J Comput Aided Mol Des (2008) 22:91–104
DOI 10.1007/s10822-008-9172-0

A comprehensive analysis of the thermodynamic events involved in ligand–receptor binding using CoRIA and its variants

Jitender Verma · Vijay M. Khedkar · Arati S. Prabhu · Santosh A. Khedkar · Alpeshkumar K. Malde · Evans C. Coutinho

Received: 11 September 2007 / Accepted: 5 January 2008 / Published online: 25 January 2008

Abstract Quantitative Structure-Activity Relationships (QSAR) are being used since decades for prediction of biological activity, lead optimization, classification, identification and explanation of the mechanisms of drug action, and prediction of novel structural leads in drug discovery. Though the technique has lived up to its expectations in many aspects, much work still needs to be done in relation to problems related to the rational design of peptides. Peptides are the drugs of choice in many situations, however, designing them rationally is a complicated task and the complexity increases with the length of their sequence. In order to deal with the problem of peptide optimization, one of our recently developed QSAR formalisms CoRIA (Comparative Residue Interaction Analysis) is being expanded and modified as: reverse-CoRIA (*r*CoRIA) and mixed-CoRIA (*m*CoRIA) approaches. In these methodologies, the peptide is fragmented into individual units and the interaction energies (van der Waals, Coulombic and hydrophobic) of each amino acid in the peptide with the receptor as a whole (*r*CoRIA) and with individual active site residues in the receptor (*m*CoRIA) are calculated, which along with other thermodynamic descriptors, are used as independent variables that are correlated to the biological activity by chemometric methods. As a test case, the three CoRIA methodologies have been validated on a dataset of diverse nonamer peptides that bind to the Class I major histocompatibility complex molecule HLA-A*0201, and for which some structure activity relationships have already been reported. The different models developed, and validated both internally as well as externally, were found to be robust with statistically significant values of r^2 (correlation coefficient) and r^2_{pred} (predictive r^2). These models were able to identify all the structure activity relationships known for this class of peptides, as well uncover some new relationships. This means that these methodologies will perform well for other peptide datasets too. The major advantage of these approaches is that they explicitly utilize the 3D structures of small molecules or peptides as well as their macromolecular targets, to extract position-specific information about important interactions between the ligand and receptor, which can assist the medicinal and computational chemists in designing new molecules, and biologists in studying the influence of mutations in the target receptor on ligand binding.

Keywords CoRIA · *r*CoRIA · *m*CoRIA · G/PLS · 3D-QSAR

J. Verma · V. M. Khedkar · A. S. Prabhu · S. A. Khedkar · A. K. Malde · E. C. Coutinho (✉)
Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (East), Mumbai 400 098, India
e-mail: evans@bcpindia.org; evans@bcp.edu.in

Introduction

Quantitative Structure-Activity Relationships builds atomistic or virtual models to establish a correlation between structural features of potential drug candidates and their binding affinity/biological activity/toxicity towards a known or hypothetical macromolecular target. Since its establishment by Hansch [1–4], the technique has undergone over the years significant modifications in almost all of its aspects. The various QSAR approaches are often categorized according to their dimensionality as 2D, 3D and so on, which refers to the structural representation or the way by which the descriptor values are derived. Because ligand–receptor interactions are inherently 3D properties, there has been much effort in developing QSAR methods that explicitly take into consideration the 3D geometries of molecules. Of

particular interest for the biomedical sciences are the 3D-QSAR techniques, a majority of which are based on the calculation of ligand–receptor interactions (usually van der Waals and Coulombic interaction energies) *indirectly* using probes positioned at intersections of a lattice (grid or box) straddling a three dimensional region resembling a binding site surrogate. Comparative Molecular Field Analysis (CoMFA) [5], Molecular Shape Analysis (MSA) [6], Molecular Similarity Matrices (for e.g., CoMSIA) [7], Distance Geometry [8], the Hypothetical Active Site Lattice method (HASL) [9], Genetically Evolved Receptor Models (GERM) [10], CoMPASS [11] etc. are the 3D-QSAR methods developed on this concept and are exclusively based on ligand information without taking into account the 3D structure of the macromolecular target. Hopfinger and workers [12] for the first time incorporated information on the receptor in a QSAR analysis to devise the 4D-QSAR methodology. Similarly, Vedani et al. have developed methods beyond the third dimension by accounting for the effect of different conformations as the fourth dimension [13], the induced-fit mechanism as the fifth dimension [14], and assessment of different solvation models as the sixth dimension [15], additionally incorporating contributions from the solvent and entropy factors into the analysis.

Significant advances have been made in recent years in realizing the rational computations of ligand–receptor binding thermodynamics [16]. Several endeavors in QSARs have attempted to use the wealth of valuable information contained in the ligand–receptor complexes, in the last few years. The imperative landmarks in receptor-ligand based QSAR methods are COMBINE (*Co*mparative *Bin*ding *E*nergy) [17], AFMoC (*A*daptation of *F*ields for *Mo*lecular *C*omparison; a reverse variant of CoMFA) [18], and CoRIA (*Co*mparative *R*esidue *I*nteraction *A*nalysis) [19]. Our newly developed CoRIA methodology, based on the descriptors that completely describe the thermodynamic events involved in ligand binding, is able to explore both the qualitative as well as the quantitative aspects of the ligand–receptor recognition process. The concept has already been validated on small organic molecules [19, 20]. In this paper we describe an extension of the CoRIA approach, meant to deal with the problems of peptide QSAR. In the new methodology, the ligand (peptide) is fragmented into individual units i.e. at the level of amino acid residues. The non-bonded (van der Waals and Coulombic) and hydrophobic interaction energies of each residue in the peptide with the receptor as a whole (termed as reverse-CoRIA or *r*CoRIA approach) and with individual residues in the active site of the receptor (called the mixed-CoRIA or *m*CoRIA method) are calculated and used as independent variables along with other thermodynamic descriptors, in statistical analysis. The advantage of this formalism is that it makes explicit use of the structures of ligand–receptor complexes to provide deeper insights into important interactions at the level of both the receptor and the ligand, which can directly be utilized in the design of new molecules and receptors. The methodology can be employed to forecast modifications in both the ligand as well as the receptor, provided structures of some ligand–receptor complexes are available.

Peptides and proteins are the essential elements in all living systems. Peptides are preferred as drugs of choice due to their high potency (low dose), specificity and selectivity (reduced side effects). Rational de novo design of a peptide is still a difficult task and their optimization is an awkward process, since the complications increase with the length of the peptide sequence. The experimental methods of optimizing a peptide include a systematic scan of the peptide by incorporation of a particular amino acid at a single position one at a time, and then comparing the effect against the wild type [21]. Several in silico approaches including 3D-QSAR and simulation methods have also been used to complement the experimental techniques in peptide optimization [22, 23]. The QSAR methods are singularly useful in solving problems related to the design of peptide ligands. Furthermore, peptides are ideal candidates for the *r*CoRIA and *m*CoRIA approaches, since it is relatively easier to fragment a peptide rather than a small organic molecule.

The ideal method of optimizing a peptide lead structure would be to examine the contribution of every possible amino acid type at every possible position in the peptide, towards the overall activity of the molecule. It is comparatively a straightforward exercise to optimize the activity of peptides through a description of the nature and location of every amino acid in the peptide sequence in the QSAR formalism, since fragmenting the peptide into individual residues is an intrinsic property of peptides. Various such attempts have been made to design more potent peptides using *descriptor*-based QSAR approaches. Sneath correlated the chemical structure and biological activity of peptides using the qualitative (interval) data of amino acids as descriptors [24]. Kidera et al. described the natural amino acids through 10 orthogonal vectors derived from Principal Component Analysis (PCA) of 188 reported properties [25]. Later, in a similar kind of work, Hellberg et al. generated the principal properties, the so called *z-scores*, by performing PCA on various descriptors of each of the 20 natural amino acids and then applied them to study the effect of variation in amino acid sequence on a set of ACE dipeptide inhibitors [26]. Collantes et al. studied the application of *isotropic surface area (ISA)* and *electronic charge index (ECI)* of the side-chains of amino acids as descriptors in a QSAR study of three peptide sets—ACE dipeptide inhibitors, bradykinin potentiating peptides and the bitter tasting dipeptides [27]. Often researchers have also used various descriptors such as

t-scores [27], *MS WHIM scores* [28] etc in peptide QSAR. Recently, we reported a *descriptor-based* QSAR approach for the optimization of peptides, assuming that each amino acid residue makes an autonomous contribution to the overall activity and that the total activity is the sum of the constituent units [29]. The location and nature of every amino acid residue in the peptide sequence was encoded in the QSAR formalism using the ideology of the Hansch (*descriptor* QSAR) and the Free–Wilson (*binary* QSAR) methodologies to deduce the most favorable sequence of amino acids in the nonamer peptides that bind to the Class I MHC (Major Histocompatibility Complex) molecule HLA-A*0201. All the above mentioned techniques of peptide optimization are limited by the scope of ligand-based QSAR methods. CoRIA and its variants which are founded on the thermodynamics of ligand–receptor interactions are better optimization tools since their implementation is not restricted only to content derived from the ligand but also incorporate receptor-rich information.

Methodology

Biological Data

The peptides that form stable complexes with Class I MHC proteins such as HLA-A*0201 help in the activation of T-cells which in turn allows the T-cell-mediated immune system to distinguish body cells from invading antigens. The prediction of peptide binding affinity to MHC molecules is an important obligation for epitope prediction and enables the identification of highly immunogenic proteins which may function as valuable putative vaccines. Several QSAR studies have investigated the binding of antigenic peptides with MHC Class I molecules, and various key interactions are now well understood [30–38]. This study uses this peptide dataset therefore to demonstrate the potential of the CoRIA approaches in substantiating what has already been recognized.

The dataset used as a test bed in this study to validate the three CoRIA formalisms (CoRIA, *r*CoRIA and *m*CoRIA) includes eighty nonapeptides with affinity for the HLA-A*0201 molecule, taken from the dataset compiled by Doytchinova and Flower [22]. This dataset was simply chosen because it has all the qualities necessary for the successful development of a good QSAR model, like a large number of molecules with good structural diversity and a modest span of activity values, besides high quality biological data. The binding affinities reported as IC_{50} values are based on a quantitative assay which determines the inhibition of binding of a radiolabeled standard peptide (FLPSDYFPSV) to detergent-solubilized MHC molecules [39, 40]. The IC_{50} values were converted to the negative logarithmic values (pIC_{50}), which cover more than 3 log orders. Table 1 lists the sequences and experimental pIC_{50} values of the peptides used in this study. The peptides were divided into a training set consisting of 55 molecules and a

Table 1 Nonapeptides used in the QSAR study along with their experimental pIC_{50} values

Mol. ID	Peptide	pIC_{50}	Mol. ID	Peptide	pIC_{50}
Training set (N = 55)					
T1	GTLVALVGL	5.34	T29	ILHNGAYSL	7.13
T2	GIGILTVIL	6.00	T30	HLYSHPIIL	7.13
T3	AIAKAAAAV	6.18	T31	VVMGTLVAL	7.17
T4	ALAKAAAAI	6.21	T32	GLSRYVARL	7.25
T5	LLSSNLSWL	6.34	T33	YMLDLQPET	7.31
T6	GLACHQLCA	6.38	T34	YLEPGPVTV	7.34
T7	LIGNESFAL	6.42	T35	YLSPGPVTA	7.38
T8	ALAKAAAAV	6.42	T36	YMNGTMSQV	7.40
T9	MLLAVLYCL	6.48	T37	SVYDFFVWL	7.44
T10	KLPQLCTEL	6.48	T38	ITWQVPFSV	7.46
T11	ALAKAAAAL	6.51	T39	ITYQVPFSV	7.48
T12	AAGIGILTV	6.58	T40	YLSPGPVTV	7.64
T13	FLGGTPVCL	6.62	T41	VLIQRNPQL	7.64
T14	ILDEAYVMA	6.62	T42	SLYADSPSV	7.66
T15	NLSWLSLDV	6.64	T43	RLLQETELV	7.68
T16	VLQAGFFLL	6.68	T44	ILSQVPFSV	7.70
T17	VILGVLLLI	6.79	T45	QLFEDNYAL	7.76
T18	VTWHRYHLL	6.79	T46	YAIDLPVSV	7.80
T19	TLGIVCPIC	6.82	T47	FVWLHYYSV	7.82
T20	HLYQGCQVV	6.83	T48	YLMPGPVTV	7.93
T21	FAFRDLCIV	6.89	T49	WLDQVPFSV	7.94
T22	FLEPGPVTA	6.90	T50	YLYPGPVTV	8.05
T23	ITDQVPFSV	6.95	T51	YLFPGPVTV	8.24
T24	YVITTQHWL	6.98	T52	ILYQVPFSV	8.31
T25	LLCLIFLLV	7.00	T53	YLFPGPVTA	8.50
T26	HLAVIGALL	7.00	T54	YLWPGPVTA	8.50
T27	YLEPGPVTL	7.06	T55	ILWQVPFSV	8.77
T28	YTDQVPFSV	7.07			
Test set (N = 25)					
S1	LLSCLGCKI	5.45	S14	VLLDYQGML	7.33
S2	DPKVKQWPL	6.18	S15	ILSPFMPLL	7.35
S3	GLGQVPLIV	6.30	S16	IIDQVPFSV	7.40
S4	LLAVGATKV	6.48	S17	KIFGSLAFL	7.48
S5	IISCTCPTV	6.58	S18	LLLCLIFLL	7.59
S6	ALIHHNTHL	6.62	S19	ALMDKSLHV	7.77
S7	YMIMVKCWM	6.66	S20	YLYPGPVTA	7.77
S8	TLDSQVMSL	6.79	S21	LLFGYPVYV	7.89
S9	CLTSTVQLV	6.83	S22	ILKEPVHGV	7.92
S10	FLCKQYLNL	6.88	S23	FLLSLGIHL	8.05
S11	ALCRWGLLL	7.00	S24	YLWPGPVTV	8.13
S12	SIISAVVGI	7.16	S25	ILFQVPFSV	8.70
S13	YLEPGPVTI	7.19			

test set of 25 molecules based on the Tanimoto coefficient using the '*select diverse*' utility in *Cerius2* (v 4.6; Accelrys Inc., USA) [41].

Molecular modeling

Almost all the molecular modeling calculations were carried out with *InsightII* (v 2005L, Accelrys Inc., USA) [42] running on a Pentium IV computer with the Linux Red Hat Enterprise 2.1 OS. Among the various X-ray crystal structures of HLA-A*0201 in the Protein Data Bank [43], the highest resolution complex of HLA Class I histocompatibility antigen with beta-2-microglobulin and the alpha and beta chains of T-cell receptor bound to the nonameric viral peptide GILGFVFTL (PDB id 1OGA) was selected for docking the various peptides by superimposition. The crystal structures of eight HLA receptor-nonapeptide complexes (PDB ids 1AKJ, 1DUZ, 1HHG, 1HHJ, 1OGA, 1QEW, 1QSE, 1QSF) were used as templates to build starting conformations of the 80 nonapeptides described in Table 1. The choice of the template was governed by the mutation matrix score of the sequence similarity between the target (peptides in Table 1) and the template sequences (the 8 PDB structures quoted above). The conformations of the peptides were then built by replacing the original amino acids in the eight template sequences with the appropriate residues. For example, the test set molecule S21 (LLFGYPVYV) in Table 1 was generated by replacing Ala at the eighth position in the template structure 1QSF (LLFGYPVAV) with Tyr. The side chain conformations were optimized using the '*Rotamer Search*' option in *InsightII* so as to minimize any possible steric clashes between them. Hydrogens were added to the molecules corresponding to a pH of 7.3, keeping in line with the conditions under which the binding assay was carried out. The geometries of the nonapeptides were optimized within the receptor complexes by subjecting them to an energy minimization protocol with the CFF91 [44] force field, using steepest descents and conjugate gradient methods, till a gradient of 0.001 kcal/mol/Å was reached.

Descriptors

The thermodynamics of ligand–receptor binding involves many events like interaction, solvation and entropy changes, all of which are taken into consideration in the CoRIA approaches and are described below.

Interaction energies

The primary input to the CoRIA methodologies comes from the specific non-bonded interactions between the ligand and receptor as a consequence of their proximity. It is an entirely enthalpic contribution and is equal to the total energy of the complex minus the energy of the free protein and free ligand. The major factors constituting the non-bonded interaction energy are van der Waals (E_{vdw}) and electrostatic (E_{ele}) interactions between the ligand and the receptor, which are functionally calculated as follows:

$$E_{vdw} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

$$E_{ele} = \frac{q_i q_j}{\varepsilon r_{ij}}$$

where A_{ij} and B_{ij} are the repulsive and attractive term coefficients between atoms i and j respectively, r_{ij} is the interatomic distance between atoms i and j, q_i and q_j are the atomic charges of interacting atoms i and j respectively, and ε is the dielectric constant. The non-bonded (van der Waals and Coulombic) interaction energies were computed using the CFF91 force field [44] in the "*Discover*" module of the program *InsightII*.

Another major parameter contributing to the thermodynamics of ligand–receptor binding is the hydrophobic interaction between the ligand and receptor. It is a complex process resulting primarily from entropic effects related to the change in the orientation of solvent molecules in the solvation shell wrapping the solute molecules, and also from the bulk form of solvent molecules. The quantified values for the hydrophobic interactions between the ligand and receptor were obtained in the form of *HINT* scores [45] through the "*HINT*" module incorporated in the *Sybyl* program (v 7.1, Tripos Inc., USA) [46]. The hydrophobicity calculation in this program is based on the fact that solubility data can be regarded as just another physical property capable of mirroring the molecular interactions between solute and solvent molecules. *HINT* calculates the hydrophobic interactions between all atom pairs in a molecule using the following equation:

$$B = \sum_i \sum_j b_{ij}$$

where, $b_{ij} = a_i a_j S_i S_j R_{ij} T_{ij}$

b_{ij} = micro-interaction constant representing the attraction/interaction between atoms i and j
a_i = the hydrophobic atom constant for atom i
S_i = the solvent accessible surface area for atom i
R_{ij} = the functional distance behavior for the interaction between atoms i and j
T_{ij} = a discriminant function designed to keep the signs of interactions consistent with the *HINT* convention that favorable interactions are positive and unfavorable interactions are negative.

All the interaction energies (van der Waals, Coulombic and hydrophobic) between the nonameric peptides and the receptor were computed separately for each of the three formalisms CoRIA, *r*CoRIA and *m*CoRIA as shown schematically in Fig. 1. The rectangular bar in the center represents the nonapeptide segmented into individual amino acids marked as P1, P2,…, P9. The nonapeptide is surrounded by the active site residues (R1, R2,…, R_n) of the receptor. The arrows indicate the interaction between the respective residues of the peptide and the receptor.

The interacting residues in the CoRIA approach include a total of 80 residues in the receptor within a radius of 10 Å from the peptide. The interaction energies of each nonameric peptide (as a whole, not partitioned) with these active site residues (R1, R2,…, R_n) of the receptor were calculated (Fig. 1a). Thus for each peptide, there were 80 entries (i.e. columns) in the QSAR table, each for the van der Waals, Coulombic and hydrophobic interactions. On the other hand in the *r*CoRIA approach, the interaction energies of each individual amino acid of the nonapeptide (P1, P2,…, P9) with the receptor as a whole were calculated (Fig. 1b), which makes a total of 9 entries each for the van der Waals, Coulombic and hydrophobic interactions. Finally for the *m*CoRIA approach, nine subsets each consisting of receptor residues lying within a 5 Å radius from the C_α atom of each of the nine residues of the peptide were constructed (Fig. 1c) and the interaction energies of each member of the subset with the respective residue in each of the 9 positions of the peptide were computed (P1 ↔ R1, R2,…, R_m, P2 ↔ R3, R4,…, R_n and so on). The total number of entries for the interaction energies (van der Waals, Coulombic and hydrophobic) in the *m*CoRIA approach is 726.

Solvation free energy

Prior to binding, both the ligand as well as the receptor are solvated, but as the interactions with water compete with protein–ligand interactions, the solvent molecules reorganize. The free energy of solvation of the ligand at physiological conditions is the hydration free energy, which is the difference between the free (e.g. cellular) and the bound state. It corresponds to the energy required to strip the solvent molecules off the ligand when changing from an aqueous environment to a hydrophobic receptor cavity. Since in many complexes, the conformation of the receptor does not change much from the uncomplexed structure, the net solvation free energy for the receptor (free and bound) is negligible as compared to that of the ligand [47]. The electrostatic contribution to the solvation free energy of the peptides was calculated using the "*Prepare*" module in the program *QUASAR* [48] which is based on the method developed and validated by Still et al. [49].

Strain energy

Another important descriptor in the CoRIA approaches is the contribution from changes in the conformation of the ligand

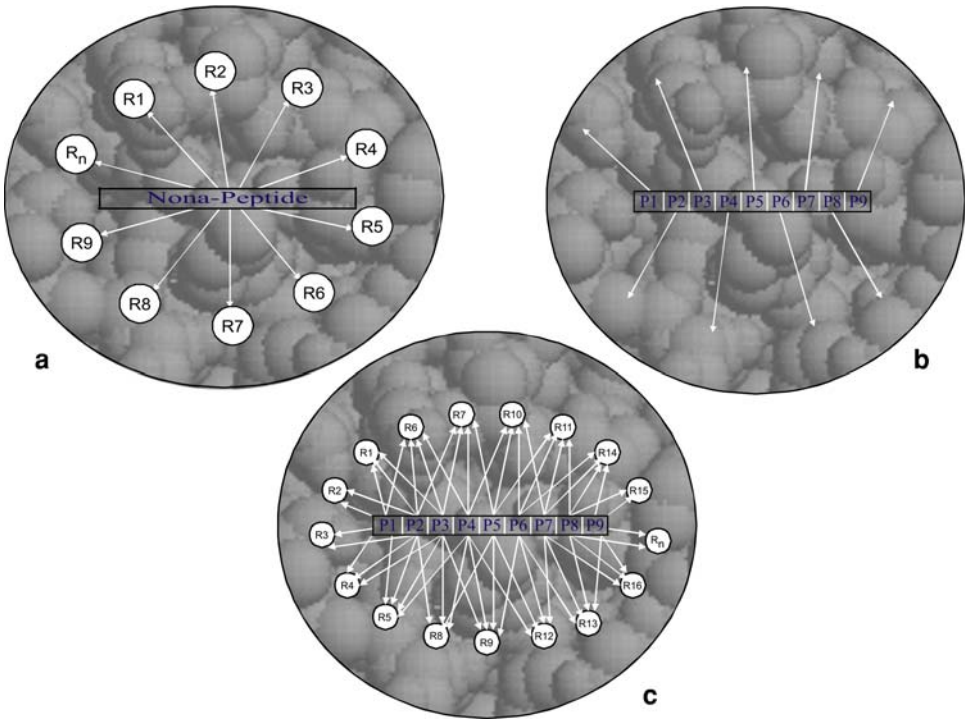

Fig. 1 Schematic representation of the strategies adopted for calculation of interaction energy fields for (**a**) CoRIA, (**b**) *r*CoRIA and (**c**) *m*CoRIA approaches

upon binding with the receptor. The conformational change in the ligand upon binding to the receptor is much more significant compared to that for the receptor and can be estimated by the strain energy upon binding. The ligand conformational energy can be calculated with a molecular mechanics potential function as the energy associated with changes in bond lengths, angles, torsions and non-bonded interactions. The peptides extracted from their complexes were minimized using a combination of 5,000 steps of steepest descents followed by 50,000 steps of conjugate gradients, to a maximum energy derivative of 0.001 kcal/mol/Å. The difference in the energy of the ligands in their docked conformations and the conformations minimized in vacuo was taken as the energy due to conformational strain.

Entropy loss

The term 'entropy loss' accounts for the loss of torsional, vibrational, rotational and translational free energies upon binding. When two molecules bind, there is a loss of three rotational and three translational degrees of freedom. The loss of entropy due to reduced conformational flexibility upon receptor binding was estimated using the "*Prepare*" module in the program *QUASAR* [48] following the philosophy of Searle and Williams [50] by assigning an amount of 0.7 kcal/mol to every freely rotatable (i.e. single) bond, excluding the terminal $-CH_3$ groups.

Solvent accessible surface area

Solvent accessible surface area (SASA) is used to define a static or dynamic solvent-accessible region as a correction factor for situations where the ligands expose a different fraction of their surface to a solvent accessible part of the binding site. In other words, the residual surface of the ligand that is still accessible to the solvent after it has bound to the receptor, is a measure of the depth of the binding in the pocket. It correlates with the tightness and more or less with the strength and number of binding interactions with the pocket. SASA was also calculated using the "*Prepare*" module in the program *QUASAR* [48].

Statistical analysis

All QSAR equations were generated with the G/PLS method as implemented in the *Cerius2* program (v 4.6) [41], which combines the best features of the Genetic Function Approximation (GFA) [51] and the Partial Least Squares (PLS) [52] methodologies. Since interaction energies are not perfectly orthogonal, pretreatment based on correlation matrix was avoided. All descriptors in the dataset were scaled according to their mean and standard deviations, where each value in a given column is subtracted from the column mean and then divided by the standard deviation of that column, such that all the scaled descriptors have a mean of zero and a standard deviation of unity. This scaling or standardization assigns equal weight to all the descriptors and puts them on the same platform for a meaningful statistical analysis. The models were developed with linear terms and the optimal number of components was selected as four for *r*CoRIA and six for CoRIA and *m*CoRIA approaches, for which the crossvalidated r^2 (i.e. q^2) was found to be the highest. To assure simple interpretation and ease of use of the equations in designing new ligands, the length of the equations was set to six terms; with a smoothness value of 1.0 (the smoothness function controls the bias in the scoring factor between equations with different number of terms). The number of generations was limited to 10,000 and population size to 500. Crossover and mutation probabilities of 50% (default settings) were used. The models developed with a training set of 55 molecules were validated internally using randomization at 99% confidence interval, Leave-One-Out (LOO), Leave-Group-Out (LGO, group of 5) and by boot-strapping procedures [53]. Externally, the models were evaluated for their predictive power on a test set of 25 molecules.

Results and discussion

The QSAR equations and statistical analysis of the best models developed for the three different approaches are reported in Tables 2 and 3 respectively. The models constructed by all three QSAR approaches were found to be statistically significant with correlation coefficients (r^2) of more than 0.8. After randomization of the activity data, the r^2 values decrease to smaller numbers, negating the possibility of a chance correlation. Cross-validation by leave-one out and leave-five-out procedures returned statistically significant q^2 values. The boot-strapping results further advocated the robustness of the models. The predictive r^2 of all the models on the test set was also found to be more than 0.6, indicating a good predictive power of the models for molecules outside the training set.

The plots of experimental vs predicted pIC_{50} values of the peptides in the training and test sets for the best CoRIA, *r*CoRIA and *m*CoRIA models are shown in Fig. 2. All 500 equations of each model were analyzed for the frequency with which a particular descriptor appears in the population of equations. The plots of the most frequently occurring descriptors for different models are shown in Fig. 3. The frequency of occurrence of different descriptors is shown on the *X*-axis as a percentage, whereas the signs of the

Table 2 Best models developed by the three CoRIA approaches

Model	Best QSAR equation
CoRIA	pIC_{50} = 7.07−0.28 (V_L160) + 0.19 (H_E58)−0.46 (V_R169) + 0.32 (H_G26) + 0.20 (V_D77)
*r*CoRIA	pIC_{50} = 7.10−0.37 (C_P8)−0.34 (V_P3) + 0.01 (C_P2)−0.29 (V_P1) + 0.08 (H_P7)
*m*CoRIA	pIC_{50} = 7.15−0.19 (C_P2_G100) + 0.22 (H_P8_A69)−0.23 (H_P2_R97)−0.35 (V_P3_L160)−0.38 (V_P1_G162)

C, V and H—Coulombic, van der Waals and Hydrophobic interactions respectively

P1, P2, P3 etc—Residue at positions 1, 2 and 3 respectively in the peptide

V_L160—van der Waals interaction of the receptor residue Leu160 with the peptide

C_P8—Coulombic interaction of the residue at position 8 in the peptide with the receptor

H_P2_R97—Hydrophobic interaction of the residue at position 2 in the peptide with the receptor residue Arg97

Table 3 Statistical analysis of the QSAR models for the three approaches

Model	r^2	r^2 (BS)	SD (BS)	PRESS	r^2 (random)	q^2 (LOO)	q^2 (LGO)	r^2_{pred}	p^2
CoRIA	0.820	0.796	0.013	6.85	0.340	0.759	0.750	0.610	0.609
*r*CoRIA	0.817	0.804	0.005	6.72	0.324	0.778	0.778	0.602	0.601
*m*CoRIA	0.844	0.835	0.003	5.33	0.305	0.789	0.737	0.619	0.618

Number of molecules in training and test sets are 55 and 25 respectively; r^2: correlation coefficient; r^2 (BS) and SD (BS): mean values of r^2 and standard deviation respectively from Boot-strap analysis; PRESS: Predictive Residual Sum of Squares; r^2 (random): mean value of r^2 after randomization at 99% confidence interval; q^2 by LOO and LGO: cross-validation correlation coefficient by Leave-One-Out and Leave-Group-Out (group of 5) respectively; r^2_{pred}: predictive correlation coefficient of test set; p^2: predictive correlation coefficient of test set as defined by Vedani et al. [13–15]

terms in the equations are shown on the *Y*-axis. Descriptors with positive coefficients in the equations are shown as positive frequency values, whereas those with negative coefficients appear with negative frequencies. A detailed analysis of the models obtained by the three CoRIA approaches is described below. However, while analyzing the results of the CoRIA methodologies one should bear in mind that more negative the value of the non-bonded (i.e. van der Waals and Coulombic) interaction energies, stronger is the interaction between the ligand and the receptor. Similarly positive values of these interaction energies imply weaker interaction between the respective groups of the ligand and the receptor. However in the case of the hydrophobicity, favorable interactions have positive values and unfavorable interactions are negative.

CoRIA analysis

In the CoRIA approach, besides other descriptors, the central elements are the interaction energies of the entire ligand (nonameric peptide) with individual residues in the active site of the receptor (Fig. 1a). An analysis of all the equations reveals that hydrophobic and van der Waals interactions are the major forces driving the binding of the peptides with the receptor. The *HINT* scores of receptor residues Gly26, Glu58 and Lys66 have positive coefficients in the QSAR equations (Fig. 3a), suggesting that enhancing the hydrophobic interaction of the peptides through these residues may significantly improve their binding with the receptor. On the other hand, the van der Waals interaction energies of the receptor residues Leu160 and Arg169 with the peptides appear as negative coefficients in the equations (Fig. 3a). This indicates that an increase in the biological activity can be gained by strengthening the van der Waals interactions of the peptides with these residues.

*r*CoRIA analysis

The *r*CoRIA formalism, in contrast to CoRIA, involves the calculation of the interaction of the individual amino acids of the nonameric peptides with the receptor (active site residues) as a single unit (Fig. 1b). An examination of the QSAR equations and the frequency plots (Fig. 3b) of the *r*CoRIA models shows that residues at positions 1, 2, 3 and 8 in the peptides have a greater influence on the biological activity than others. Residues at remaining positions in the peptides show relatively smaller contributions towards binding. The results indicate that the N-terminal residues 1, 2 and 3 and the C-terminal residue 8 in the peptides act as anchors imparting high-affinity for binding to MHC class-I molecules, which is in line with earlier studies [39, 54–57]. The Coulombic interaction of the receptor with the residue at position 2 in the peptide has positive coefficient in the pool of equations (Fig. 3b), suggesting that an overall positive value of the Coulombic interaction energy of the receptor with the amino acid at position 2 will favor

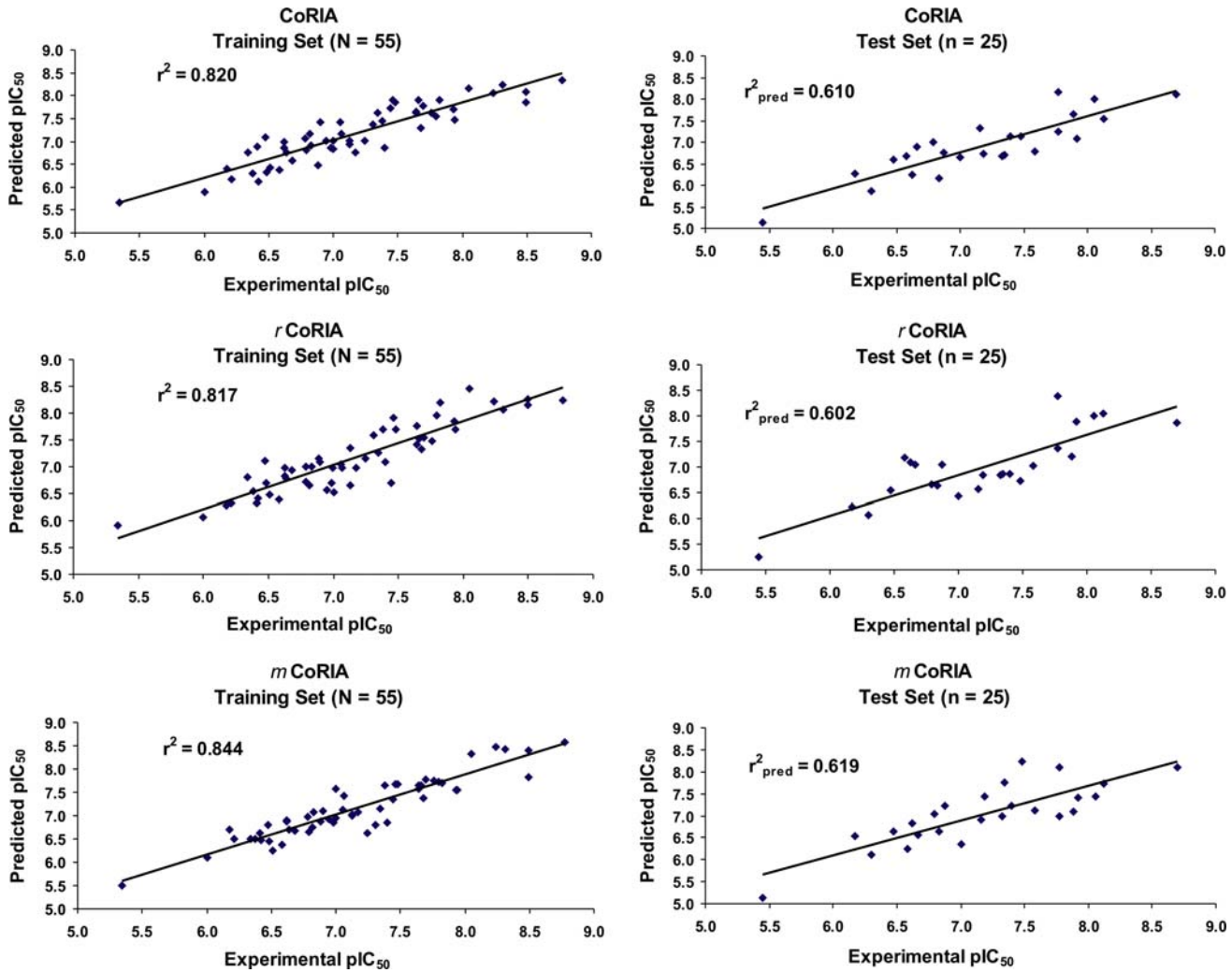

Fig. 2 Plots of experimental vs predicted pIC_{50} values of the peptides in the training and test sets for the best CoRIA, *r*CoRIA and *m*CoRIA models

binding. On the hydrophobic front, owing to the positive coefficient of the *HINT* score of the residue at position 2 in the peptide with the receptor (Fig. 3b), increasing the hydrophobic character of the amino acid at this position will favor binding. Concerning position 8 in the peptide, strengthening the Coulombic interaction while simultaneously reducing the hydrophobic interaction is predicted to amplify binding, as suggested from the negative signs of the coefficients for the Coulombic and *HINT* terms for this residue in the population of equations. The *HINT* score of the residue at position 1 in the peptide with the receptor has a positive coefficient in the equations (Fig. 3b), recommending an increase in the hydrophobic interaction of this position in the peptide with the receptor to favor binding as discussed earlier. In tandem, an increase in the van der Waals interaction of the residue at position 1 in the peptide with the receptor is suggested to improve binding, as shown by the negative coefficient of this interaction in the equations (Fig. 3b). The *r*CoRIA analysis also indicates that the hydrophobic interaction of the residues at position 7 in the peptide with the receptor should be improved to enhance binding, owing to the positive coefficients of this interaction in the equations (Fig. 3b). Similarly, the van der Waals interaction of the amino acids at positions 3 in the peptides with the receptor should be strengthened to ensure tighter binding, due to the negative coefficients of this interaction in the equations (Fig. 3b).

A careful examination of the models shows that hydrophobic amino acid residues with bulky extended side chains like Trp, Phe, Tyr, Leu and Ile are ideal at positions 1 and 3 in the peptides. Neutral or hydrophobic amino acids like Val, Leu, Ile, Met and Pro are recommended at positions 2 and 7, whereas charged residues like Arg, Lys, His, Glu, and Asp are favored at position 8 in the peptide. Most of these preferences for ligand residues are consistent with the previous studies [22, 29, 35–38] except for position 8 where

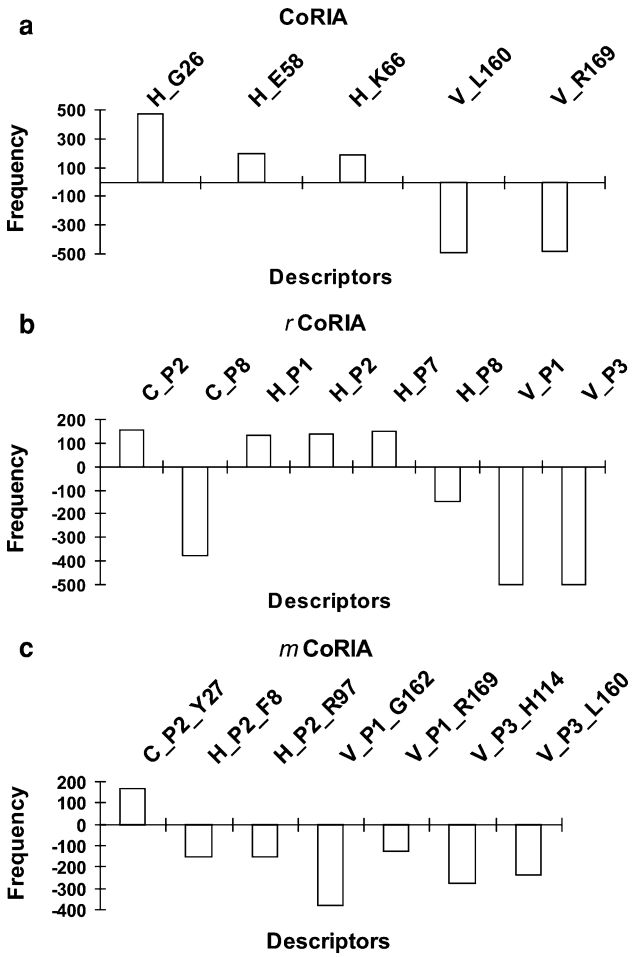

Fig. 3 Frequency plots of descriptors appearing in the equations of the (**a**) CoRIA, (**b**) *r*CoRIA and (**c**) *m*CoRIA models. C, V and H—Coulombic, van der Waals and hydrophobic interactions respectively. P1, P2, P3 etc—Residue at positions 1, 2 and 3 respectively in the peptide. H_G26—Hydrophobic interaction of the receptor residue Gly26 with the peptide. C_P2—Coulombic interaction of the residue at position 2 in the peptide with the receptor. V_P1_G162—van der Waals interaction of the residue at position 1 in the peptide with the receptor residue Gly162

hydrophilic short chain amino acids have been suggested, instead of the charged ones gleaned from this study. Based on the *r*CoRIA results, a hydrophilic charged residue at position 8 in the peptide will be most suitable.

*m*CoRIA analysis

The major drive behind the *m*CoRIA formalism is the fact that a greater detail of the thermodynamics of binding can be uncovered when both the receptor and the ligand are broken down into small units and the thermodynamic properties evaluated for each of these individual units (Fig. 1c). This fragmented receptor-ligand approach gives explicit information of the crucial interactions of amino acids in the peptide with specific residues of the receptor within the binding cavity. The frequency analysis of the descriptors appearing in the QSAR equations (Fig. 3c) highlights various interactions that rule the binding of amino acids at particular positions in the peptide, with specific residues in the receptor. Like the *r*CoRIA results, the *m*CoRIA models also suggest that the interaction of the receptor with the amino acids at positions 1, 2 and 3 in the peptides dictates the overall strength of binding. For example according to the *m*CoRIA analysis, the strength of the Coulombic interaction between the amino acid at position 2 in the peptides and the receptor residue Tyr27 needs to be reduced for tighter ligand binding, due to the positive coefficients of this interaction in the equations (Fig. 3c). This observation is partly supported by the *r*CoRIA equations which show a positive coefficient for the Coulombic interaction of the receptor with the residue at position 2 in the peptides (Fig. 3b). As highlighted by the models, the *HINT* scores of the interactions between the amino acid at position 2 in the peptides with the receptor residues Phe8 and Arg97 have negative coefficients in the equations (Fig. 3c), indicating that reducing these hydrophobic interactions will favor the binding process. In conjunction with the *r*CoRIA results, to improve the peptide-receptor binding it is necessary to enhance the hydrophobic interaction of the residue at position 2 in the peptide with the receptor as a whole (Fig. 3b), but the interaction is required to be abridged specifically with the receptor residues Phe8 and Arg97.

The van der Waals interactions of the residue at position 1 in the peptide with receptor residues Gly162 and Arg169 have negative coefficients in the QSAR equations (Fig. 3c), indicating that binding can be enhanced by increasing the strength of these interactions. This observation is justified partly by the *r*CoRIA results which suggest that van der Waals interaction between the receptor and the ligand residue at position 1 should be improved (vide supra), and also partly by the CoRIA results which recommends an increase in the van der Waals interaction between the peptide and the receptor residue Arg169 (vide supra). Similarly in order to improve binding according to the *m*CoRIA analysis, the van der Waals interaction of the residue at position 3 in the peptide with receptor residues His114 and Leu160 should be strengthened, due to the negative coefficients of these interactions in the equations (Fig. 3c). This observation is validated partly by the *r*CoRIA results which display a negative coefficient for van der Waals interaction between the receptor and the residue at position 3 in the peptide (Fig. 3b), and partly by the CoRIA results which show a negative coefficient for the van der Waals interaction of the peptide with the receptor residue Leu160 (Fig. 3a).

It is worth mentioning that descriptors like free energy of solvation, strain energy, entropy etc also appear in the QSAR models derived for the three CoRIA approaches, but their

frequency of occurrence is too low to be considered significant for peptide optimization. One explanation for the lack of appearance of these terms in the final equations could be due to the fact that for the present dataset of peptides with similar length and character, the free energy of solvation may not be a major determinant in the overall binding of these peptides to the MHC molecules. Another possibility is that some more in-depth theory or advanced methodology needs to be incorporated in the calculation of these properties, so that they can be picked up quickly and more frequently by a statistical tool, to be considered as important as other interaction terms in designing new compounds. Also, there is ample scope for improvement in the CoRIA methodologies by taking into consideration solvation of the entire ligand–receptor complexes followed by extensive sampling of configurations using molecular dynamics or Monte Carlo simulations prior to evaluation of the thermodynamic descriptors, and inclusion of the ligand–receptor intermolecular hydrogen-bonding terms preferably at the level of individual unit of the receptor and the ligand.

Combined analysis of QSAR models

The important descriptors that appear in the CoRIA, *r*CoRIA and *m*CoRIA models are presented in Table 4a–c

Table 4 Important (a) CoRIA, (b) *r*CoRIA, (c) *m*CoRIA, descriptors and their values for some selected molecules

(a) CoRIA

Mol	pIC_{50}	H_G26 (+)[@]	H_E58 (+)	H_K66 (+)	V_L160 (−)	V_R169 (−)
T55	8.770	−0.552	−0.386	−0.205	−1.892	−2.099
T39	7.480	−2.209	−0.252	−1.597	−1.537	−2.099
T31	7.174	−0.850	0.468	−0.813	0.082	0.161
T23	6.947	−1.529	0.468	−1.143	0.663	−0.969
T18	6.793	−0.850	−0.252	−1.061	−0.160	0.839
T2	6.000	−0.170	0.468	0.252	0.844	1.743
T1	5.342	−1.529	0.468	0.355	0.364	1.743

(b) *r*CoRIA

Mol	pIC_{50}	C_P2 (+)	C_P8 (−)	H_P1 (+)	H_P2 (+)	H_P7 (+)	H_P8 (−)	V_P1 (−)	V_P3 (−)
T55	8.770	−0.396	−0.568	−0.786	0.897	−0.066	−0.626	−0.315	−2.741
T39	7.480	−2.552	−0.584	−2.225	−1.046	−0.331	−0.701	−0.941	−1.817
T31	7.174	0.440	0.329	−0.805	−0.514	0.510	−0.236	0.121	−0.136
T23	6.947	−2.029	−0.528	−0.283	0.091	1.285	−0.424	−0.372	1.443
T18	6.793	−3.352	0.559	0.032	−1.674	0.909	0.455	0.258	−1.487
T2	6.000	0.527	0.430	−0.037	−0.489	−0.131	0.211	1.919	1.123
T1	5.342	−2.772	0.226	0.139	−1.038	0.627	1.344	1.894	0.011

*m*CoRIA

Mol	pIC_{50}	C_P2_Y27 (+)	H_P2_F8 (−)	H_P2_R97 (−)	V_P1_G162 (−)	V_P1_R169 (−)	V_P3_H114 (−)	V_P3_L160 (−)
T55	8.770	0.005	−1.122	−2.107	−0.489	−0.861	−1.746	−1.991
T39	7.480	−0.965	0.735	0.775	−0.489	−1.125	−2.069	−1.607
T31	7.174	1.260	0.735	0.775	−0.032	−0.069	0.709	−0.016
T23	6.947	−0.337	1.354	1.063	−0.945	−0.861	0.106	0.801
T18	6.793	−3.417	1.354	−1.243	0.728	0.723	−1.617	−0.081
T2	6.000	0.803	0.735	0.775	1.945	2.043	1.290	0.910
T1	5.342	−4.786	1.354	1.063	1.489	1.515	0.569	0.125

[@] Signs within the parenthesis refer to the signs of the coefficients of the respective descriptor in the QSAR equations

C, V and H—Coulombic, van der Waals and hydrophobic interactions respectively

P1, P2, P3 etc—Residue at positions 1, 2 and 3 respectively in the peptide

H_G26—Hydrophobic interaction of the receptor residue Gly26 with the peptide

C_P2—Coulombic interaction of the residue at position 2 in the peptide with the receptor

V_P1_G162—van der Waals interaction of the residue at position 1 in the peptide with the receptor residue Gly162

respectively, along with their values for some selected molecules. In this section, we look at how the values of these descriptors for some molecules are a reflection of their activity.

Molecule T55, the most active in the set has been taken as the reference for a comparison of the descriptors with other molecules, to understand the effect of the descriptors on the biological activity. According to the CoRIA model (Table 4a), the lower activity of molecule T39 compared to molecule T55 is the result of its reduced (more negative/less positive) hydrophobic interaction with the receptor residues Gly26 and Lys66 as well as due to its abridged (less negative/more positive) van der Waals interaction with the receptor residue Leu160.

According to the *r*CoRIA model, the lower activity of molecule T39 is partly due to strong (more negative/less positive) Coulombic interaction energy of its residue at position 2, reduced (more negative/less positive) hydrophobic interaction of its residues at positions 1, 2 and 7, as well as decreased (less negative/more positive) van der Waals interaction of its residue at position 3 with the receptor.

As explained by the *m*CoRIA model, molecule T39 is less active than molecule T55, partly because of stronger [increased (less positive/more negative)] Coulombic interaction of its residue at position 2 with the receptor residue Tyr27. The decrease in its activity also ensues from the increased (more positive/less negative) hydrophobic interaction of its residue at position 2 with the receptor residues Phe8 and Arg97. Weaker [reduced (less negative/more positive)] van der Waals interaction of its residue at position 3 with the receptor residue Leu160 is also responsible to a certain extent for the lower activity of molecule T39 compared to molecule T55. Figure 4 shows a stereo view of molecule T55 surrounded by important active site residues reflected in the *m*CoRIA equations. The interactions between specific residues of the ligand and the receptor, that are required to be strengthened and weakened according to the *m*CoRIA model, are shown by green and red arrows respectively. Interestingly, many of the electrostatic (and some of the van der Waals) interactions shown to be imperative by the CoRIA approaches are distant from the peptide molecule (up to 10 Å). This indicates that along with the direct interactions, indirect long-range interactions also significantly contribute towards the stability of the ligand–receptor complexes. Such observations have also been described in the literature [34, 58–60].

Likewise, the activity of the remaining molecules can be rationalized on the basis of the CoRIA equations (i.e. as a result of weak hydrophobic interaction with the receptor residues Gly26, Glu58 and Lys66, and/or weak van der Waals interaction with receptor residues Leu160 and Arg169) or as explained by the *r*CoRIA and *m*CoRIA models.

Application of CoRIA approaches in peptide optimization

In an attempt to demonstrate the usefulness of different CoRIA methodologies in designing new peptides with improved binding affinity for Class I MHC molecule HLA-A*0201, the most active peptide in the dataset T55 (with sequence ILWQVPFSV and pIC_{50} 8.770) was structurally modified based on the results of CoRIA, *r*CoRIA and *m*CoRIA approaches. For example, the residue (Ile) at position 1 in molecule T55 was replaced with the bulkier amino acid Phe, in order to strengthen its hydrophobic and van der Waals interaction with the receptor as a whole (as suggested by *r*CoRIA, Fig. 3b) and more precisely with the receptor residues Gly162 and Arg169 (as recommended partly by CoRIA as well as *m*CoRIA models, Fig. 3a and c respectively). Both single as well as double amino acid substitutions were made to generate new HLA binding peptides as shown in Table 5. After structural modifications, the peptide-receptor complexes were subjected to a thorough minimization procedure as discussed in the

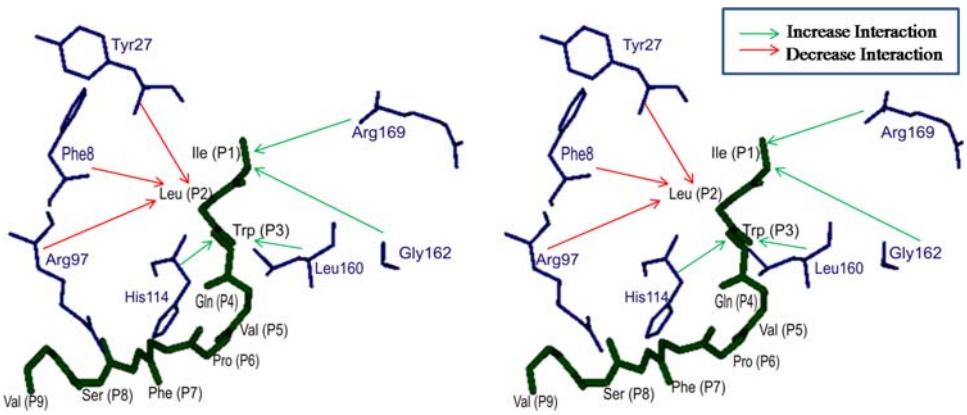

Fig. 4 A stereoview of the active site of HLA-A*0201 showing molecule T55 (green color, backbone atoms drawn) with important receptor residues (blue color, heavy atoms only) appearing in the *m*CoRIA equations. Green and red arrows indicate the interactions between specific residues of the ligand and the receptor, that are required to be increased and decreased respectively as per the *m*CoRIA model

Table 5 Predicted activities of molecule T55 and newly designed peptides by CoRIA approaches and online servers

Mol ID	Sequence	CoRIA pIC_{50}	*r*CoRIA pIC_{50}	*m*CoRIA pIC_{50}	SVMHC		PREDEP Score	MHC_PRED		MULTIPRED		MHCBPS Score	NETMHC Score
					SYFPEITHI Score	MHCPEP Score		AA pIC_{50}	AA + Int pIC_{50}	ANN Score	HMM Score		
T55	*ILWQVPFSV*	*8.66*	*8.34*	*8.70*	*1.30*	*1.42*	*−6.07*	*8.33*	*8.61*	*5.95*	*7.60*	*1.16*	*0.83*
1	**F**LWQVPFSV	9.11	8.83	9.16	1.17	1.71	−6.22	8.73	8.81	5.95	7.50	1.07	0.91
2	I**M**WQVPFSV	8.59	8.66	8.87	0.73	0.98	−5.95	8.28	8.32	5.85	7.66	1.28	0.85
3	IL**M**QVPFSV	8.76	8.70	9.25	1.31	1.26	−5.73	8.18	8.29	5.88	7.74	1.05	0.85
4	ILWQVPF**E**V	8.61	8.50	8.90	1.25	1.44	−5.98	8.02	8.21	5.95	7.80	0.14	0.82
5	**Y**LF**F**QVPFSV	8.90	8.80	8.95	1.31	1.51	−6.17	8.81	9.01	5.95	7.26	1.28	0.91
6	**Y**L**Y**QVPFSV	8.96	8.85	9.11	1.25	1.43	−6.14	9.02	9.03	5.96	7.29	1.28	0.91
7	**F**L**Y**QVPFSV	9.13	9.00	9.16	1.17	1.45	−5.97	8.83	8.81	5.95	7.25	1.20	0.91
8	**Y**LWQVPF**E**V	8.79	8.78	8.88	1.24	1.69	−6.29	8.61	8.73	5.96	7.73	0.25	0.91
9	**Y**LWQVPF**L**V	8.71	8.70	8.80	1.06	1.70	−7.02	8.86	8.80	5.96	7.28	0.75	0.89
10	**F**LWQVPF**L**V	8.92	8.82	9.00	1.03	1.77	−6.85	8.67	8.49	5.94	7.25	0.91	0.89

Prediction end points

CoRIA Methodologies: pIC_{50} (the higher the better)

SVMHC: Prediction scores based on training data from the SYFPEITHI and MHCPEP databases (the higher the better)

PREDEP: Energy scores (the lower the better)

MHCPRED: pIC_{50} from amino acids model (AA) and the model based on amino acids and interactions (AA + Int) (the higher the better)

MULTIPRED: Binding scores from Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) (the higher the better)

MHCBPS: Prediction scores (the higher the better)

NETMHC: Prediction score (the higher the better)

methodology section, and the non-bonded (van der Waals and Coulombic) and hydrophobic (in terms of *HINT* scores) interaction energies calculated separately for each of the three methodologies CoRIA, *r*CoRIA and *m*CoRIA. The activities of the newly designed peptides were then predicted by substituting the interaction energies into the best QSAR equations of the three models (Table 2). For comparison, the activities of the new peptides were also predicted from various available online servers like SVMHC [61], PREDEP [62], MHCPRED [33, 63, 64], MULTIPRED [65], MHCBPS [66] and NETMHC [67], specifically for their binding affinity towards Class I MHC molecule HLA-A*0201. Since the prediction end points (as shown in Table 5) corresponding to the activities/binding affinities of the peptides are different in case of some of these servers, a direct comparison may not be possible between different approaches (inter-methodology comparison), but the activities of the new peptides can certainly be compared within an approach/methodology (intra-methodology comparison) with the predicted values of the most active molecule, T55 (highlighted as italic). The predicted activity values by all the approaches for molecule T55 and the newly designed peptides are listed in Table 5, with the substituted amino acids highlighted as bold. Interestingly, with the exception of peptide 1 (FLWQVPFSV) which is predicted to be better than molecule T55, by nearly all the online servers, almost none of the other singly substituted peptides are predicted by the online servers to be as active as predicted by the three CoRIA methodologies. However, the new peptides generated by dual amino acid substitution are predicted to be much more active than molecule T55 by practically all the servers. It is apparent from this table that new peptides designed in line with the suggestions of the three CoRIA approaches by modifying more than one amino acid has improved binding affinity for Class I MHC molecule HLA-A*0201 compared to those designed on the basis of single amino acid substitutions.

Conclusions

In the present work, the recently developed QSAR formalism CoRIA [19], has been explored and extended further as two related methodologies—the reverse-CoRIA (*r*CoRIA) and mixed-CoRIA (*m*CoRIA) approaches. In the *r*CoRIA technique, the ligand is fragmented into its constituent units and the interaction of each individual unit is calculated with the receptor as a whole. In the *m*CoRIA approach, both the ligand as well as the receptor are fragmented into smaller units or residues, and the interaction of each unit of the ligand is calculated with individual active site residues of the receptor. The efficiency of the three approaches (CoRIA, *r*CoRIA and *m*CoRIA) has been tested on a standard dataset of diverse nonamer peptides that bind to the Class I major histocompatibility complex molecule HLA-A*0201. The QSAR models developed from the three approaches yield statistically significant results and throw deep insight into the factors that govern ligand–receptor binding. The methodologies have been able to reveal all structure activity relationships reported for this class of molecules as well uncover some that were hitherto unknown. Thus, the approach can confidently be used on other datasets for which nothing or very little SAR is known.

The CoRIA, *r*CoRIA and *m*CoRIA approaches work in tandem to successfully dig out all crucial interactions that modulate the binding of the nonameric antigens to the HLA receptor for which some information was available from earlier studies. The equations derived by the three approaches have also uncovered various other aspects that have not yet been explored and may have a hidden role in ligand–receptor recognition process. The methodologies can be used to extract position-specific information about the type of residues or the nature of interactions that are important for binding and can also serve as a guide for conducting mutation studies directed towards understanding ligand–receptor thermodynamics and for optimizing lead molecules. The *r*CoRIA and *m*CoRIA methodologies involve fragmentation of the ligand (in addition to the receptor) into smaller units, and in this study was illustrated for the peptide class, as peptides can very logically be broken down into individual units. Application of these techniques to small organic molecules is underway. In conclusion, the present QSAR techniques draw out all the major thermodynamic events that govern ligand–receptor binding and can be used as a powerful tool to support the drug design process.

Acknowledgements This work was made possible by a grant [01(1986)/05/EMR-II] from the Council of Scientific and Industrial Research (CSIR, New Delhi). The Department of Science and Technology (DST, New Delhi) is also thanked for providing some of the computational facilities under the FIST program (SR/FST/LSI-163/2003). Jitender Verma, S. A. Khedkar and A. K. Malde thank CSIR for financial support. V. M. Khedkar thanks the Amrut Mody Research Foundation (AMRF) for the support.

References

1. Hansch C, Maloney PP, Fujita T, Muir RM (1962) Nature 194:178
2. Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M (1963) J Am Chem Soc 85:2817
3. Hansch C, Fujita T (1964) J Am Chem Soc 86:1616
4. Hansch CA (1969) Acc Chem Res 2:232
5. Cramer RD III, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959
6. Tokarski JS, Hopfinger AJ (1994) J Med Chem 37:3639
7. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130

8. Havel TF, Kuntz ID, Crippen GM (1983) Bull Math Biol 45:665
9. Doweyko AM (1988) J Med Chem 31:1396
10. Walters DE, Hinds RM (1994) J Med Chem 37:2527
11. Jain AN, Koile K, Chapman D (1994) J Med Chem 37:2315
12. Hopfinger A, Wang S, Tokarski J, Baiqiang J, Albuquerque M, Madhav P, Duraiswami C (1997) J Am Chem Soc 119:10509
13. Vedani A, Dobler M, Prog Drug Res (2000) 55:105
14. Vedani A, Dobler M (2002) J Med Chem 45:2139
15. Vedani A, Dobler M, Lill MA (2005) J Med Chem 48:3700
16. Tokarski JS, Hopfinger AJ (1997) J Chem Inf Comput Sci 37:792
17. Ortiz AR, Pisabarro TM, Gago F, Wade RC (1995) J Med Chem 38:2681
18. Gohlke H, Klebe G (2002) J Med Chem 45:4153
19. Datar PA, Khedkar SA, Malde AK, Coutinho EC (2006) J Comput Aided Mol Des 20:343
20. Khedkar SA, Malde AK, Coutinho EC (2007) J Chem Inf Model 47:1839
21. Marshall GR (1993) Tetrahedron 49:3547
22. Doytchinova IA, Flower DR (2001) J Med Chem 44:3572
23. Doytchinova IA, Flower DR (2002) J Comput Aided Mol Des 16:535
24. Sneath PH (1966) J Theor Biol 12:157
25. Kidera A, Konishi Y, Oka M, Ooi T, Scheraga HA (1985) J Prot Chem 4:23
26. Hellberg S, Sjostrom M, Skagerberg B, Wold S (1987) J Med Chem 30:1126
27. Collantes ER, Dunn WJ (1995) J Med Chem 38:2705
28. Zaliani A, Gancia E (1999) J Chem Inf Comput Sci 39:525
29. Pissurlenkar RRS, Malde AK, Khedkar SA, Coutinho EC (2007) QSAR Comb Sci 26:189
30. Doytchinova IA, Walshe VA, Jones NA, Gloster SE, Borrow P, Flower DR (2004) J Immunol 172:7495
31. Hattotuwagama CK, Doytchinova IA, Flower DR (2005) J Chem Inf Mod 45:1415
32. Hattotuwagama CK, Guan P, Doytchinova IA, Flower DR (2004) Org Biomolec Chem 2:3274
33. Hattotuwagama CK, Guan P, Doytchinova IA, Zygouri C, Flower DR (2004) J Mol Graph Model 22:195
34. Davies MN, Hattotuwagama CK, Moss DS, Drew MGB, Flower DR (2006) BMC Structural Biology 6:5
35. Guan P, Doytchinova IA, Walshe VA, Borrow P, Flower DR (2005) J Med Chem 48:7418
36. Doytchinova IA, Flower DR (2007) J Chem Inf Model 47:234
37. Doytchinova IA, Flower DR (2002) Proteins 48:505
38. Doytchinova IA, Walshe V, Borrow P, Flower DR (2005) J Comput Aided Mol Des 19:203
39. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A (1993) Cell 74:929
40. Sette A, Sidney J, del Guercio M-F, Southwood S, Ruppert J, Dalberg C, Grey HM, Kubo RT (1994) Mol Immunol 31:813
41. Cerius2, version 4.6 (1998) Accelrys Inc., San Diego, CA, USA
42. InsightII, version 2005L (2005) Accelrys Inc., San Diego, CA, USA
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
44. Maple JR, Hwang M-J, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) J Comput Chem 15:162
45. Kellogg GE, Semus SF, Abraham DJ (1991) J Comput Aided Mol Des 5:545
46. Sybyl, version 7.1 (2005) Tripos Associates Inc., 1699 S Hanley Rd., St. Louis, MO 631444, USA
47. Murcko MA, Stouten PFW (1997) In: Charifson PS (ed) Practical application of computer-aided drug design, Marcel Dekker Inc., New York, USA, p 355
48. Quasar, version 5.0 (2005) Biographics laboratory 3R, Basel, Switzerland
49. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) J Am Chem Soc 112:6127
50. Searle MS, Williams DH (1992) J Am Chem Soc 114:10690
51. Rogers D, Hopfinger AJ (1994) J Chem Inf Comput Sci 34:854
52. Wold S, Johansson E, Cocchi M (1993) In: Kubinyi H (ed) 3D QSAR in drug design: theory, methods and applications, ESCOM Science Publishers, Leiden, p 523
53. Richard D, Cramer RD III, Bunce JD, Patterson DE, Frank IE (1988) Quant Struct-Act Relat 7:18
54. Madden DR (1995) Annu Rev Immunol 13:587
55. Madden DR, Garboczi DN, Wiley DC (1993) Cell 75:693
56. Falk K, Rötzschke O, Stefanovic S, Jung G, Rammensee H-G (1991) Nature 351:290
57. Holtje H-D, Sippl W, Rognan D, Folkers G (2003) In: Holtje H-D, Folkers G (eds) Molecular modeling—basic principles and applications, WILEY—VCH GmbH & Co. KGaA, Weinheim, p 179
58. Binz AK, Rodriguez RC, Biddison WE, Baker BM (2003) Biochemistry 42:4954
59. Marvin JS, Hellinga HW (2001) Nat Struct Biol 8:795
60. Oelschlaeger P, Mayo SL, Pleiss J (2005) Protein Sci 14:765
61. Dönnes P, Elofsson A (2002) BMC Bioinformatics 3:25
62. Altuvia Y, Sette A, Sidney J, Southwood S, Margalit H (1997) Hum Immunol 58:1
63. Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) Appl Bioinformatics 2:63
64. Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) Nucleic Acids Res 31:3621
65. Zhang GL, Khan AM, Srinivasan KN, August JT, Brusic V (2005) Nucleic Acids Res 33:W172
66. Bhasin M, Singh H, Raghava GPS (2003) Bioinformatics 19:665
67. Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S (2003) Tissue Antigens 62:378