopy and others [53–57]. In a very recent work Gonzalez-Diaz et al. [58] formulated a general-purpose perturbation theory or model for multiple-boundary QSPR/QSAR problems. However, there is not report in the immunoinformatics literature of a general QSPR perturbation model for IEDB B-epitopes. Here we report the first example of QSPR-perturbation model for B-epitopes reported in IEDB able to predict the probability of occurrence of an epitope after a perturbation in the sequence, the experimental technique, the exposition process, and/or the source or host organisms.

2. Materials and Methods

2.1. Molecular Descriptors for Peptides. We calculated the molecular descriptors of the structure of peptides using the software MARCH-INSIDE (MI) based on the algorithm with the same name [59]. The MI approach uses a Markov Chain method to calculate the k th mean values of different physico-chemical molecular properties $\lambda(m_i)$ for i th molecules (m). These $\lambda(m_i)$ values are calculated as an average of ${}^k\lambda(m_i)$ values for all atoms placed at topological distance $d \leq k$; which are in turn the means of atomic properties (λ_j) for all atoms in the molecule and its neighbors placed at $d = k$. For instance, it is possible to derive average estimations of molecular refractivities ${}^kMR(m_i)$, partition coefficients ${}^kP(m_i)$, and hardness ${}^k\eta(m_i)$ for atoms placed at different topological distances $d \leq k$. In this first work, we calculated only one type of $\lambda(m_i)$ values. We calculated for all peptides the average value $\chi(m_i)$ of all the atomic electronegativities χ_i for all δ_i atoms connected to the i th atom ($i \rightarrow j$) and their neighbors placed at a distance $d \leq 5$ [59]:

$$\chi(m_i) = \frac{1}{6} \sum_{k=0}^5 {}^k\chi_j = \frac{1}{6} \sum_{k=0}^5 \sum_{i \rightarrow j}^{\delta_i} P_k(\chi_j) \cdot \chi_j. \quad (1)$$

We calculate the probabilities ${}^kP(\lambda_j)$ for any atomic property including ${}^kP(\chi_j)$ using a Markov Chain model for the gradual effects of the neighboring atoms at different distances in the molecular backbone. This method has been explained in detail in many previous works so we omit the details here [59].

2.2. Electronegativity Perturbation Model for Prediction of B-Epitopes. Very recently Gonzalez-Diaz et al. [58] formulated a general-purpose perturbation theory or model for multiple-boundary QSPR/QSAR problems. We adapted here this new theory or modeling method to approach to the peptide prediction problem from the point of view of perturbation theory. Let be a set of i th peptide molecules denoted as m_i with a value of efficiency ϵ_{ij} as epitopes experimentally determined under a set of boundary conditions $c_j \equiv (c_0, c_1, c_2, c_3, \dots, c_n)$. We put the main emphasis here on peptides reported in the database IEDB. In this sense, the boundary conditions c_j used here are the same reported in this database, $c_0 =$ is the specific

peptide, $c_1 = so_j$, $c_2 = ho_j$, $c_3 = ip_j$, and $c_4 = tq_j$. In general, so is the organism that expresses the peptide (but it can include also artificial peptides, cellular lines, etc.), ho is the host organism exposed to the peptide by means of the bp detected with tq . As our analysis, based on the data reported by IEDB we are unable to work with continuous values of epitope activity ε_{ij} . Consequently, we have to predict the discrete function of B-epitope efficiency $\lambda(\varepsilon_{ij}) = 1$ for epitopes reported in the conditions c_j and $\lambda(\varepsilon_{ij}) = 0$, otherwise. Our main aim is to predict the shift or change in a function of the output efficiency $\Delta\lambda(\varepsilon_{ij}) = \lambda(\varepsilon_{ij})_{\text{ref}} - \lambda(\varepsilon_{ij})_{\text{new}}$ that takes place after a change, variation, or perturbation (ΔV) in the structure and/or boundary conditions of a peptide of reference. But we know the efficiency of the process of reference $\lambda(\varepsilon_{ij})_{\text{ref}}$ in addition to the molecular structure and the set of conditions c_j for initial (reference) and final processes (new). Consequently, to predict $\Delta\lambda(\varepsilon_{ij})$ we have to predict only $\lambda(\varepsilon_{ij})_{\text{new}}$ the efficiency function of the new state obtained by a change in the structure of the peptide and/or the boundary conditions. Let ΔV be a perturbation in a function λ ; we can define V_{ij} as the state information function for the reference and new states. According to our recent model [58], we can write V_{ij} as a function of the conditions and structure of the peptide m_i as follows. In fact, the variational state functions V_{ij} have to be written in pairs in order to describe the initial (reference) and final (new) states of a perturbation, as follow:

$$V_{ij} = \lambda(\varepsilon_{ij})_{\text{new}} - \sum_{j=1}^4 \left(\lambda(m_i) - \lambda(c_{ij})_{\text{avg}} \right), \quad (2)$$

$$V_{qr} = \lambda(\varepsilon_{qr})_{\text{ref}} - \sum_{r=1}^4 \left(\lambda(m_q) - \lambda(c_{qr})_{\text{avg}} \right).$$

The state function ${}^n V_{ij}$ is for the i th peptide measured under a set of c_{ij} boundary conditions in output, final, or new state. The conjugated state function ${}^r V_{qr}$ is for the q th peptide measured under a set of c_{qr} boundary conditions for the input, initial, or reference state. The difference ΔV between the new (output) state and the reference (input) state is the additive perturbation [58]. Consider

$$\Delta V = V_{ij} - V_{qr} = \left[\lambda(\varepsilon_{ij})_{\text{new}} - \sum_{j=1}^4 \left(\lambda(m_i) - \lambda(c_{ij})_{\text{avg}} \right) \right] - \left[\lambda(\varepsilon_{qr})_{\text{ref}} - \sum_{r=1}^4 \left(\lambda(m_q) - \lambda(c_{qr})_{\text{avg}} \right) \right]. \quad (3)$$

Equation (3) described before opens the door to test different hypothesis. A simple hypotheses is H_0 : existence of one small and constant value of the perturbation function $\Delta V = e_0$ for all the pairs of peptides and a linear relationship

TABLE 1: Results of QSPR-perturbation model for IEDB B-Epitopes.

Data subset	Stat. param.	Pred. %	Predicted epitope perturbations $\lambda(\varepsilon_{ij}) = 1$	$\lambda(\varepsilon_{ij}) = 0$
$\lambda(\varepsilon_{ij}) = 1$	Sp	97.0	84607	2660
$\lambda(\varepsilon_{ij}) = 0$	Sn	93.6	4354	63548
Total train	Ac	95.5		
$\lambda(\varepsilon_{ij}) = 1$	Sp	97.1	28060	840
$\lambda(\varepsilon_{ij}) = 0$	Sn	93.3	1485	20641
Total cv	Ac	95.4		

Bold font is used to highlight the number of cases correctly classified by the model.

between perturbations of input/output boundary conditions with coefficients a_{ij} , b_{ij} , c_{qr} , and d_{ij} . Consider

$$e_0 = \Delta V = \left[a_{ij} \cdot \lambda(\varepsilon_{ij})_{\text{new}} - \sum_{j=1}^4 b_{ij} \cdot \left(\lambda(m_i) - \lambda(c_j)_{\text{avg}} \right) \right] - \left[c_{qr} \cdot \lambda(\varepsilon_{qr})_{\text{ref}} - \sum_{r=1}^4 d_{qr} \cdot \left(\lambda(m_q) - \lambda(c_r)_{\text{avg}} \right) \right]. \quad (4)$$

We can use elemental algebraic operations to obtain from these equations an expression for efficiency as epitope of the peptide $\lambda(\varepsilon_{ij})_{\text{new}}$. In this case, considering $b_{ij} \approx d_{qr}$, we can obtain the different expressions; the last may be very useful to solve the QSRR problem for the large datasets formed by IEDB B-epitopes. Consider

$$\lambda(\varepsilon_{ij})_{\text{new}} = \left(\frac{c_{qr}}{a_{ij}} \right) \cdot \lambda(\varepsilon_{qr})_{\text{ref}} + \left[\sum_{j=1}^4 \left(\frac{b_{qr}}{a_{ij}} \right) \cdot \left(\lambda(m_i) - \lambda(c_j)_{\text{avg}} \right)_{\text{new}} \right] - \left[\sum_{r=1}^4 \left(\frac{d_{qr}}{a_{ij}} \right) \cdot \left(\lambda(m_q) - \lambda(c_r)_{\text{avg}} \right)_{\text{ref}} \right] + \left(\frac{e_0}{a_{ij}} \right), \quad (5)$$

$$\lambda(\varepsilon_{ij})_{\text{new}} = {}^r c_0 \cdot \lambda(\varepsilon_{qr})_{\text{ref}} + \sum_{j=1}^4 {}^r d_{ij} \cdot \Delta \left(\lambda(m_i) - \lambda(c_j)_{\text{avg}} \right) + {}^r e_0,$$

$$\lambda(\varepsilon_{ij})_{\text{new}} = {}^r c_0 \cdot \lambda(\varepsilon_{qr})_{\text{ref}} + \sum_{j=1}^4 {}^r d_{ij} \cdot \Delta \Delta \lambda_{ijqr} + {}^r e_0,$$

$$\lambda(\varepsilon_{ij})_{\text{new}} = {}^r c_0 \cdot \lambda(\varepsilon_{qr})_{\text{ref}} + \sum_{j=1}^4 {}^r d_{ij} \cdot \Delta \Delta \chi_{ijqr} + {}^r e_0.$$

TABLE 2: Average values and count of input-output cases for different organisms, process, and techniques.

Source organism (so)	N_{in}	N_{out}	* χ
<i>Homo sapiens</i>	38920	39274	2.685
<i>Plasmodium falciparum</i>	10669	9446	2.704
<i>Hepatitis C virus</i>	9935	10239	2.683
<i>Bos taurus</i>	5671	5780	2.690
<i>Canine parvovirus</i>	5655	5637	2.693
<i>Foot-mouth disease virus</i>	4062	4176	2.676
<i>Triticum aestivum</i>	3769	3887	2.703
<i>Bacillus anthracis</i>	3602	3600	2.699
<i>Human papillomavirus</i>	3316	3414	2.693
<i>Human herpesvirus</i>	3026	3132	2.684
<i>Gallus gallus</i>	2850	2829	2.689
<i>Arachis hypogaea</i>	2648	2670	2.687
<i>Mycobacterium tuberculosis</i>	2637	2593	2.688
<i>Clostridium botulinum</i>	2588	2722	2.685
<i>SARS coronavirus</i>	2550	2704	2.686
<i>Mus musculus</i>	2334	2287	2.682
<i>Hepatitis B virus</i>	2007	2066	2.680
<i>Helicobacter pylori</i>	1958	1796	2.695
<i>Hevea brasiliensis</i>	1938	1958	2.697
<i>Hepatitis E virus</i>	1928	1941	2.685
<i>Shigella flexneri</i>	1878	1701	2.699
<i>Dengue virus 2</i>	1767	1828	2.679
<i>Staphylococcus aureus</i>	1757	1661	2.694
<i>Treponema pallidum</i>	1739	1755	2.691
<i>Escherichia coli</i>	1721	1678	2.689
<i>Murine hepatitis virus</i>	1575	1603	2.692
<i>Haemophilus influenzae</i>	1545	1587	2.695
<i>Streptococcus mutans</i>	1523	1537	2.697
<i>Puumala virus (strain)</i>	1505	1574	2.689
<i>Chlamydia trachomatis</i>	1402	1546	2.704
<i>Human respiratory virus</i>	1347	1398	2.682
<i>Borrelia burgdorferi</i>	1228	1237	2.698
<i>Hepatitis delta virus</i>	1182	1199	2.690
<i>Streptococcus pyogenes</i>	1181	1251	2.697
<i>Porphyromonas gingivalis</i>	1143	1085	2.688
<i>Human enterovirus</i>	1106	1132	2.689
<i>Influenza A virus</i>	1085	1086	2.687
<i>Mycoplasma hyopneumoniae</i>	1044	1024	2.695
<i>Rattus norvegicus</i>	1025	1039	2.689
<i>Bordetella pertussis</i>	1011	960	2.685
<i>Human T-lymphotropic virus</i>	996	1031	2.680
<i>Anaplasma marginale</i>	977	857	2.707
<i>Measles virus strain</i>	804	810	2.688
<i>Fasciola hepatica</i>	803	857	2.685
<i>Neisseria meningitidis</i>	789	853	2.696
<i>Human poliovirus</i>	766	780	2.690
<i>Tityus serrulatus</i>	764	775	2.680
<i>Torpedo californica</i>	752	788	2.687

TABLE 2: Continued.

Source organism (so)	N_{in}	N_{out}	* χ
<i>Cryptomeria japonica</i>	719	794	2.680
<i>Mycobacterium bovis</i>	717	733	2.688
<i>Trypanosoma cruzi</i>	691	777	2.704
<i>Andes virus CHI-7913</i>	679	687	2.690
<i>Bovine papillomavirus</i>	672	665	2.692
<i>Human hepatitis</i>	670	696	2.688
<i>Leishmania infantum</i>	659	735	2.688
<i>Human parvovirus</i>	649	691	2.683
<i>Poa pratensis</i>	648	664	2.692
<i>Aspergillus fumigatus</i>	642	709	2.677
<i>Duck hepatitis</i>	587	603	2.688
<i>Olea europaea</i>	571	577	2.692
<i>Porcine reproductive</i>	515	514	2.681
<i>Fagopyrum esculentum</i>	509	497	2.685
<i>Juniperus ashei</i>	505	568	2.672
<i>Mycobacterium leprae</i>	489	542	2.690
<i>Glycine max</i>	477	509	2.685
<i>D. pteronyssinus</i>	455	464	2.680
<i>Plasmodium vivax</i>	453	446	2.690
<i>Chlamydomydia pneumoniae</i>	446	462	2.690
<i>Pseudomonas aeruginosa</i>	443	454	2.691
<i>Vibrio cholera</i>	427	426	2.694
<i>Streptococcus sp.</i>	426	425	2.691
<i>Mycobacterium avium</i>	425	415	2.689
<i>Dermatophagoides farinae</i>	410	390	2.693
<i>Human coxsackievirus</i>	406	392	2.694
<i>Equine infectious virus</i>	404	419	2.688
<i>Babesia equi</i>	383	371	2.696
<i>Prunus dulcis</i>	383	379	2.708
<i>Human adenovirus</i>	375	405	2.686
<i>Theileria parva</i>	366	371	2.713
<i>Candida albicans</i>	365	370	2.690
<i>Porcine endogenous</i>	355	351	2.692
<i>Ovis aries</i>	352	350	2.683
<i>Chironomus thummi</i>	347	338	2.691
<i>Sus scrofa</i>	343	362	2.686
<i>Bovine leukemia virus</i>	333	329	2.676
<i>Ricinus communis</i>	329	314	2.692
<i>Androctonus australis</i>	322	357	2.685
<i>Renibacterium salmoninarum</i>	319	350	2.690
<i>Orientia tsutsugamushi</i>	309	372	2.705
<i>Anacardium occidentale</i>	293	306	2.693
<i>Conus geographus</i>	289	295	2.660
Host organism (ho)	N_{in}	N_{out}	* χ
<i>Homo sapiens</i>	257293	91093	2.6856
<i>Mus musculus</i>	107867	51466	2.6873
<i>Oryctolagus cuniculus</i>	65053	31433	2.6900
<i>Bos taurus</i>	15333	2072	2.6909
<i>Rattus norvegicus</i>	9450	3562	2.6876
<i>Aotus sp.</i>	9044	3933	2.6879
<i>Sus scrofa</i>	7725	3464	2.6873
<i>Gallus gallus</i>	7507	997	2.6790

TABLE 2: Continued.

Source organism (so)	N_{in}	N_{out}	$^*\chi$
<i>Canis lupus</i>	6604	3334	2.6906
<i>Macaca mulatta</i>	5261	2569	2.6993
<i>Ovis aries</i>	3953	1653	2.6836
<i>Equus caballus</i>	3943	2099	2.6842
<i>Cavia porcellus</i>	3458	1688	2.6833
<i>Capra hircus</i>	2182	1127	2.6830
<i>Aotus nancymaae</i>	1659	852	2.6837
<i>Pan troglodytes</i>	1614	732	2.6757
<i>Marmota monax</i>	1100	509	2.7011
<i>Felis catus</i>	901	279	2.6838
<i>Myodes glareolus</i>	814	388	2.6863
<i>Anas platyrhynchos</i>	688	342	2.6880
<i>Homo sapiens (human)</i>	508	270	2.6851
<i>Trichosurus vulpecula</i>	456	126	2.6921
<i>Mesocricetus auratus</i>	438	104	2.6909
<i>Macaca cyclopis</i>	382	193	2.6871
<i>O. tshawytscha</i>	333	159	2.6929
<i>Macaca fuscata</i>	188	100	2.6667
<i>Cricetulus migratorius</i>	171	142	2.7008
<i>Camelus dromedarius</i>	171	89	2.6886
<i>Dicentrarchus labrax</i>	121	55	2.6759
<i>Macaca fascicularis</i>	96	52	2.6793
<i>Saimiri sciureus</i>	92	44	2.6900
<i>Canis familiaris</i>	77	42	2.6850
<i>Rattus rattus</i>	72	31	2.6760
<i>Callithrix pygmaea</i>	67	30	2.6920
<i>Chinchilla lanigera</i>	41	24	2.6729
<i>Aotus lemurinus</i>	30	19	2.6860
<i>Papio cynocephalus</i>	27	13	2.7267
<i>Aotus griseimembra</i>	26	12	2.7000
<i>Mustela vison</i>	18	10	2.7000
<i>Chlorocebus aethiops</i>	15	10	2.6875
<i>Bos indicus</i>	13	4	2.6925
<i>Oncorhynchus mykiss</i>	10	4	2.6700
<i>M. macquariensis</i>	9	6	2.6600
<i>Cricetulus griseus</i>	8	4	2.6900
<i>Aotus trivirgatus</i>	7	4	2.7000
Process type (pt)	N_{in}	N_{out}	$^*\chi$
AID	111197	108536	2.6876
OOD	32419	32617	2.6868
OAID	19210	18954	2.6801
OOA	15863	16303	2.6902
NI	13430	15206	2.6845
EWEIR	4818	4864	2.6843
EEE	3113	3546	2.6906
OOD	2806	2799	2.6887
AICD	1077	1095	2.6812
EWED	696	686	2.6879
DEWED	280	337	2.6804
TT	260	215	2.6806
OOC	153	137	2.6800

TABLE 2: Continued.

Source organism (so)	N_{in}	N_{out}	$^*\chi$
Technique (tq)	N_{in}	N_{out}	$^*\chi$
ELISA	133458	135109	2.6871
WI	33627	33292	2.6887
ACAbB	7780	9068	2.6862
PhDIP	7450	4496	2.6771
RIA	5241	5218	2.6858
IFAIH	4454	4581	2.6879
NIAA	4222	4316	2.6892
FIA	2255	2276	2.6897
PAC	1312	1219	2.6837
IP	1127	1089	2.6886
SPR	758	639	2.6860
FACS	608	647	2.6907
Other	502	495	2.6813
SAC	484	393	2.6878
ELISPOT	396	412	2.6979
RDAT	366	323	2.6859
EDAT	284	330	2.6800
XRC	231	227	2.6880
MS	209	179	2.6849
PFF	171	153	2.6820
AbDPO	162	295	2.6968
CdC	146	205	2.6895
IAbBA	144	183	2.6940
IOT	124	106	2.6835
HAGGI	115	122	2.6834
IgMHR	89	90	2.6929
EAAA	84	139	2.6922
HS	82	67	2.6791
AbdCC	73	118	2.6897
AGG	50	60	2.6980
CM	50	57	2.6863

The * indicates that quantities like $^*\chi$ is the average value of the mean electronegativity (m_i) for all the peptides in IEDB that are epitopes for the same boundary condition.

3. Results and Discussion

We propose herein, for the first time, a QSRR-perturbation model able to predict variations in the propensity of a peptide to act as B-epitope taking into consideration the propensity of a peptide of reference and the changes in peptide sequence, immunological process, host organism, source organisms, and the experimental technique used. The best QSPR-perturbation model found here with LDA was

$$\begin{aligned}
\lambda(\varepsilon_{ij})_{new} = & 4.979 \cdot \lambda(\varepsilon_{ij})_{ref} - 221.642 \cdot \Delta\chi_{seq} \\
& + 8.770 \cdot \Delta\Delta\chi_{ho} + 63.572 \cdot \Delta\Delta\chi_{so} \\
& - 55.387 \cdot \Delta\Delta\chi_{ip} + 201.919 \cdot \Delta\Delta\chi_{tq} - 2.149, \\
N = & 155169, \quad Rc = 0.92, \\
U = & 0.15, \quad p < 0.01.
\end{aligned}
\tag{6}$$

TABLE 3: Top100 values of p1 for positive perturbations in training series.

IDE	New experiment				Experiment of reference				Input perturbation terms						
	Sequence	ho	so		Sequence	ho	so		$\Delta\chi$	ip	tq	$\Delta\chi_{ho}$	$\Delta\chi_{ip}$	$\Delta\chi_{so}$	$\Delta\chi_{tq}$
115153	MKGVVC TRIYEKV	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	<i>Mus musculus</i>	<i>Dengue virus 2</i>	AID	WI	0.01	0.012	0.004	0.017	0.012	0.012
52124	QQQPP	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OOA	ELISA	<i>Homo sapiens</i>	<i>Glycine max</i>	OOA	WI	0	0	-0.018	0	0.002	0.002
3639	APLGVT	<i>Homo sapiens</i>	<i>Hepatitis E virus</i>	EWEIR	ELISA	<i>Homo sapiens</i>	<i>Human herpesvirus</i>	OOID	ELISA	0.04	0.04	0.039	0.043	0.04	0.04
135959	LTRAYA KDVKFG	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	<i>Mus musculus</i>	MD	AID	ELISA	-0.04	-0.038	-0.047	-0.033	-0.04	-0.04
108075	PREPQVY	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	<i>Oryctolagus cuniculus</i>	<i>Homo sapiens</i>	AID	ACAbB	-0.01	-0.006	-0.01	-0.003	-0.011	-0.011
25113	HVVLDLP	<i>Homo sapiens</i>	<i>Hepatitis E virus</i>	EWEIR	ELISA	<i>Mus musculus</i>	MD	AID	ELISA	0.02	0.022	0.013	0.023	0.02	0.02
48780	PPFSPQ	<i>Homo sapiens</i>	<i>Hepatitis delta virus</i>	OOID	ELISA	<i>Mus musculus</i>	MD	AID	SAC	0.02	0.022	0.008	0.021	0.021	0.021
40988	LYVVAYQA	<i>Mus musculus</i>	<i>Viscum album</i>	AID	ELISA	<i>Homo sapiens</i>	<i>Hepatitis B virus</i>	OOID	ELISA	0.02	0.018	0.006	0.019	0.02	0.02
50439	QDAYNAAG	<i>Mus musculus</i>	<i>Mycobacterium scrofulaceum</i>	AID	ELISA	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0.04	0.038	0.028	0.039	0.04	0.04
98849	KIPAVFKIDA	<i>Homo sapiens</i>	<i>Bos taurus</i>	DEWED	WI	<i>Homo sapiens</i>	<i>Arachis hypogaea</i>	OOA	IFAIH	-0.05	-0.05	-0.053	-0.04	-0.051	-0.051
116171	TQDQDP BBHFFK NIVTPR	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ACAbB	<i>Oryctolagus cuniculus</i>	MD	AID	MS	0.01	0.014	0.008	0.017	0.009	0.009
123442	LLKDLRKN	<i>Homo sapiens</i>	<i>Borna disease virus</i>	EWEIR	WI	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OOID	ELISA	-0.02	-0.02	-0.025	-0.017	-0.022	-0.022
47858	PHVVDL	<i>Homo sapiens</i>	<i>Hepatitis E virus</i>	EWEIR	ELISA	<i>Capra hircus</i>	<i>Human T-lymphotropic virus</i>	AID	ACAbB	0.01	0.007	0.005	0.013	0.009	0.009
61783	STNKAVVSL	<i>Bos taurus</i>	<i>Bovine respiratory</i>	AID	ELISA	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0.03	0.025	0.019	0.029	0.03	0.03
118210	VMLYQISEE	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	WI	<i>Oryctolagus cuniculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.05	0.054	0.05	0.057	0.048	0.048
130944	LFKHS	<i>Oryctolagus cuniculus</i>	<i>Rattus norvegicus</i>	AID	ELISA	<i>Homo sapiens</i>	MD	OOD	WI	0.01	0.006	-0.002	0.011	0.012	0.012

TABLE 3: Continued.

IDE	Sequence	New experiment			Experiment of reference			Input perturbation terms								
		ho	so	ip	ho	so	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$			
23028	GVKYA	<i>Homo sapiens</i>	MD	OAID	WI	23032	GVLAKD VRFSQV	<i>Homo sapiens</i>	MD	OOID	ELISA	0	0	0	0.007	-0.002
51199	QKKAIE	<i>Oryctolagus cuniculus</i>	<i>Vibrio cholerae</i>	AID	ELISA	51204	QKKNK RNTNRR PQDV	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0.03	0.026	0.018	0.029	0.03
144783	SHVVT MLDNF	<i>Homo sapiens</i>	<i>Homo sapiens</i>	NI	ELISA	144786	SMNRGRG THPSLIWM	<i>Mus musculus</i>	MD	AID	ACAbB	0.03	0.032	0.023	0.033	0.029
134343	DLYIK	<i>Mus musculus</i>	<i>Human papillomavirus</i>	AID	NIAA	134344	DMAQV TVGPGLL GVSTL	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	WI	0	0	-0.009	0	0
38321	LNQLAGRM	<i>Anas platyrhynchos</i>	<i>Duck hepatitis</i>	AID	ELISA	38323	LNQTAR AFDCAI CWEPSP	<i>Oryctolagus cuniculus</i>	<i>Bovine leukemia virus</i>	AID	ACAbB	-0.01	-0.008	-0.022	-0.01	-0.011
144657	GQITVD MMYG	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	144661	ADGGCA WPACYC	<i>Oryctolagus cuniculus</i>	MD	AID	WI	0.02	0.024	0.013	0.027	0.022
21084	GLQN	<i>Mus musculus</i>	<i>Chlamydia trachomatis</i>	AID	ELISA	21093	GLRAQD DFSGWDI NTPAFEW	<i>Mus musculus</i>	<i>Mycobacterium tuberculosis</i>	AID	WI	0.03	0.03	0.013	0.03	0.032
98453	SGFSGVQFV	<i>Oryctolagus cuniculus</i>	<i>Neisseria meningitidis</i>	AID	ELISA	98456	SICSNN PTCWAIC KRIPNKK	<i>Mus musculus</i>	<i>Human respiratory virus</i>	AID	IFAIH	0.04	0.037	0.027	0.04	0.041
98453	SGFSGVQFV	<i>Mus musculus</i>	<i>Neisseria meningitidis</i>	AID	ELISA	98456	SICSNN PTCWAIC KRIPNKK	<i>Mus musculus</i>	<i>Human respiratory virus</i>	AID	IFAIH	0.04	0.04	0.027	0.04	0.041
107107	EAIQP	<i>Rattus norvegicus</i>	<i>Homo sapiens</i>	AID	ELISA	107110	EKERRP SPIGTATLL	<i>Homo sapiens</i>	MD	OOA	ELISA	0.05	0.048	0.043	0.053	0.05
110857	FTGEAY SYWSAK	<i>Homo sapiens</i>	<i>Mycoplasma penetrans</i>	EWEIR	ELISA	110859	GEEERIS LPLPNF SSLNRE	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	FIA	0	0.002	-0.017	0.003	0.003
36315	LGSAYP	<i>Mus musculus</i>	<i>Mycobacterium leprae</i>	MD	ELISA	36317	LGSAGFG TIVYKG	<i>Mus musculus</i>	<i>Avian erythroblastosis virus</i>	AID	ACAbB	0.01	0.01	0	0.011	0.009
122034	WNPAD	<i>Rattus norvegicus</i>	<i>Torpedo californica</i>	AID	ELISA	122035	WNPAD YGGIKWN PADYGGIK	<i>Rattus norvegicus</i>	MD	AID	RIA	0.01	0.01	0.001	0.01	0.009
25013	HVADIDKLLID	<i>Mus musculus</i>	<i>Puumala virus Kazan</i>	AID	ELISA	25021	HVAPTH YV'TESDA SQRVTQL	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0	-0.002	-0.014	-0.001	0
36162	LGIHE	<i>Oryctolagus cuniculus</i>	<i>Candida albicans</i>	AID	ELISA	36166	LGIMGE YRGTPRN QDLYDAA	<i>Mus musculus</i>	<i>Human respiratory virus</i>	AID	RIA	0	-0.003	-0.007	0	-0.001
67253	TWEVLH	<i>Mus musculus</i>	<i>Plasmodium vivax</i>	AID	ELISA	67257	TWGEN ETDVLIL NNTTRPPQ	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ACAbB	-0.02	-0.022	-0.027	-0.021	-0.021

TABLE 3: Continued.

IDE	New experiment				Experiment of reference				Input perturbation terms							
	Sequence	ho	so	ip	tq	IDE	Sequence	ho	so	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$
50990	QGYRVSSYLP	<i>Homo sapiens</i>	<i>Hevea brasiliensis</i>	OOA	WI	50998	QHEQDR PTPSPAP SRPFSVL	<i>Homo sapiens</i>	<i>Hepatitis E virus</i>	OOID	ELISA	0.01	0.01	-0.002	0.007	0.008
100458	RDVLIQYAPE	<i>Mus musculus</i>	<i>Bacillus anthracis</i>	AID	ELISA	100462	RESTRY GNQNGRI RVLQRED	<i>Homo sapiens</i>	<i>Arachis hypogaea</i>	EWED	ELISA	0.03	0.028	0.018	0.03	0.03
111036	TESTFT GEAYSV	<i>Homo sapiens</i>	<i>Mycoplasma penetrans</i>	EWEIR	ELISA	111039	TGVPID PAVPDSS IVPLES	<i>Bos taurus</i>	<i>Bovine papillomavirus</i>	AID	ELISA	0.03	0.035	0.02	0.033	0.03
117919	IFIEME	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	WI	117921	IGHDLIE KRKFNQ	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	WI	0.03	0.032	0.03	0.037	0.03
7127	CTDTDKLF	<i>Oryctolagus cuniculus</i>	<i>Shigella flexneri</i>	AID	ELISA	7128	CTDVST AIHADQL TPAW	<i>Homo sapiens</i>	<i>SARS coronavirus</i>	OOID	ELISA	0.01	0.006	-0.003	0.009	0.01
112253	PGQSPKL	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	112255	PIRALV GDEVELP CRISPGK	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.01	0.012	0.01	0.017	0.01
122034	WNPAD	<i>Rattus norvegicus</i>	<i>Torpedo californica</i>	AID	ELISA	122038	WNPDDY GGYKWNP DDYGGVK	<i>Rattus norvegicus</i>	MD	AID	RIA	0	0	-0.009	0	-0.001
131878	FLMLVG GSTL	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ACAAB	131879	FLVAHT RARAPSA GERARRS	<i>Mus musculus</i>	<i>Mus musculus</i>	AID	NIAA	0.03	0.032	0.028	0.037	0.033
70664	VQVVDYDQ	<i>Homo sapiens</i>	<i>Treponema pallidum</i>	OOID	ELISA	70667	VQWMNR LIAFAFAG NHVSP	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0.05	0.05	0.041	0.05	0.05
71545	VTV	<i>Homo sapiens</i>	<i>Helicobacter pylori</i>	OOID	ELISA	71559	VTVRGGL RILSPDRK	<i>Homo sapiens</i>	<i>Arachis hypogaea</i>	OOA	WI	0.04	0.04	0.032	0.043	0.042
127856	TDVRYKD	<i>Mus musculus</i>	<i>Mus musculus</i>	AID	ACAAB	127857	TDVRYK DDMYHFF CPAFAQQ	<i>Mus musculus</i>	<i>Mus musculus</i>	AID	PFF	0.01	0.01	0.01	0.01	0.006
112149	GVGWIRQ	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	112152	HHPART AHYGSLP QKSHGRT	<i>Homo sapiens</i>	<i>Homo sapiens</i>	AID	ELISA	0	0	0	0.007	0
119581	FSCVMHE	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	119582	GLQLJQL INVDVFNQI	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	RIA	-0.01	-0.008	-0.01	-0.003	-0.011
144657	GQITVD MMYG	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	144659	GREGYP ADGGAA GYCNTE	<i>Oryctolagus cuniculus</i>	MD	AID	WI	-0.01	-0.006	-0.017	-0.003	-0.008
25013	HVADI DKLID	<i>Mus musculus</i>	<i>Puumala virus Kazan</i>	AID	ELISA	25022	HVAPTH YVVESDA SQRVTQV	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0	-0.002	-0.014	-0.001	0

TABLE 3: Continued.

IDE	New experiment		Experiment of reference					Input perturbation terms								
	Sequence	ho	so	ip	tq	IDE	Sequence	ho	so	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$
144652	GMRGM KGLVY	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	144654	GPHPTLE VVPVMGRGS	<i>Mus musculus</i>	MD	AID	ELISA	-0.02	-0.018	-0.027	-0.013	-0.02
104515	HDCRPKKI	<i>Mus musculus</i>	<i>La Crosse virus</i>	AID	IFAIH	104521	IGTLKKIL DETVKD KIAKEQ	<i>Rattus norvegicus</i>	<i>Streptococcus pyogenes</i>	AID	ELISA	-0.04	-0.04	-0.053	-0.04	-0.041
7367	CYGDWA	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OOA	ELISA	7374	EYKLPDS EPTKTNGK	<i>Mus musculus</i>	<i>Tityus serrulatus</i>	AID	WI	-0.02	-0.018	-0.043	-0.023	-0.018
7367	CYGDWA	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OOA	ELISA	7374	CYGLPDS EPTKTNGK	<i>Mus musculus</i>	<i>Tityus serrulatus</i>	AID	WI	-0.02	-0.018	-0.043	-0.023	-0.018
144610	DFFTYK	<i>Mus musculus</i>	<i>Porcine transmissible</i>	AID	WI	144611	DFNGSF DMNGTITA	<i>Oryctolagus cuniculus</i>	<i>Escherichia coli</i>	AID	ELISA	-0.01	-0.007	-0.018	-0.01	-0.012
112047	ASTRESG	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	112048	ATASTM DHARHGF LPRHRDT	<i>Homo sapiens</i>	<i>Homo sapiens</i>	AID	ELISA	0.05	0.05	0.05	0.057	0.05
36136	LGGVFT	<i>Homo sapiens</i>	<i>Dengue virus 2</i>	OOID	ELISA	36137	LGGWKIQ SDPRAYAL	<i>Homo sapiens</i>	<i>Ambrosia artemisiifolia</i>	OOA	RIA	0.01	0.01	0.007	0.013	0.009
115256	FRELKD LKGY	<i>Homo sapiens</i>	<i>Bos taurus</i>	DEWED	WI	115261	GDLEILL QKWENG ECAQKKI	<i>Homo sapiens</i>	<i>Bos taurus</i>	OOA	FIA	-0.01	-0.01	-0.01	0	-0.009
129024	KADQLYK	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	129026	KAKKP AAAAGA KKAKS	<i>Oryctolagus cuniculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.03	0.034	0.03	0.037	0.03
148481	YTRDLVYK	<i>Rattus norvegicus</i>	<i>Homo sapiens</i>	AID	WI	148483	YVPIVT FYSEIM HSSRAIP	<i>Oryctolagus cuniculus</i>	MD	AID	ELISA	0	0.002	-0.007	0	-0.002
150850	GY	<i>Mus musculus</i>	<i>Human papillomavirus</i>	AID	ELISA	150853	HIGGLSI LDPIFGVL	<i>Homo sapiens</i>	<i>Dermatophagoides farinae</i>	OOA	ACAbB	0.04	0.038	0.04	0.043	0.039
107366	FPPKPKD	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	107376	GDRSGYS SPGSPG	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	ACAbB	-0.03	-0.028	-0.03	-0.023	-0.031
114859	ICGTD GVTYT	<i>Homo sapiens</i>	<i>Gallus gallus</i>	OOA	WI	114865	IVERETR QSENPL WHALRR	<i>Rattus norvegicus</i>	<i>Human herpesvirus</i>	AID	ELISA	0.04	0.042	0.035	0.037	0.038
62149	SVHLF	<i>Homo sapiens</i>	MD	OAID	WI	62150	SVIALGS QEGALHQ ALAGAI	<i>Equus caballus</i>	<i>West Nile virus</i>	AID	IFAIH	-0.02	-0.021	-0.012	-0.013	-0.021
98455	SGSVQFVIQ	<i>Mus musculus</i>	<i>Neisseria meningitidis</i>	AID	ELISA	98456	SICSNP TCWAICK RIPNKK	<i>Mus musculus</i>	<i>Human respiratory virus</i>	AID	IFAIH	0.04	0.04	0.027	0.04	0.041
61783	STNKAV VSL	<i>Bos taurus</i>	<i>Bovine respiratory</i>	AID	ELISA	61791	STNPKPQ RKTGRNT NRRPQ	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	EWEIR	ACAbB	0.04	0.035	0.029	0.037	0.039
100318	NAPKT FQFIN	<i>Mus musculus</i>	<i>Bacillus anthracis</i>	AID	ELISA	100319	NASSELH LLGFGIN AENHR	<i>Homo sapiens</i>	<i>Arachis hypogaea</i>	EWED	ELISA	0	-0.002	-0.012	0	0

TABLE 3: Continued.

IDE	Sequence	New experiment			Experiment of reference			Input perturbation terms							
		ho	so	ip	ho	so	ip	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$	
118947	PFSAPPPA	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	118948	<i>Homo sapiens</i>	<i>Human herpesvirus</i>	NI	ACAbB	-0.04	-0.04	-0.041	-0.036	-0.041
78323	YFRD	<i>Mus musculus</i>	<i>Bluetongue virus 1</i>	AID	ELISA	78341	<i>Homo sapiens</i>	<i>Streptococcus mutans</i>	EEE	ELISA	0.01	0.008	0.007	0.013	0.01
145831	IPLGTRP	<i>Mus musculus</i>	<i>Human papillomavirus</i>	AID	ELISA	145841	<i>Sus scrofa</i>	<i>Classical swine</i>	AID	PAC	-0.04	-0.04	-0.045	-0.04	-0.043
119592	HTFPAVLQ	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	119596	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	RIA	0.01	0.012	0.01	0.017	0.009
58780	SKAANLSIK	MD	<i>Beet necrotic</i>	MD	WI	58783	<i>Oryctolagus cuniculus</i>	<i>Influenza A virus</i>	AID	RIA	-0.01	-0.004	-0.014	-0.009	-0.013
115153	MKGVVC TRIVEKV	<i>Homo sapiens</i>	<i>Homo sapiens</i>	NI	ELISA	115155	<i>Mus musculus</i>	<i>Dengue virus 2</i>	AID	ELISA	0.01	0.012	0.004	0.013	0.01
133629	LPLRF	<i>Oryctolagus cuniculus</i>	<i>Gallus gallus</i>	AID	ACAbB	133630	<i>Mus musculus</i>	MD	AID	SPR	-0.01	-0.013	-0.021	-0.01	-0.01
96215	EEEEAE DKED	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	96216	<i>Mus musculus</i>	<i>Mus musculus</i>	AID	ELISA	0.08	0.082	0.078	0.087	0.08
39782	LTAASV	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OAA	ELISA	39788	<i>Oryctolagus cuniculus</i>	<i>Yersinia pestis</i>	AID	ACAbB	0.01	0.014	-0.001	0.007	0.009
107479	KFNWYVD	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	107482	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	ACAbB	0	0.002	0	0.007	-0.001
63569	TETVNSDI	<i>Macaca mulatta</i>	<i>Shigella flexneri</i>	AID	ELISA	63573	<i>Homo sapiens</i>	<i>Mycobacterium leprae</i>	Ooid	ELISA	0.05	0.036	0.041	0.049	0.05
134028	DDTIS	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	NIAA	134029	<i>Oryctolagus cuniculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.05	0.053	0.05	0.05	0.048
23028	GVKYA	<i>Homo sapiens</i>	MD	OAID	WI	23032	<i>Homo sapiens</i>	MD	EWEIR	ACAbB	0	0	0	0.004	-0.003
115293	IMCVKK ILDK	<i>Homo sapiens</i>	<i>Bos taurus</i>	DEWED	WI	115295	<i>Homo sapiens</i>	<i>Bos taurus</i>	OAA	FIA	-0.03	-0.03	-0.03	-0.02	-0.029
65105	TLTPENTL	<i>Mus musculus</i>	<i>Shigella flexneri</i>	AID	ELISA	65110	<i>Oryctolagus cuniculus</i>	<i>SARS coronavirus</i>	AID	ELISA	0.01	0.013	-0.003	0.01	0.01
134471	PKPEQ	<i>Mus musculus</i>	<i>Streptococcus pneumoniae</i>	AID	FACS	134472	<i>Homo sapiens</i>	<i>Hepatitis B virus</i>	AID	ELISA	0.02	0.018	0.019	0.02	0.016

TABLE 3: Continued.

IDE	New experiment		Experiment of reference				Input perturbation terms									
	Sequence	ho	so	ip	tq	IDE	Sequence	ho	so	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$
142228	LYCYEQ LNDSSE	<i>Homo sapiens</i>	<i>Human papillomavirus</i>	NI	ELISA	142250	NWGDEP SKRRDRS NSRGRKN	<i>Felis catus</i>	<i>Feline infectious</i>	AID	ELISA	0.07	0.068	0.071	0.073	0.07
21549	GNYPDFW YQS	<i>Homo sapiens</i>	<i>Staphylococcus aureus</i>	OOID	ELISA	21553	GNYNKY RYLRHGK LRPFER	<i>Mus musculus</i>	<i>SARS coronavirus</i>	AID	ELISA	0.03	0.032	0.021	0.031	0.03
21549	GNYPDFWYQS	<i>Homo sapiens</i>	<i>Staphylococcus aureus</i>	OOID	ELISA	21553	GNYNKY RYLRHGK LRPFER	<i>Oryctolagus cuniculus</i>	<i>SARS coronavirus</i>	AID	ELISA	0.03	0.034	0.021	0.031	0.03
34908	LAPLGE	<i>Homo sapiens</i>	<i>Hepatitis E virus</i>	EWEIR	ELISA	34914	LAPSTL RSLRKR RLSSP	<i>Homo sapiens</i>	<i>Human herpesvirus</i>	OOID	ELISA	0.06	0.06	0.059	0.063	0.06
130454	CLFPNNSYC	<i>Mus musculus</i>	MD	AID	NIAA	130456	CRPQVN NPKEWS CAAC	<i>Homo sapiens</i>	MD	OOD	ACAbB	0.01	0.008	0.01	0.011	0.007
19644	GFVPSM	<i>Homo sapiens</i>	<i>Hepatitis delta virus</i>	OOID	ELISA	19647	GFVSASI FGFOAEV GPNTR	<i>Oryctolagus cuniculus</i>	<i>Vaccinia virus WR</i>	AID	ELISA	-0.01	-0.006	-0.013	-0.009	-0.01
20678	GKRPE	<i>Mus musculus</i>	<i>Streptococcus pyogenes</i>	AID	WI	20680	GKSKRD AKNNAAK LAVDKLL	<i>Mus musculus</i>	<i>Vaccinia virus WR</i>	AID	WI	0.01	0.01	0.001	0.01	0.01
123278	GYLKDLPTT	<i>Ovis aries</i>	<i>Fasciola hepatica</i>	AID	ELISA	123282	HACQKK LLKFEAL QQEEGEE	<i>Rattus norvegicus</i>	<i>Gallus gallus</i>	AID	PFF	-0.02	-0.016	-0.016	-0.02	-0.025
141067	PLSLEPDP	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	FACS	141073	REGVRW RVMAIQ	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.06	0.06	0.06	0.06	0.056
139248	SFAGTVIE	<i>Mus musculus</i>	<i>Classical swine</i>	AID	WI	139305	TAAQITQ RKWEAA REAEQRR	<i>Oryctolagus cuniculus</i>	<i>Homo sapiens</i>	AID	IP	0.03	0.033	0.026	0.03	0.03
104515	HDCRPKKI	<i>Mus musculus</i>	<i>La Crosse virus</i>	AID	IFAIH	104520	IAKEQE NKEITGT LKKILDE	<i>Rattus norvegicus</i>	<i>Streptococcus pyogenes</i>	AID	ELISA	-0.06	-0.06	-0.073	-0.06	-0.061
113517	HLYADGLTD	<i>Mus musculus</i>	<i>Human papillomavirus</i>	AID	IFAIH	113518	HNKIQA IELEDLLR YSKLYR	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	ELISA	0.02	0.02	0.011	0.02	0.019
156970	VERHQ	<i>Homo sapiens</i>	<i>Homo sapiens</i>	OAID	ELISA	156975	WSSTV LRVSPT RTVP	<i>Mus musculus</i>	MD	AID	ELISA	0.01	0.012	0.003	0.017	0.01
78252	PVQNLT	<i>Mus musculus</i>	<i>Porphyromonas gingivalis</i>	AID	WI	78253	QGGCGR GWAFSA TGAEA	<i>Mus musculus</i>	<i>Glycine max</i>	AID	ELISA	0.02	0.02	0.017	0.02	0.018
6068	CCPDKNKS	<i>Mus musculus</i>	<i>Human herpesvirus</i>	AID	WI	6074	CCRHKQ KDVGDYK QTLPPS	<i>Ovis aries</i>	MD	AID	ELISA	0.01	0.006	0.004	0.01	0.008

TABLE 3: Continued.

IDE	Sequence	New experiment			Experiment of reference			Input perturbation terms								
		ho	so	ip	tq	IDE	Sequence	ho	so	ip	tq	$\Delta\chi$	$\Delta\Delta\chi_{ho}$	$\Delta\Delta\chi_{so}$	$\Delta\Delta\chi_{ip}$	$\Delta\Delta\chi_{tq}$
53109	RAGVCY	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OOA	ELISA	53116	RALTA FSPAQDI WGTS	<i>Oryctolagus cuniculus</i>	SARS coronavirus	AID	ELISA	-0.02	-0.016	-0.037	-0.023	-0.02
147041	IPEQ	<i>Homo sapiens</i>	<i>Triticum aestivum</i>	OOA	WI	147064	KHQGA QYVWN RTA	<i>Homo sapiens</i>	<i>Bos taurus</i>	OOID	WI	0.06	0.06	0.047	0.057	0.06
70664	VQVVYDYQ	<i>Oryctolagus cuniculus</i>	<i>Treponema pallidum</i>	AID	ELISA	70667	VQWMN RLIAFAF AGNHVSP	<i>Homo sapiens</i>	<i>Hepatitis C virus</i>	OOID	ELISA	0.05	0.046	0.041	0.049	0.05
134343	DLYIK	<i>Mus musculus</i>	<i>Human papillomavirus</i>	AID	PhDIP	134344	DMAQV TVGPGLL GVSTL	<i>Mus musculus</i>	<i>Homo sapiens</i>	AID	PhDIP	0	0	-0.009	0	0

The first input term is the value $\lambda(\varepsilon_{ij})_{\text{ref}}$ is the scoring function λ of the efficiency of the initial process ε_{ij} (known solution). The function $\lambda(\varepsilon_{ij})_{\text{ref}} = 1$ if the i th peptide could experimentally be demonstrated to be a B-epitope in the assay of reference (reference) carried out in the conditions c_j , $\lambda(\varepsilon_{ij})_{\text{ref}} = 0$ otherwise. The variational-perturbation terms $\Delta\Delta\chi_{cj}$ are at the same time terms typical of perturbation theory and moving average (MA) functions used in Box-Jenkin models in time series [60]. These new types of terms account both for the deviation of the electronegativity of all amino acids in the sequence of the new peptide with respect to the peptide of reference and with respect to all boundary conditions. In Table 1, we give the overall classification results obtained with this model. Speck-Planche et al. [61–63] introduced different multitarget/multiplexing QSAR models that incorporate this type of information based on MAs. The results obtained with the present model are excellent compared with other similar models in the literature useful for other problems including moving average models [64, 65] or perturbation models [58]. Notably, this is also the first model combining both perturbation theory and MAs in a QSPR context.

The other input terms are the following. The first $\Delta\chi_{\text{seq}} = \chi(m_q)_{\text{ref}} - \chi(m_i)_{\text{new}}$ is the perturbation term for the variation or in the mean value of electronegativity for all amino acids in the sequence of the peptide of reference. The remnant input variables of the model $\Delta\Delta\chi_{cj} = \Delta\chi_{cj-\text{ref}} - \Delta\chi_{cj-\text{new}} = [\chi(m_q)_{\text{ref}} - {}^*\chi(c_{qr})_{\text{ref}}] - [\chi(m_i)_{\text{new}} - {}^*\chi(c_{ij})_{\text{new}}]$ quantify values of the conditions of the new assay c_j -new that represent perturbations with respect to the initial conditions c_j -ref of the assay of reference. The quantities ${}^*\chi(c_{ij})$ and ${}^*\chi(c_{qr})$ are the average values of the mean electronegativity values $\chi(m_i)$ and $\chi(m_q)$ for all new and reference peptides in IEDB that are epitopes under the j th or r th boundary condition. The values of these terms have been tabulated for >500 source organisms, >50 host organisms, >10 biological process, and >30 experimental techniques. We must substitute the values of $\chi(m_i)$ and $\chi(m_q)$ of the new and reference peptides and the tabulated values of ${}^*\chi(c_{ij})$ and ${}^*\chi(c_{qr})$ for all combinations of boundary conditions to predict the perturbations of the action as epitope of peptides. In doing so we can found the optimal sequence and boundary conditions towards the use of the peptide in the development of a vaccine. In Table 2 we give some of these values of ${}^*\chi(c_{ij})$ and ${}^*\chi(c_{qr})$.

In Table 3 we depict the sequences and input-output boundary conditions for top perturbations present in IEDB. All these perturbations have observed value of $\lambda(\varepsilon_{ij})_{\text{new}} = 1$ and predicted value also equal to 1 with a high probability. See Supplementary Material available online at <http://dx.doi.org/10.1155/2014/768515> file contains a full list of >200,000 cases of perturbations.

4. Conclusions

It is possible to develop general models for vaccine design able to predict the results of multiple input-output perturbations

in peptide sequence and experimental assay boundary conditions using ideas of QSPR analysis, perturbation theory, and Box and Jenkins MA operators. The electronegativity values calculated with MARCH-INSIDE seem to be good molecular descriptors for this type of QSPR-perturbation models.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The present study was partially supported by Grants AGL2010-22290-C02 and AGL2011-30563-C03 from Ministerio de Ciencia e Innovación, Spain, and Grant CN 2012/155 from Xunta de Galicia, Spain.

References

- [1] B. Peters, J. Sidney, P. Bourne et al., “The design and implementation of the immune epitope database and analysis resource,” *Immunogenetics*, vol. 57, no. 5, pp. 326–336, 2005.
- [2] B. Peters, J. Sidney, P. Bourne et al., “The immune epitope database and analysis resource: from vision to blueprint,” *PLoS Biology*, vol. 3, no. 3, p. e91, 2005.
- [3] P. Wang, A. A. Morgan, Q. Zhang, A. Sette, and B. Peters, “Automating document classification for the Immune Epitope Database,” *BMC Bioinformatics*, vol. 8, article 269, 2007.
- [4] A. Sette, “The immune epitope database and analysis resource: from vision to blueprint,” *Genome Informatics*, vol. 15, no. 2, p. 299, 2004.
- [5] N. Salimi, W. Fleri, B. Peters, and A. Sette, “The immune epitope database: a historical retrospective of the first decade,” *Immunology*, vol. 137, no. 2, pp. 117–123, 2012.
- [6] Y. Kim, A. Sette, and B. Peters, “Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource,” *Journal of Immunological Methods*, vol. 374, no. 1-2, pp. 62–69, 2011.
- [7] Y. Kim, J. Ponomarenko, Z. Zhu, D. Tamang et al., “Immune epitope database analysis resource,” *Nucleic Acids Research*, pp. W525–W530, 2012.
- [8] A. M. Helguera, R. D. Combes, M. P. Gonzalez, and M. N. D. S. Cordeiro, “Applications of 2D descriptors in drug design: a DRAGON tale,” *Current Topics in Medicinal Chemistry*, vol. 8, no. 18, pp. 1628–1655, 2008.
- [9] G. M. Casañola-Martín, Y. Marrero-Ponce, M. T. H. Khan et al., “Dragon method for finding novel tyrosinase inhibitors: *Biosilico* identification and experimental in vitro assays,” *The European Journal of Medicinal Chemistry*, vol. 42, no. 11-12, pp. 1370–1381, 2007.
- [10] I. V. Tetko, J. Gasteiger, R. Todeschini et al., “Virtual computational chemistry laboratory—design and description,” *Journal of Computer-Aided Molecular Design*, vol. 19, no. 6, pp. 453–463, 2005.
- [11] A. R. Katritzky, A. Oliferenko, A. Lomaka, and M. Karelson, “Six-membered cyclic ureas as HIV-1 protease inhibitors: a QSAR study based on CODESSA PRO approach,” *Bioorganic and Medicinal Chemistry Letters*, vol. 12, no. 23, pp. 3453–3457, 2002.

- [12] A. R. Katritzky, S. Perumal, R. Petrukhin, and E. Kleinpeter, "CODESSA-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities," *Journal of Chemical Information and Computer Sciences*, vol. 41, no. 3, pp. 569–574, 2001.
- [13] S. Vilar, G. Cozza, and S. Moro, "Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery," *Current Topics in Medicinal Chemistry*, vol. 8, no. 18, pp. 1555–1572, 2008.
- [14] G. Marzaro, A. Chilin, A. Guiotto et al., "Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors," *The European Journal of Medicinal Chemistry*, vol. 46, no. 6, pp. 2185–2192, 2011.
- [15] S. Vilar, E. Estrada, E. Uriarte, L. Santana, and Y. Gutierrez, "In silico studies toward the discovery of new anti-HIV nucleoside compounds through the use of tops-mode and 2D/3D connectivity indices. 2. Purine derivatives," *Journal of Chemical Information and Modeling*, vol. 45, no. 2, pp. 502–514, 2005.
- [16] E. Estrada, J. A. Quincoces, and G. Patlewicz, "Creating molecular diversity from antioxidants in Brazilian propolis. combination of TOPS-MODE QSAR and virtual structure generation," *Molecular Diversity*, vol. 8, no. 1, pp. 21–33, 2004.
- [17] E. Estrada and H. González, "What are the limits of applicability for graph theoretic descriptors in QSPR/QSAR? modeling dipole moments of aromatic compounds with TOPS-MODE descriptors," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 1, pp. 75–84, 2003.
- [18] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal et al., "Tomocomd-Caridd, a novel approach for computer-aided "rational" drug design: I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds," *Journal of Computer-Aided Molecular Design*, vol. 18, no. 10, pp. 615–634, 2004.
- [19] Y. Marrero Ponce, R. Medina Marrero, E. A. Castro et al., "Protein quadratic indices of the "macromolecular pseudograph's α -carbon atom adjacency matrix". 1. prediction of arc repressor alanine-mutant's stability," *Molecules*, vol. 9, no. 12, pp. 1124–1147, 2004.
- [20] H. González-Díaz, F. Prado-Prado, and F. M. Ubeira, "Predicting antimicrobial drugs and targets with the MARCH-INSIDE approach," *Current Topics in Medicinal Chemistry*, vol. 8, no. 18, pp. 1676–1690, 2008.
- [21] R. García-Domenech, J. Gálvez, J. V. de Julián-Ortiz, and L. Pogliani, "Some new trends in chemical graph theory," *Chemical Reviews*, vol. 108, no. 3, pp. 1127–1169, 2008.
- [22] E. Estrada and E. Uriarte, "Recent advances on the role of topological indices in drug discovery research," *Current Medicinal Chemistry*, vol. 8, no. 13, pp. 1573–1588, 2001.
- [23] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, 2008.
- [24] E. Estrada, E. J. Delgado, J. B. Alderete, and G. A. Jaña, "Quantum-connectivity descriptors in modeling solubility of environmentally important organic compounds," *Journal of Computational Chemistry*, vol. 25, no. 14, pp. 1787–1796, 2004.
- [25] E. Besalu, X. Girones, L. Amat, and R. Carbo-Dorca, "Molecular quantum similarity and the fundamentals of QSAR," *Accounts of Chemical Research*, vol. 35, no. 5, pp. 289–295, 2002.
- [26] D. A. Rincón, M. N. D. S. Cordeiro, and R. A. Mosquera, "On the electronic structure of cocaine and its metabolites," *Journal of Physical Chemistry A*, vol. 113, no. 50, pp. 13937–13942, 2009.
- [27] M. Mandado, M. J. González-Moa, and R. A. Mosquera, "Chemical graph theory and n-center electron delocalization indices: a study on polycyclic aromatic hydrocarbons," *Journal of Computational Chemistry*, vol. 28, no. 10, pp. 1625–1633, 2007.
- [28] T. Hill and P. Lewicki, *STATISTICS: Methods and Applications: A Comprehensive Reference for Science, Industry and Data Mining*, StatSoft, Tulsa, Okla, USA, 2006.
- [29] E. Frank, M. Hall, L. Trigg, G. Holmes, and I. H. Witten, "Data mining in bioinformatics using Weka," *Bioinformatics*, vol. 20, no. 15, pp. 2479–2481, 2004.
- [30] G. Patlewicz, N. Ball, E. D. Booth, E. Hulzebos, E. Zvinavashe, and C. Hennes, "Use of category approaches, read-across and (Q)SAR: general considerations," *Regulatory Toxicology and Pharmacology*, vol. 67, no. 1, pp. 1–12, 2013.
- [31] D. W. Roberts and G. Y. Patlewicz, "Updating the skin sensitization in vitro data assessment paradigm in 2009—a chemistry and QSAR perspective," *Journal of Applied Toxicology*, vol. 30, no. 3, pp. 286–288, 2010.
- [32] D. W. Roberts and G. Y. Patlewicz, "Nonanimal alternatives for skin sensitization: letter to the editor," *Toxicological Sciences*, vol. 106, no. 2, pp. 572–574, 2008.
- [33] E. Estrada, G. Patlewicz, and Y. Gutierrez, "From knowledge generation to knowledge archive. a general strategy using TOPS-MODE with DEREK to formulate new alerts for skin sensitization," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 2, pp. 688–698, 2004.
- [34] G. F. Gerberick, C. A. Ryan, P. S. Kern et al., "A chemical dataset for evaluation of alternative approaches to skin-sensitization testing," *Contact Dermatitis*, vol. 50, no. 5, pp. 274–288, 2004.
- [35] G. Y. Patlewicz, D. A. Basketter, C. K. Smith Pease et al., "Further evaluation of quantitative structure–activity relationship models for the prediction of the skin sensitization potency of selected fragrance allergens," *Contact Dermatitis*, vol. 50, no. 2, pp. 91–97, 2004.
- [36] E. Estrada, G. Patlewicz, M. Chamberlain, D. Basketter, and S. Larbey, "Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach," *Chemical Research in Toxicology*, vol. 16, no. 10, pp. 1226–1235, 2003.
- [37] G. Patlewicz, R. Rodford, and J. D. Walker, "Quantitative structure–activity relationships for predicting skin and eye irritation," *Environmental Toxicology and Chemistry*, vol. 22, no. 8, pp. 1862–1869, 2003.
- [38] G. Y. Patlewicz, R. A. Rodford, G. Ellis, and M. D. Barratt, "A QSAR model for the eye irritation of cationic surfactants," *Toxicology In Vitro*, vol. 14, no. 1, pp. 79–84, 2000.
- [39] E. Tenorio-Borroto, C. G. Penuelas Rivas, J. C. Vasquez Chagoyan, N. Castanedo, F. J. Prado-Prado, and X. Garcia-Mera, "ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen," *Bioorganic & Medicinal Chemistry*, vol. 20, no. 20, pp. 6181–6194, 2012.
- [40] D. R. Flower, H. McSparron, M. J. Blythe et al., "Computational vaccinology: quantitative approaches," *Novartis Foundation Symposium*, vol. 254, pp. 102–120, 2003.
- [41] I. A. Doytchinova, P. Guan, and D. R. Flower, "Quantitative structure-activity relationships and the prediction of MHC supermotifs," *Methods*, vol. 34, no. 4, pp. 444–453, 2004.
- [42] Y. Xiao and M. R. Segal, "Prediction of genomewide conserved epitope profiles of HIV-1: classifier choice and peptide representation," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, article 25, 2005.

- [43] T. Fagerberg, V. Zoete, S. Viatte, P. Baumgaertner, P. M. Alves, P. Romero et al., "Prediction of cross-recognition of peptide-HLA A2 by melan-a-specific cytotoxic T lymphocytes using three-dimensional quantitative structure-activity relationships," *PLoS ONE*, vol. 8, no. 7, Article ID e65590, 2013.
- [44] D. Barh, A. N. Misra, A. Kumar, and A. Vasco, "A novel strategy of epitope design in *Neisseria gonorrhoeae*," *Bioinformation*, vol. 5, no. 2, pp. 77–82, 2010.
- [45] J. Bi, R. Song, H. Yang et al., "Stepwise identification of HLA-A*0201-restricted CD8⁺ T-cell epitope peptides from herpes simplex virus type 1 genome boosted by a steprank scheme," *Biopolymers*, vol. 96, no. 3, pp. 328–339, 2011.
- [46] R. D. Bremel and E. J. Homan, "An integrated approach to epitope analysis II: a system for proteomic-scale prediction of immunological characteristics," *Immunome Research*, vol. 6, no. 1, article 8, 2010.
- [47] C. M. Diez-Rivero, B. Chenlo, P. Zuluaga, and P. A. Reche, "Quantitative modeling of peptide binding to TAP using support vector machine," *Proteins*, vol. 78, no. 1, pp. 63–72, 2010.
- [48] E. Martínez-Naves, E. M. Lafuente, and P. A. Reche, "Recognition of the ligand-type specificity of classical and non-classical MHC I proteins," *FEBS Letters*, vol. 585, no. 21, pp. 3478–3484, 2011.
- [49] M. Bhasin, E. L. Reinherz, and P. A. Reche, "Recognition and classification of histones using support vector machine," *Journal of Computational Biology*, vol. 13, no. 1, pp. 102–112, 2006.
- [50] P. A. Reche and E. L. Reinherz, "PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands," *Nucleic Acids Research*, vol. 33, no. 2, pp. W138–W142, 2005.
- [51] E. M. Lafuente and P. A. Reche, "Prediction of MHC-peptide binding: a systematic and comprehensive overview," *Current Pharmaceutical Design*, vol. 15, no. 28, pp. 3209–3220, 2009.
- [52] A. Gaulton, L. J. Bellis, A. P. Bento et al. et al., "A large-scale bioactivity database for drug discovery," *Nucleic Acids Research*, pp. D1100–D1107, 2012.
- [53] W. H. Cropper, *Great Physicists: The Life and Times of Leading Physicists from Galileo to Hawking*, Oxford University Press, New York, NY, USA, 2004.
- [54] E. L. Bouzarth, A. Brooks, R. Camassa et al., "Epicyclic orbits in a viscous fluid about a precessing rod: theory and experiments at the micro-and macro-scales," *Physical Review E*, vol. 76, no. 1, part 2, Article ID 016313, 2007.
- [55] M. Laberge, "Intrinsic protein electric fields: basic non-covalent interactions and relationship to protein-induced Stark effects," *Biochimica et Biophysica Acta*, vol. 1386, no. 2, pp. 305–330, 1998.
- [56] F. M. Fernandez and J. A. Morales, "Perturbation theory without wave functions for the Zeeman effect in hydrogen," *Physical Review A*, vol. 46, no. 1, pp. 318–326, 1992.
- [57] B. D. Marshall and W. G. Chapman, "Thermodynamic perturbation theory for associating fluids with small bond angles: effects of steric hindrance, ring formation, and double bonding," *Physical Review E*, vol. 87, no. 5, Article ID 052307, 12 pages, 2013.
- [58] H. Gonzalez-Diaz, S. Arrasate, A. Gomez-San Juan et al., "New theory for multiple input-output perturbations in complex molecular systems. I. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry," *Current Topics in Medicinal Chemistry*, 2013.
- [59] H. Gonzalez-Diaz, S. Arrasate, N. Sotomayor et al., "MIANN models in medicinal, physical and organic chemistry," *Current Topics in Medicinal Chemistry*, vol. 13, no. 5, pp. 619–641, 2013.
- [60] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1970.
- [61] A. Speck-Planche, V. V. Kleandrova, F. Luan, and M. N. Cordeiro, "Unified multi-target approach for the rational in silico design of anti-bladder cancer agents," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 13, no. 5, pp. 791–800, 2013.
- [62] A. Speck-Planche, V. V. Kleandrova, and M. N. Cordeiro, "New insights toward the discovery of antibacterial agents: multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs," *The European Journal of Pharmaceutical Sciences*, vol. 48, no. 4-5, pp. 812–818, 2013.
- [63] A. Speck-Planche, V. V. Kleandrova, F. Luan, and M. N. Cordeiro, "Multi-target inhibitors for proteins associated with Alzheimer: in silico discovery using fragment-based descriptors," *Current Alzheimer Research*, vol. 10, no. 2, pp. 117–124, 2013.
- [64] A. Speck-Planche, V. V. Kleandrova, F. Luan, and M. N. Cordeiro, "In silico discovery and virtual screening of multi-target inhibitors for proteins in Mycobacterium tuberculosis," *Combinatorial Chemistry & High Throughput Screening*, vol. 15, no. 8, pp. 666–673, 2012.
- [65] A. Speck-Planche, V. V. Kleandrova, F. Luan, and M. N. D. S. Cordeiro, "Chemoinformatics in anti-cancer chemotherapy: multi-target QSAR model for the in silico discovery of anti-breast cancer agents," *The European Journal of Pharmaceutical Sciences*, vol. 47, no. 1, pp. 273–279, 2012.