Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 209–220
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus

Vladimír Frecer [a,b], Martin Kabeláč [a,1], Piergiuseppe De Nardi [c], Sabrina Pricl [c], Stanislav Miertuš [a,*]

[a] *International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, I-34012 Trieste, Italy*
[b] *Cancer Research Institute, Slovak Academy of Sciences, SK-83391 Bratislava, Slovak Republic*
[c] *Department of Chemical, Environmental and Raw Materials Engineering, University of Trieste, Piazzale Europa 1, I-34127 Trieste, Italy*

Received 25 January 2002; received in revised form 15 May 2003; accepted 15 July 2003

Abstract

We have designed small focused combinatorial library of hexapeptide inhibitors of NS3 serine protease of the hepatitis C virus (HCV) by structure-based molecular design complemented by combinatorial optimisation of the individual residues. Rational residue substitutions were guided by the structure and properties of the binding pockets of the enzyme's active site. The inhibitors were derived from peptides known to inhibit the NS3 serine protease by using unusual amino acids and α-ketocysteine or difluoroaminobutyric acid, which are known to bind to the S_1 pocket of the catalytic site. Inhibition constants (K_i) of the designed library of inhibitors were predicted from a QSAR model that correlated experimental K_i of known peptidic inhibitors of NS3 with the enthalpies of enzyme–inhibitor interaction computed via molecular mechanics and the solvent effect contribution to the binding affinity derived from the continuum model of solvation. The library of the optimised inhibitors contains promising drug candidates—water-soluble anionic hexapeptides with predicted K_i^* in the picomolar range.

Keywords: Hepatitis C virus; NS3 serine protease; Structure-based molecular design; Molecular modelling; Peptidic inhibitors; Combinatorial optimisation

1. Introduction

More than 170 million people world-wide are chronically infected with hepatitis C virus (HCV), which causes hepatitis that may eventually develop over a longer period of time into liver cirrhosis, hepatocellular carcinoma or liver failure [1,2]. Current therapy using a combination of α-interferon and ribavirin is effective only in about 50% of cases and exhibits severe adverse side effects [3,4]. Replication of the HCV crucially depends on the maturation of the viral polyprotein precursor encoded in the positive-sense single-stranded RNA of the HCV NH_2–C–E1–E2–p7–NS2–NS3–NS4A–NS4B–NS5A–NS5B–COOH, which is cleaved into 10 viral proteins [5]. The viral non-structural protein 3 (NS3) is a multifunctional enzyme possessing serine protease activity in the N-terminal third of the protein and a RNA helicase/NTPase activity in the C-terminal portion. The serine protease domain of NS3, which needs to associate with the cofactor NS4A for efficient processing, is responsible for four out of five cleavage events in the non-structural region of the HCV polyprotein [6,7].

Crystal structure of the serine protease domain of the NS3 with its essential cofactor NS4A [8,9] revealed that the NS3/4A complex adopts a chymotrypsin/trypsin-like fold with structurally conserved regions typical of small chymotrypsin-like proteases [10,11]. The N-terminal region of NS3/4A (residues 1–93 of NS3 and residues 21–34 of NS4A) contains an eight-stranded β-barrel motif, with one

Abbreviations: Ac, acetyl; Asa, β-carboxyaspartic acid; *t*Bu, *tert*-butylglycine; Cha, β-cyclohexylalanine; Cyo, α-ketocysteine; Dif, β,β-diphenylalanine; Fab, δ,δ-difluoro-β-amino-α-ketopentanoic acid; Gla, γ-carboxyglutamic acid; Glr, glutaric acid; Cpa, β-carboxypropionylalanine; Nal, β-napthylalanine; Nap, napthylglycin; Nva, norvaline; Phg, α-phenylglycine; Suc, succinic acid; Tro, 7-hydroxytryptofan; Trc, 4-carboxytryptofan; HCV, hepatitis C virus; NS3, viral non-structural protein number 3 of HCV; NS4A, viral non-structural protein number 4 of HCV, cofactor of the NS3; NS3/4A, complex of NS3 with cofactor NS4A; MM, molecular mechanics; QSAR, quantitative structure–activity relationships; K_i, inhibition constant; $\log P_{o/w}$, log of partitioning coefficient in octanol/water system.

* Corresponding author. Tel.: +39-040-922-8114; fax: +39-040-922-8115.
E-mail address: miertus@ics.trieste.it (S. Miertuš).
[1] Present address: J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of Czech Republic, CZ-18223 Prague, Czech Republic.

1093-3263/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1093-3263(03)00161-X

of the strands contributed by the NS4A cofactor [10]. The C-terminal region (residues 94–175) contains a six-stranded β-barrel that ends with a helix. The active site (His:57, Asp:81 and Ser:139) is located between these two regions and is formed by a shallow solvent exposed pocket requiring many interaction points for binding of substrates or inhibitors [8–11]. Thus, the NS3 protease displays substrate specificity that requires relatively large peptides spanning the active site from S_6 to $S_{4'}$ pocket [12–15,20]. The catalytic site contains an oxyanion hole, which stabilizes the hemiketal quaternary cleavage intermediate by hydrogen bonds with amide protons of the catalytic serine residue Ser:139 and glycine Gly:137 [11]. Conserved features of the substrates recognized by the NS3 protease include acidic residue in P_6 and P_5 positions, preference for cysteine in P_1 and hydrophobic residues in $P_{4'}$ [13,16,17]. Substrates and inhibitors typically bind to the active site of the NS3 in an extended conformation and form an antiparallel β-sheet with the protease with one strand contributed by the inhibitor and the other strand contributed by the protease [18].

The complex NS3/4A has been identified as a promising target for antiviral drugs effective against the HCV [15,19,20]. Recently, it has been reported that N-terminal cleavage products of the substrate form competitive inhibitors of the NS3 protease activity. These native inhibitors (typically hexapeptides) served as the basis for designing substrate-based inhibitors, sequences of which were derived from the polyprotein precursor sites cleaved by the NS3 protease [19–23]. Llinàs-Brunet et al. [21,22] prepared a highly potent hexapeptide that displayed a low nanomolar inhibition constant (K_i) towards the NS3/4A and was selective with respect to other serine proteases. Other inhibitors take advantage of the fact that serine proteases can be inhibited by strong electrophiles located at the position of the scissile amide bond. Dundson et al. [24] designed two heptapeptides containing aminoboronic acid with K_i values against the NS3 protease of about 80 nM. Narjes et al. [25] reported that α-ketoacids such as difluoroaminobutyric acid where the difluorcarbon group of the side chain mimics the thiol group of cysteine residue present at the native substrate cleavage sites are potent slow binding inhibitors of the NS3 protease of HCV. Peptidyl trifluoromethyl ketones have been described as inhibitors of chymotrypsin and other serine proteases by Cassidy et al. [26]. Also rhodanine derivates were known to possess biological activities such as antibacterial, antivirial and antidiabetic. Sudo et al. [27] reported rhodanine derivatives that make NS3 protease inhibitors, however, these compounds possess higher activity towards other serine proteases such as chymotrypsin and plasmin. Sing et al. [28] have shown that bulkier and hydrophobic functional groups in arylalkylidene rhodanines increase selectivity to the NS3 protease of HCV. However, their high molecular weight decreases their potential for future therapeutic use. Yeung et al. [29] identified a novel type of bisbenzimidazole-based inhibitors of the HCV NS3 protease. Vertex pharmaceuticals together with Eli Lilly are developing a small molecule inhibitor of NS3/4A protease VX-950 (LY570310), which has entered pre-clinical studies.

In this work, we aim to design new potent pseudo-peptidic inhibitors of the NS3 protease of HCV by using structure-based molecular design and combinatorial optimisation of a focused library of potential inhibitors. The inhibitors were derived from known peptidic inhibitors specific to the NS3/4A protease of HCV and their experimentally determined inhibition constants [23], which permitted to predict the inhibitory effect of newly designed and optimised derivatives.

2. Materials and methods

2.1. Model building

Crystal structures of NS3 protease with NS4A cofactor [9] and of the ternary complex of NS3 and NS4A co-crystallised with tetrapeptide inhibitors [30] obtained from the *protein data bank* [31] were used for the structure-based design of novel pseudo-peptide inhibitors. The bound conformations of the designed inhibitors were modelled by prolongation of the backbone of the original inhibitors up to hexapeptides with the side-chains filling the S_6–S_1 pockets [12] of the NS3 protease active site. An approximately extended backbone conformation was used [18] in accordance with the observation that bound peptide inhibitors form an antiparallel β-sheet structure with one strand provided by the enzyme and the other by the inhibitor [20,21,30]. The P_1 position of the inhibitor was occupied by electrophiles such as α-ketocysteine or trifluoromethyl ketones [25,26], which were modelled in their non-covalently bound form. To accommodate the side chains of the individual residues of the designed inhibitors in the specificity pockets of the NS3/4A binding site we carried out a complete torsion force scan of potential energy hypersurface over all rotatable bonds of the considered side chain. The force constant of the torsion force driving the scanned torsion angles to values 0–360° with an increment of 30°, was set equal to 100 kcal mol^{-1} deg^{-2}. All possible combinations of individual torsion angles of the side chain were generated and each structure was fully optimised using molecular mechanics. The structures of free inhibitors (hydrophilic anionic peptides) were generated from their bound (extended) conformations by geometry optimisation. Modelling of the NS3/4A protease, designed inhibitors and the enzyme–inhibitor complexes was done using *Insight II* molecular modelling software [32].

2.2. Molecular mechanics

Simulations of the models of inhibitors, NS3/4A serine protease and their complexes were carried out using all-atom representation in the class II consistent force field and charge parameters cff91 [33]. A dielectric constant of 4 was used for all molecular mechanics (MM) calculations in order to

take into account the shielding in proteins. Minimisations of the enzyme–inhibitor complexes, free enzyme and free inhibitors were carried out by relaxing the structures gradually, starting with the side chains and followed by protein or peptide backbone relaxation. In all the geometry optimisations, a sufficient number of steepest descent and conjugate gradient iterative cycles were used with the convergence criterion for the average gradient set to 0.01 kcal mol^{-1} $Å^{-1}$. All structures were considered to be at neutral pH with the protonizable and ionisable groups being charged. Thus total molecular charge of the serine protease domain of the NS3/4A protein–cofactor receptor (Q_E) was equal to $7e^-$.

2.3. Solvation

Inclusion of solvent effects in the theoretical prediction of inhibition constants improves significantly the predictability of enzyme–inhibitor binding especially for charged inhibitors [34]. It was shown previously that solvent effects of ionic species are closely related to their molecular charge [35]. Continuum models of solvation have proven useful in biological applications where the description of bulk solvent effects on larger solutes via explicit solvent models is limited by the size or prohibitive simulation times [36]. We computed the solvation energy of the enzyme–inhibitor complexes, free enzyme and free inhibitors using the version of *polarizable continuum model* (PCM) [37] adapted for calculations on biopolymers [38]. This solvation model considers solvent as a homogeneous medium characterised by macroscopic properties, such as permittivity, polarisability density and molar volume. It employs rigorous treatment of solute–solvent interactions including the electrostatic, dispersion and repulsion terms and involves the cavitation term that accounts for the creation of a realistic cavity reproducing van der Waals molecular surface of the solute. A dielectric constant of 1 was used for the solute and 78.5 for water. Atomic radii and charges were taken from the cff91 force field [33] and atomic polarisabilities of Thole [39] were employed.

2.4. Calculation of enzyme–inhibitor binding affinities

The association of peptidic inhibitor (I) with its target enzyme (NS3) in solution to form a molecular complex (NS3:I) is a reversible process, which can be represented by the following equilibrium:

$$\{\mathrm{NS3}\}_{\mathrm{aq}} + \{\mathrm{I}\}_{\mathrm{aq}} \Leftrightarrow \{\mathrm{NS3:I}\}_{\mathrm{aq}} \quad (1)$$

Gibbs free energy change connected with the enzyme–inhibitor complex formation, ΔG_{comp}, can be obtained from standard Gibbs free energies of the associating particles at equilibrium:

$$\Delta G_{\mathrm{comp}} = G\{\mathrm{NS3:I}\}_{\mathrm{aq}} - [G\{\mathrm{NS3}\}_{\mathrm{aq}} + G\{\mathrm{I}\}_{\mathrm{aq}}] \quad (2)$$

We approximated the exact thermodynamic values of the Gibbs free energy for larger systems such as the complex {NS3:I} by the expression [34,40]:

$$G\{\mathrm{NS3:I}\}_{\mathrm{aq}} \cong E_{\mathrm{MM}}\{\mathrm{NS3:I}\} + RT + G_{\mathrm{solv}}\{\mathrm{NS3:I}\} \quad (3)$$

where E_{MM}{NS3:I} stands for the MM total energy and G_{solv}{NS3:I} is the Gibbs free energy of solvation. Changes in the conformational entropies of the associating particles were neglected. Thermal averaging of the enzyme–inhibitor complexes was taken into account only approximately as the complexes were modelled from thermally averaged X-ray structure by inhibitor residue replacements. Using Eq. (3), the Gibbs free energy change of the NS3:I complex formation (binding affinity, BA) in water can be evaluated as the sum of interaction and solvent effect contributions:

$$\Delta G_{\mathrm{comp}} = -\mathrm{BA} \cong \Delta H_{\mathrm{int}} + \Delta G_{\mathrm{solv}} \quad (4)$$

where ΔH_{int} is the enthalpic part of the enzyme–inhibitor interaction and ΔG_{solv} is the solvent effect contribution. The complexation Gibbs free energy thus takes into account not only the interactions in the complex but also the stability of the free inhibitor, free enzyme and the effect of solvent upon the enzyme and inhibitor association. Comparison between different inhibitors was done via relative changes in the complexation Gibbs free energy, $\Delta\Delta G_{\mathrm{comp}} \cong \Delta\Delta H_{\mathrm{int}} + \Delta\Delta G_{\mathrm{solv}}$, where

$$\Delta\Delta H_{\mathrm{int}} = [E_{\mathrm{MM}}\{\mathrm{NS3:I}\} - (E_{\mathrm{MM}}\{\mathrm{NS3}\} + E_{\mathrm{MM}}\{\mathrm{I}\}) - RT] - \Delta H_{\mathrm{int}}\{\mathrm{NS3:I_{ref}}\} \quad (5)$$

$$\Delta\Delta G_{\mathrm{solv}} = [G_{\mathrm{solv}}\{\mathrm{NS3:I}\} - (G_{\mathrm{solv}}\{\mathrm{NS3}\} + G_{\mathrm{solv}}\{\mathrm{I}\})] - \Delta G_{\mathrm{solv}}\{\mathrm{NS3:I_{ref}}\} \quad (6)$$

with respect to the interaction and solvent effect terms of a reference inhibitor (I_{ref}). The evaluation of relative changes is preferable as it is expected to lead to cancellation of errors caused by the approximate nature of the MM method and the PCM solvent effect description as well as the neglect of conformational entropy changes. The enzyme inhibition constant (K_i) is related to the relative changes in the binding affinity contributions of an inhibitor I to the NS3/4A protease as: $\ln K_i\{\mathrm{I}\} \approx -a(\Delta\Delta H_{\mathrm{int}} + \Delta\Delta G_{\mathrm{solv}})/RT + b$, where a and b are regression coefficients.

3. Results and discussion

3.1. Validation of computational approach

To verify the validity of the structure-based computational procedure for prediction of binding affinities of newly designed peptidic inhibitors of the NS3/4A protease we first modelled the interactions with the enzyme for a series of 16 known inhibitors, for which the experimental inhibition constants were previously determined [23] (Table 1).

Table 1
Relative interaction and solvent effect contributions to the enzyme–inhibitor complexation Gibbs free energy computed for a series of known peptidic inhibitors of the NS3/4A protease of HCV. Experimental inhibition constants of the inhibitors towards the serine protease domain of the NS3/4A were taken from [23]

Inhibitor	Chemical structure[a] P_6–P_5–P_4–P_3–P_2–P_1	M_w[b] (Da)	Q_i[c] (e^-)	$\Delta G_{solv}\{I\}$[d] (kcal mol^{-1})	$\Delta\Delta H_{int}$[e] (kcal mol^{-1})	$\Delta\Delta G_{solv}$[f] (kcal mol^{-1})	K_i^{exp}[g] (μM)
I1	AcAsp–Glu–Dif–Glu–Cha–Cys	936	−4	197.4	24.6	−190.7	0.025
I2	AcGlu–Dif–Glu–Cha–Cys	822	−3	350.3	69.9	−378.3	0.7
I3	AcDif–Glu–Cha–Cys	694	−2	467.9	122.4	−588.5	15
I4	AcGlu–Cha–Cys	471	−2	466.8	137.8	−596.4	230
I5	AcAsp–Glu–Dif–Ile–Cha–Cys	921	−3	346.6	71.5	−399.2	0.03
I6	AcGlu–Dif–Ile–Cha–Cys	807	−2	474.0	123.9	−586.4	1.2
I7	AcDif–Ile–Cha–Cys	679	−1	609.0	175.2	−794.7	50
I8	AcIle–Cha–Cys	456	−1	563.8	187.8	−794.2	–[h]
I9	AcAsp–D-Glu–Leu–Glu–Cha–Cys	826	−4	195.4	29.0	−187.0	0.023
I10	AcAsp–Glu–Leu–Glu–Cha–Cys	826	−4	190.3	29.5	−184.7	0.06
I11	AcAsp–D-Gla–Leu–Ile–Cha–Cys	854	−4	158.8	12.9	−147.5	0.00075
I12	AcAsp–Glu–Met–Glu–Cha–Cys	844	−4	189.6	31.2	−188.9	0.18
I13	AcAsp–Glu–Dif–Lys–Cha–Cys	936	−2	444.9	139.7	−625.8	–
I14	AcAsp–Glu–Met–Glu–Nal–Cys	874	−4	192.7	34.4	−198.4	0.4
I15	AcAsp–Glu–Met–Glu–Glu–Cys	819	−5	0.0	0.0	0.0	0.5
I16	Asp–D-Glu–Leu–Glu–Cha–Cys	894	−3	334.2	20.5	−446.9	–

[a] N-terminal residue of the peptide inhibitor is at the P_6 position [12], for acronyms used for unusual amino acids, see *Abbreviations*.
[b] Molecular weight of the inhibitor.
[c] Molecular charge of the inhibitor. Total molecular charge of the serine protease domain of the NS3/4A protein–cofactor complex is $7e^-$.
[d] Relative Gibbs free energy of hydration of free inhibitors calculated using *polarizable continuum model* [36,37]. The $\Delta G_{solv}\{I\}$ quantity is useful for predicting water solubility of the inhibitors. $\Delta G_{solv}\{I\}$ was taken with respect to the $G_{solv}\{\mathbf{I}15\}$ of the reference inhibitor.
[e] Relative interaction contribution to the enzyme–inhibitor binding affinity was calculated using Eq. (5) and taken with respect to $\Delta H_{int}\{\mathbf{I}15\}$ of the reference inhibitor, which displayed the lowest computed value of ΔH_{int}.
[f] Relative solvent effect contribution to the enzyme–inhibitor binding affinity was calculated using Eq. (6) and taken with respect to $\Delta G_{solv}\{\mathbf{I}15\}$ of the reference inhibitor.
[g] Experimental inhibition constants of the modelled peptidic inhibitors towards the NS3/4A serine protease domain were taken from [23].
[h] Experimental data were not available.

The models of the known inhibitors were derived from the crystal structures of enzyme–inhibitor complexes containing tetrapeptide inhibitors [30]. Spatial arrangement of the side chains in the model of the most potent inhibitor from the considered series, **I**11, bound to the active site of the NS3/4A is depicted in Fig. 1. The acidic hexapeptide **I**15 was selected as the reference inhibitor (I_{ref}) since it exhibited the strongest computed interaction with the NS3/4A from the training set of modelled inhibitors. Closer inspection of the relative interaction and solvent effect contributions to the enzyme–inhibitor binding affinity reveals that both the enzyme–inhibitor ($\Delta\Delta H_{int}$) as well as the solute–solvent ($\Delta\Delta G_{solv}$) interaction terms are dominated by electrostatic interactions, both correlate well with the molecular charge of the inhibitor (Q_i) which ranges from $-1e^-$ to $-5e^-$ and $\Delta\Delta H_{int}$ and $\Delta\Delta G_{solv}$ mutually compensate their contributions to the total enzyme–inhibitor binding affinity (Table 1). This is not surprising since binding of the highly charged acidic inhibitors to the basic charged catalytic site of the NS3/4A enzyme with overall charge $Q_E = +7e^-$ was considered (Table 1). Besides the inhibitor charge, the relative $\Delta\Delta H_{int}$ and $\Delta\Delta G_{solv}$ terms vary also with the amino acid sequences of individual inhibitors at constant Q_i.

Computed interaction and solvent effect contributions to the enzyme inhibitor binding affinity, shown in Table 1, were correlated with the experimental pK_i values using multivariate linear regression. The resulting correlation equations and their statistical characteristics are given in Table 2. Direct correlation between the pK_i and $\Delta\Delta G_{comp}$ for the series of considered inhibitors did not yield a QSAR model with high predictive value since the magnitudes of the interaction and solvent effect contributions to the $\Delta\Delta G_{comp}$ were not well balanced. This was caused by utilisation of different computational approaches where the MM total energies were derived from a force field simulation at the atomic level of resolution while the solvent effects were computed via a continuum model of solvation. However, a significant correlation was achieved between pK_i and the individual contributions to the relative binding affinity:

$$pK_i = a \cdot \Delta\Delta H_{int} + b \cdot \Delta\Delta G_{solv} + c \quad (7)$$

which was able to correlate the computed $\Delta\Delta H_{int}$ and $\Delta\Delta G_{solv}$ quantities of anionic peptidic training set inhibitors with different polarities (Q_i), sequences and sizes (tripeptides to hexapeptides) with the experimental inhibition constants ranging over seven orders of magnitude (K_i from 0.75 nM to 230 μM, Table 1) with a *leave-one-out* cross-validated correlation coefficient, $cvr^2 = 0.91$ (Eq. (b) in Table 2). This relationship validates our computational approach and indicates high ability of the binding model to

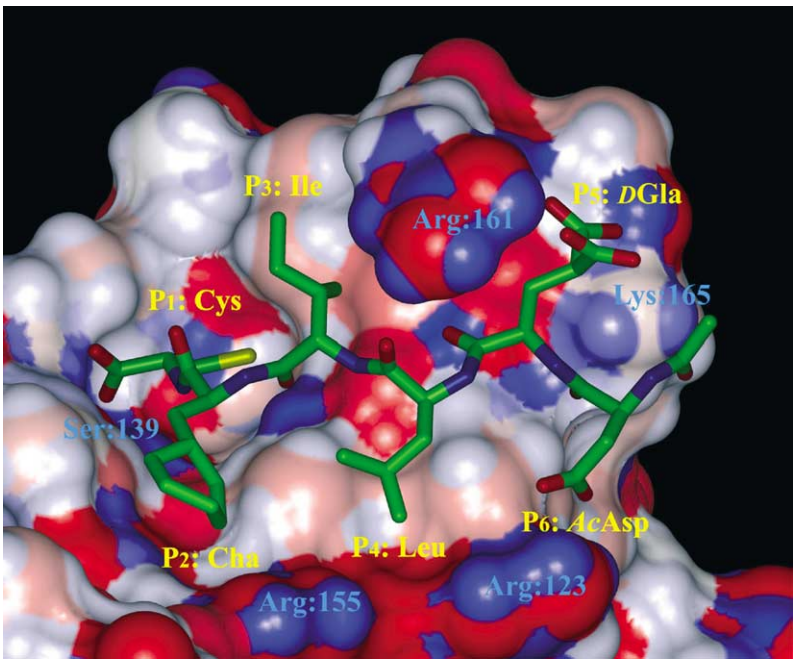

Fig. 1. Connolly surface representation of the active site of the serine protease domain of the NS3/4A protease of HCV with the bound known inhibitor **I**11 [23] with inhibition constant of the NS3/4A in the picomolar range (stick model coloured by atom types, hydrogen atoms were omitted for better clarity). Connolly surface of the active site of the protease is coloured according to a charge spectrum: acidic groups are red, basic groups are blue and neutral groups are white. The inhibitor residues are labelled in yellow, while the positions of basic resides in the active site of the protease are indicated with blue labels.

predict the inhibitory potencies of newly designed derivatives with the same mode of action.

Since an easy-to-calculate parameter such as the molecular charge of the inhibitor Q_i correlates very well with the $\Delta\Delta G_{solv}$ contribution (Eq. (c) in Table 2) we replaced the solvent term in the regression equation by the Q_i descriptor to facilitate rational structure-based design and residue optimisation for a larger series (hundreds) of designed and screened peptidic inhibitors. In this approximation to the rigorous binding model (Eq. (b) in Table 2), we assumed that the side chains of all considered natural and unusual acidic amino acid residues will be ionised both in solution and in the complex with the NS3/4A at neutral pH, irrespective of the molecular charge of the inhibitor and possible coupling between pK_a constants of the individual residues. Thus, we obtained a simple predictive QSAR model applicable to structure-based design of a library of inhibitors in the simplified form:

$$pK_i = a' \cdot \Delta\Delta H_{int} + b' \cdot Q_i + c' \quad (8)$$

(Eq. (d) in Table 2, Fig. 2), which was then used for the estimate of inhibitory potencies of newly designed inhibitors of the NS3/4A of HCV.

Higher activity of a designed derivative is frequently connected with higher or lower magnitudes of relevant molecular property/properties compared to the property/properties

Table 2
QSAR analysis of the training set of known peptidic inhibitors **I**1–**I**16 of serine protease domain of the NS3/4A of HCV given in Table 1

Eq. no.	Correlation[a]	r^2[b]	σ[c]	F-test[d]	cvr^2[e]
(a)	$pK_i = 0.006\Delta\Delta G_{comp} + 2.069$	0.52	1.09	11.73	0.55
(b)	$pK_i = -0.133\Delta\Delta H_{int} - 0.027\Delta\Delta G_{solv} + 0.150$	0.90	0.53	43.51	0.91
(c)	$\Delta\Delta G_{solv} = -201.640Q_i - 993.043$	1.00[f]	14.16	3306.95	1.00[f]
(d)	$pK_i = -0.109\Delta\Delta H_{int} + 4.226Q_i + 21.289$	0.90	0.49	50.83	0.92

[a] Experimental inhibition constants ($pK_i = -\log_{10} K_i$) of inhibitors **I**1–**I**16 (Table 1) towards the serine protease domain of NS3/4A [23] were correlated with computed quantities (Table 1). The QSAR correlation equations were derived by linear multivariate regression analysis, regression coefficients a, b, c and a', b', c' of Eqs. (7) and (8) are given. Number of experimental points $n = 13$, level of statistical significance >95% ($\alpha = 0.05$), range of activity was >6 orders of magnitude. Q_i is in e^-, $\Delta\Delta G_{comp}$, $\Delta\Delta H_{int}$ and $\Delta\Delta G_{solv}$ are in kcal mol^{-1} and K_i is in μM.

[b] Squared correlation coefficient of the regression.

[c] Standard error of the regression.

[d] Fisher F-test value of the regression.

[e] *Leave-one-out* cross-validated squared correlation coefficient of the regression.

[f] Rounded off.

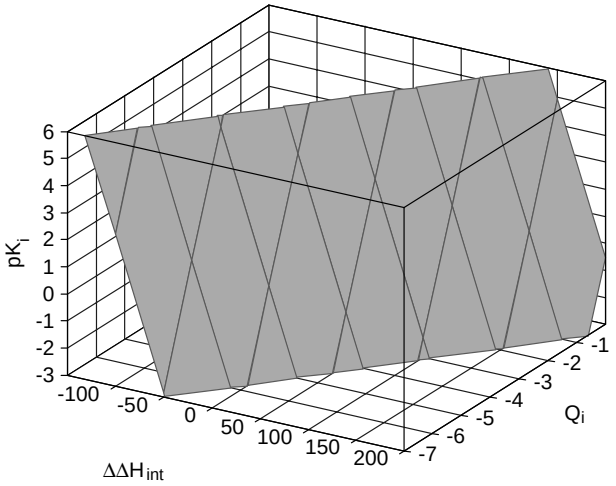

Fig. 2. QSAR regression model for 16 known peptidic inhibitors of the NS3/4A protease of HCV (Table 1) was obtained by correlating the experimental K_i^{exp} [23] with computed relative interaction contribution to the enzyme–inhibitor binding affinity $\Delta\Delta H_{int}$ and molecular charge of the inhibitor Q_i (which is proportional to the solvent effect contribution $\Delta\Delta G_{solv}$ to the relative enzyme–inhibitor binding affinity, Table 2) as: $pK_i = -0.109\Delta\Delta H_{int} + 4.226Q_i + 21.289$. The units used in the derivation of the regression coefficients: K_i (μM), $\Delta\Delta H_{int}$ (kcal mol^{-1}) and Q_i (e$^-$).

ranges occurring in the training set. Therefore, some degree of extrapolation is usually necessary in order to design derivatives with higher predicted activities. In our binding model, the QSAR equations were trained on a set of acidic peptides with molecular charge ranging from $-1e^-$ to $-5e^-$, i.e. spanning a charge window of $4e^-$. Extrapolation to more acidic peptide inhibitors with molecular charge of up to $-7e^-$, i.e. exceeding by $-2e^-$ or 50% the lower range limit of molecular charge of inhibitors considered in the training set is in our opinion still acceptable. Due to the involved extrapolation the predicted K_i^* constants of the designed inhibitor candidates may be considered as semi-quantitative.

3.2. Design of new potent inhibitors

For tight binding of peptidic inhibitors to the NS3/4A two binding *anchors* are important: the P_1, P_2 electrophile positioned near the catalytic site and the acidic residues in P_5, P_6 positions [15,20]. Therefore, a dramatic decrease in the inhibitory potency of shorter less acidic peptides was observed after removing the acidic residues in P_5 and P_6 positions of known inhibitors (**I**1–**I**3 and **I**5–**I**7) (Table 1). Rational design strategy of new more active inhibitors of NS3/4A protease should therefore concern longer peptides (hexapeptides) with a strong electrophile in the P_1 position. It should be combined with residue optimisation guided by structure-based considerations involving the occupied S_6–S_1 specificity pockets of the NS3/4A binding site while keeping in mind that any residue replacement affects the solvent effect contribution to the enzyme–inhibitor binding. Namely, increased stabilisation of the enzyme–inhibitor complexes by stronger enzyme–inhibitor interactions (more negative $\Delta\Delta H_{int}$) for longer more acidic (highly charged) peptide sequences is to a large extent counterbalanced by enhanced destabilisation of the resulting complexes due to unfavourable solvent effect contribution (more positive $\Delta\Delta G_{solv}$) originating from buried polar surface of charged amino acid side chains in certain parts of the specificity pockets of the NS3 protease binding site.

The predictive QSAR model (Eq. (8), Fig. 2) suggests that highly acidic hexapeptides ($Q_i \in \langle -7\ ,\ -5\rangle\ e^-$), which interact strongly with the cationic binding site of the NS3/4A protease ($\Delta\Delta H_{int} \in \langle -140\ ,\ -40\rangle$ kcal mol^{-1}) can form tight binding inhibitors of the NS3/4A of HCV with estimated K_i^* in the picomolar range. According to the model, the stabilising effects of the negative interaction contribution to the relative binding affinity of highly charged hexapeptides can exceed the destabilising solvent effects upon the enzyme–inhibitor complex formation. Due to an increased charge complementarity to the basic binding site of the serine protease, the highly charged acidic peptides are expected to form inhibitors specific to the NS3/4A of HCV. There are altogether five basic amino acid residues (including three arginines) at the active site of the NS3/4A within 5 Å distance form the reference inhibitor **I**15 available for ion-pairs formation with the inhibitor. In fact, Koch et al. [41] observed that site directed mutagenesis of basic residues located in the vicinity of the protease active site led to changes of k_{cat} values of the substrate indicating that these residues play important role in the stabilisation of the transition state.

The results of the design of new peptidic inhibitors of the NS3/4A protease are summarised in Table 3. The first new derivative (**N**1) was based on a combination of sequences of hexapeptides with the lowest K_i values measured by Ingallinella et al. [15,20] and Johansson et al. [23] (Table 1): AcAsp–D-Gla–Dif–Glu–Cha–Cyo (**N**1) with the α-ketocysteine (Cyo) anchor in the P_1 position. Using the QSAR model derived for the training set of peptidic inhibitors of the NS3/4A, Eq. (8), we have predicted for the derivative **N**1 an inhibition constant K_i^* of 0.04 nM towards the NS3/4A protease, which suggests for **N**1 higher inhibitory potency then that of the most effective inhibitor **I**11 considered in the training set with the K_i of 0.75 nM [23] (Table 1). The computed interaction component of the binding affinity of **N**1 towards the NS3/4A exceed by almost -40 kcal mol^{-1} that of the reference inhibitor **I**15 (Table 1), which indicated that further structure-based modifications of the inhibitors may lead to a discovery of more potent and more specific derivatives inhibiting the NS3/4A serine protease of HCV. Therefore, we carried out independent structure-based optimisation of the individual residues in the P_6 to P_1 positions while considering the 3D-structure of the individual specificity pockets and comparing the predicted inhibition constants K_i^* derived from the computed $\Delta\Delta H_{int}$ and Q_i of the new derivatives.

Table 3
Sequences and computed properties of designed hexapeptide inhibitors of the NS3 protein of HCV with individually optimised residues in the P_6–P_1 positions and their predicted inhibition constants (K_i^*) towards the serine protease domain of the NS3/4A

Inhibitor[a]	P_6	P_5	P_4	P_3	P_2	P_1	M_w[b] (Da)	Q_i[c] (e^-)	$\Delta G_{solv}\{I\}$[d] (kcal mol^{-1})	$\Delta\Delta H_{int}$[e] (kcal mol^{-1})	K_i^{*}[f] (nM)
						P_1					
N1	AcAsp	D-Gla	Dif	Glu	Cha	**Cyo**	983	**−5**	−4.8	**−39.5**	**0.04**
N2	AcAsp	D-Gla	Dif	Glu	Cha	**Fab**	1001	**−5**	−8.2	**−42.8**	**0.02**
					P_2						
N3	AcAsp	D-Gla	Dif	Glu	**Asp**	Cyo	941	**−6**	−232.1	**−76.4**	**0.05**
N4	AcAsp	D-Gla	Dif	Glu	**Glu**	Cyo	955	**−6**	−222.8	**−75.4**	**0.06**
N5	AcAsp	D-Gla	Dif	Glu	*Val*	Cyo	925	−5	−12.6	−32.7	0.2
N6	AcAsp	D-Gla	Dif	Glu	*Leu*	Cyo	939	−5	−7.7	−35.8	0.1
N7	AcAsp	D-Gla	Dif	Glu	*Ile*	Cyo	939	−5	−13.8	−31.8	0.2
N8	AcAsp	D-Gla	Dif	Glu	Cha	*Cyo*	983	−5	−4.8	−39.5	0.05
N9	AcAsp	D-Gla	Dif	Glu	*Phg*	Cyo	959	−5	−14.6	−36.0	0.09
N10	AcAsp	D-Gla	Dif	Glu	*Phe*	Cyo	973	−5	−14.9	−40.4	0.04
N11	AcAsp	D-Gla	Dif	Glu	*Tyr*	Cyo	989	−5	−5.5	−41.1	0.03
N12	AcAsp	D-Gla	Dif	Glu	*Nal*	Cyo	1009	−5	−5.7	−30.9	0.3
N13	AcAsp	D-Gla	Dif	Glu	*Trp*	Cyo	1012	−5	−7.3	−40.9	0.03
N14	AcAsp	D-Gla	Dif	Glu	*Dif*	Cyo	1049	−5	−7.2	−40.7	0.04
				P_3							
N15	AcAsp	D-Gla	Dif	**Asp**	Cha	Cyo	965	**−5**	−11.7	**−41.9**	**0.02**
N16	AcAsp	D-Gla	Dif	**Glu**	Cha	*Cyo*	983	**−5**	−4.8	**−39.5**	**0.04**
N17	AcAsp	D-Gla	Dif	*Leu*	Cha	Cyo	964	−4	172.0	−0.7	0.04
N18	AcAsp	D-Gla	Dif	*tBu*	Cha	Cyo	964	−4	171.2	1.7	0.06
			P_4								
N19	AcAsp	D-Gla	*Dif*	Glu	Cha	*Cyo*	983	−5	−4.8	−39.5	0.04
N20	AcAsp	D-Gla	**Trc**	Glu	Cha	Cyo	986	**−6**	−259.9	**−81.4**	**0.02**
N21	AcAsp	D-Gla	*Val*	Glu	Cha	Cyo	855	−5	−20.4	−14.3	19
N22	AcAsp	D-Gla	*tBu*	Glu	Cha	Cyo	869	−5	−24.2	−14.2	20
N23	AcAsp	D-Gla	*Phg*	Glu	Cha	Cyo	889	−5	−29.7	−25.7	1
N24	AcAsp	D-Gla	*Phe*	Glu	Cha	Cyo	903	−5	−14.5	−35.8	0.09
N25	AcAsp	D-Gla	*Cha*	Glu	Cha	Cyo	909	−5	−25.5	−31.6	0.3
N26	AcAsp	D-Gla	*Tyr*	Glu	Cha	Cyo	919	−5	−11.5	−38.2	0.05
N27	AcAsp	D-Gla	Nap	Glu	Cha	Cyo	939	−5	−20.2	−36.9	0.07
N28	AcAsp	D-Gla	*Nal*	Glu	Cha	Cyo	939	−5	−19.2	−34.5	0.1
N29	AcAsp	D-Gla	*Trp*	Glu	Cha	Cyo	942	−5	−19.1	−41.3	0.02
N30	AcAsp	D-Gla	*Tro*	Glu	Cha	Cyo	958	−5	−8.4	−18.2	7
		P_5									
N31	AcAsp	D-**Gla**	Dif	Glu	Cha	*Cyo*	983	**−5**	−4.8	**−39.5**	**0.04**
N32	AcAsp	D-*Cpa*	Dif	Glu	Cha	Cyo	933	−5	−7.2	0.9	870
N33	AcAsp	D-*Asa*	Dif	Glu	Cha	Cyo	965	−5	−25.3	−14.5	18
N34	AcAsp	D-*Asp*	Dif	Glu	Cha	Cyo	922	−4	201.0	50.4	12700
N35	AcAsp	D-*Val*	Dif	Glu	Cha	Cyo	907	−3	364.2	91.4	22000
N36	AcAsp	*Phg*	Dif	Glu	Cha	Cyo	941	−3	357.4	99.3	160000
N37	AcAsp	D-*Phg*	Dif	Glu	Cha	Cyo	941	−3	362.0	82.3	2200
N38	AcAsp	*Phe*	Dif	Glu	Cha	Cyo	955	−3	364.6	98.3	125000
N39	AcAsp	D-*Phe*	Dif	Glu	Cha	Cyo	955	−3	362.4	90.4	17000
N40	AcAsp	*Cha*	Dif	Glu	Cha	Cyo	961	−3	367.6	97.8	109000
N41	AcAsp	D-*Cha*	Dif	Glu	Cha	Cyo	961	−3	366.6	101	244000
N42	AcAsp	*Tyr*	Dif	Glu	Cha	Cyo	971	−3	354.5	98.9	145000
N43	AcAsp	D-*Tyr*	Dif	Glu	Cha	Cyo	971	−3	362.7	95.7	65000
N44	AcAsp	*Trp*	Dif	Glu	Cha	Cyo	994	−3	361.5	102.2	330000
N45	AcAsp	D-*Trp*	Dif	Glu	Cha	Cyo	994	−3	363.1	80.3	1400
N46	AcAsp	D-*Dif*	Dif	Glu	Cha	Cyo	1031	−3	369.3	88.5	11000
	P_6										
N47	**AcAsp**	D-Gla	Dif	Glu	Cha	*Cyo*	983	**−5**	−4.8	**−39.5**	**0.04**
N48	***Glr***	D-Gla	Dif	Glu	Cha	Cyo	922	−5	−25.4	−22.4	2
N49	*Suc*	D-Gla	Dif	Glu	Cha	Cyo	936	−5	−8.6	−14.2	20
N50	*AcGlu*	D-Gla	Dif	Glu	Cha	Cyo	993	−5	−2.8	−15.5	14

[a] N-terminal residue of the peptide inhibitor is at the P_6 position [12], for acronyms used for unusual amino acids see *Abbreviations*.
[b] Molecular weight of the inhibitor.
[c] Molecular charge of the inhibitor. Total molecular charge of the serine protease domain of the NS3/4A protein–cofactor complex is equal to $7e^-$.
[d] Relative solvation Gibbs free energy of the inhibitors taken with respect to the best binding inhibitor **I**15 of the training set (Table 1).
[e] Relative enzyme–inhibitor interaction energies taken with respect to the interaction energy of the best binding training set inhibitor **I**15 (Table 1).
[f] Inhibition constants of the designed peptidic inhibitors towards the NS3/4A serine protease domain were predicted using the QSAR correlation Eq. (d) in Table 2.

3.2.1. Optimisation of P_1 residue

The carbon atom of the activated carbonyl group of the α-ketoacid in P_1 position becomes the hemiketal quaternary carbon upon binding to the γ-O of the catalytic Ser:139 [11] at the enzyme's active site and thus contributes significantly to the inhibitor binding. The carbonyl group itself is oriented towards the His:57 and is solvent exposed while the carboxyl group of the α-ketoacid forms hydrogen bonds with the oxyanion hole amide groups of Gly:137 and Ser:139 [30]. The S_1 specificity pocket of the NS3/4A protease is lined mainly by the side chains of non-polar residues Val:132, Leu:135 and Phe:154 [30] and displays strong preference for Cys residue in the P_1 position of substrates [13,16,17,42].

From literature, it is known that α-ketoacids mimicking α-ketocysteine such as α-keto-δ,δ-difluoroaminobutyric acid (Fab) are very effective in the P_1 position of potent inhibitors of NS3/4A [25,30]. Thus, the predicted inhibition constants K_i^* of the two inhibitors containing Cyo (**N**1) and Fab (**N**2) residues in P_1 were almost identical with only a minor preference for the Fab residue (Table 3).

3.2.2. Optimisation of P_2 residue

The S_2 is a solvent accessible pocket formed by the catalytic residues His:57 and Asp:81 and the basic Arg:155. Non-polar residues were effective at the P_2 position in the training set of known NS3/4A inhibitors (Table 1) as they are thought to protect and shield the hydrogen bond between the carboxylate group of Asp:81 and the δ-NH of the His:57 from the solvent [43]. In fact, inhibitors **I**12 with β-cyclohexylalanine (Cha) and **I**14 with β-napthylalanine (Nal) in the P_2 position showed somewhat lower K_i than **I**15 with Glu in the P_2 [23].

Several substitutions of the original Cha residue in the **N**1 by hydrophobic residues such as Val, Leu and Ile (**N**5–**N**7), aromatic residues as Phg, Phe, Tyr, Nal, Trp and Dif (**N**8–**N**14) and acidic residues as Asp and Glu (**N**3, **N**4), were assessed (Table 3). Higher K_i^* were predicted for the aliphatic residues and no significant improvements in the inhibitory potency were estimated for the aromatic resides. On the other hand, presence of acidic residues in the P_2 position led to an increased interaction stabilisation of the complexes of the NS3/4A protease with the derivatives **N**3 and **N**4 by about $-35\,\mathrm{kcal\,mol^{-1}}$ mainly due to additional hydrogen bonding with the amine group of nearby Gln:41 residue. The enhanced interaction was however counterbalanced by the destabilising solvent effects arising from the higher negative molecular charge of the derivatives. Among several derivatives from the P_2 series with similar predicted K_i^* the candidates **N**3 and **N**4 with the highest interaction stabilisation were preferred because of they are expected to increased the specificity of inhibitor binding to the NS3/4A of HCV compared to other serine proteases.

3.2.3. Optimisation of P_3 residue

The backbone amino and carbonyl groups of the P_3 residue and of the residue Ala:157 form an intermolecular hydrogen bond pattern typical for an anti-parallel β-sheet. The S_3 pocket is lined by non-polar residues as Val:132, Ala:157 and Cys:159, which suggests that non-polar residues can be useful in the P_3 position of the inhibitors. However, the P_3 side chain is solvent exposed and thus may not contribute significantly to the enzyme–inhibitor interaction as observed in a number of known serine protease-inhibitor complexes [11,44]. Lys:136 is positioned close to the S_3 pocket, therefore, also negatively charged residues can fit into the P_3 position. This could be the reason why the Glu residue at the P_3 position of the inhibitor **I**1 showed slightly lower K_i value than the **I**5 with Ile at P_3 [23] (Table 1).

We replaced the acidic Glu residue in the P_3 position of **N**1 by hydrophobic residues such as Leu (**N**17) and *t*Bu (**N**18), which resulted in slightly higher predicted K_i^*. Replacement of Glu by another acidic residue, Asp, confirmed the preference of the NS3/4A for anionic residues in the P_3 position, since this substitution in the derivative **N**15 was predicted to lower the K_i^* to 0.02 nM. This was caused by better interaction stabilisation of the **N**15 by $-2.4\,\mathrm{kcal\,mol^{-1}}$ (Table 3) as compared to the **N**1 due to more favourable contact of Asp with the Lys:136.

3.2.4. Optimisation of P_4 residue

Experimental observation [30] revealed that the small shallow solvent exposed S_4 pocket is formed by the residues Val:158, Ala:156 and in part also by the side chains of Arg:123, Arg155 and Asp:168. From three similar known hexapeptide inhibitors **I**1, **I**10 and **I**12 (Table 1) the lowest K_i value was determined for the **I**1 derivative with the bulkiest β,β-diphenylalanine (Dif) residue at the P_4 position, which probably shielded the intramolecular hydrogen bonding and electrostatic interaction of the neighbouring residues Arg:123–Asp:168–Arg:155 from the solvent.

We tested hydrophobic aliphatic residues such as Val and *t*Bu (**N**21, **N**22), aliphatic cyclic residues as Cha (**N**25) and various aromatic systems such as Dif, Phg, Phe, Tyr, Nap, Nal, Trp and Tro (**N**19–**N**30) in the P_4 position of the NS3/4A inhibitor candidates (Table 3) with no significant predicted improvement in the inhibition potency. Some modifications of the bulky aromatic natural amino acids by attachment of additional function groups such as 7-hydroxyl (Tro, **N**30) or 4-carboxy (Trc, **N**20) to tryptophan residues derived from the structure-based considerations involving the Arg:155 residue of the S_4 specificity pocket, were also proposed. In particular, addition of the 4-COO^- group to Trp was predicted to cause a significant decrease in the interaction term by almost $-42\,\mathrm{kcal\,mol^{-1}}$, making the **N**20 derivative with K_i^* of 0.02 nM one of the most specific tight binding NS3/4A inhibitor candidates of all the P_4 optimised derivatives.

3.2.5. Optimisation of P_5 residue

The anionic P_5–P_6 anchor was found essential for tight binding of inhibitors to the NS3/4A [15,20] due to the

structure of the S_5 subsite, which contains polar and cationic residues such as Thr:160, Arg:161 and Lys:165. It is therefore not surprising that introduction of dicarboxylic acid as D-γ-carboxyglutamic acid (D-Gla) into the P_5 position in the inhibitor **I**11 improved the observed K_i by almost two orders of magnitude compared to the derivative **I**5, which contains Glu residue in the P_5 position (Table 1). The D-stereoisomer of Gla can acquire more favourable contacts with Arg:161 and Lys:165 than its L-form.

We truncated and prolonged the side chain of the D-Gla residue by one methylene group (**N**32, **N**33), however, in both cases, a significant increase in the K_i^* was predicted (Table 3) due to the different mode of binding of the anionic residues D-Gla, D-Cpa and D-Asa to Arg:161 and Lys:165 of the S_5 pocket. Elimination of one carboxylic group by replacing the D-Gla with D-Asp (**N**34) increased the predicted K_i^* by several orders of magnitude. We also substituted the D-Gla residue by both L- and D-forms of aromatic, cyclic and non-cyclic aliphatic residues such as Phg, Phe, Tyr, Trp, Cha and Val (**N**35–**N**46). However, a strong decrease in the inhibition potency by five to eight orders of magnitude was predicted for all such substitutions (Table 3).

3.2.6. Optimisation of P_6 residue

Hindering of the N-terminal protonation by addition of acetyl group increases the stability of the enzyme–inhibitor complex. The required anionic P_5–P_6 anchor of inhibitors is also related also to the hydrogen bonding interactions of the P_6 residue with Arg:123 present in the S_6 pocket of NS3/4A.

We have replaced the acetylated N-terminal aspartic acid in the P_6 position of **N**47 gradually by glutaric acid (Glr, **N**48), succinic acid (Suc, **N**49) and acetylated glutamic acid (AcGlu, **N**50). These replacements led to an up to three orders of magnitude large increase in the predicted K_i^* of these derivatives towards the NS3/4A due to diminished interactions between the side chain of P_6 residue and Arg:123 and between the acetyl group of the P_6 residue and Lys:165 in the S_6 subsite.

3.3. Combinatorial optimisation of inhibitors

From the set of 50 designed inhibitor candidates with individually optimised residues in the P_6 to P_1 positions (Table 3), we selected a set of nine best residues (*building blocks*) useful in specific positions of hexapeptide sequences derived from **N**1, which are predicted to compose potent inhibitors of the NS3/4A protease of HCV. This set of building blocks was used in a highly focused combinatorial *mini*-library of the size $\{P_6 \times P_5 \times P_4 \times P_3 \times P_2 \times P_1\} =$ {AcAsp × D-*Gla* × Trc × (Asp, Glu) × (Asp, Glu) × (Cyo, Fab)} = eight hexapeptide inhibitors. The inhibition constants of the combinatorial analogues towards the NS3/4A protease were predicted using the QSAR model (or virtual screening), which relies on the computed enzyme–inhibitor interaction component of the binding affinity of analogues **C**1 to **C**8 towards the NS3/4A and the estimated solvent effects that are proportional to the molecular charge Q_i of the analogues (Table 4). The predicted K_i^* towards the NS3/4A protease of all the **C** analogues lie in low picomolar range up to two orders of magnitude below the K_i^* of the individually optimised **N** derivatives (Table 3) and about three orders of magnitude below the K_i^{exp} of the best binding training set hexapeptide inhibitor considered, **I**11 (Table 1).

The combinatorial analogues represent highly anionic hexapeptides with the net charge of $-7e^-$ composed of acidic residues in the P_2 to P_6 positions and the charged C-terminal α-ketoacid and are predicted to be highly water-soluble (low negative solvation Gibbs free energies, $\Delta G_{solv}\{I\}$, Table 4). The overall preference for the anionic residues in effective inhibitors of NS3/4A is not surprising since there are five basic residues (mainly arginines) of

Table 4
Sequences and computed properties of the $\{P_6 \times P_5 \times P_4 \times P_3 \times P_2 \times P_1\} = \{1 \times 1 \times 1 \times 2 \times 2 \times 2\}$ combinatorial *mini*-library of hexapeptide inhibitors of the NS3 protein of HCV highly focused by structure-based combinatorial residue optimisation and predicted inhibition constants (K_i^*) towards the serine protease domain of the NS3/4A

Inhibitor[a]	P_6	P_5	P_4	P_3	P_2	P_1	M_w[b] (Da)	Q_i[c] (e^-)	$\Delta G_{solv}\{I\}$[d] (kcal mol^{-1})	$\Delta\Delta H_{int}$[e] (kcal mol^{-1})	K_i^*[f] (pM)
C1	AcAsp	D-Gla	Trc	Asp	Asp	Cyo	933	−7	−524.1	−131.6	0.9
C2	AcAsp	D-Gla	Trc	Asp	Asp	Fab	951	−7	−522.4	−135.7	0.3
C3	AcAsp	D-Gla	Trc	Asp	Glu	Cyo	947	−7	−528.7	−131.5	0.9
C4	AcAsp	D-Gla	Trc	Asp	Glu	Fab	965	−7	−510.7	−130.3	1.3
C5	AcAsp	D-Gla	Trc	Glu	Asp	Cyo	947	−7	−517.3	−127.2	2.7
C6	AcAsp	D-Gla	Trc	Glu	Asp	Fab	965	−7	−513.7	−131.4	1.0
C7	AcAsp	D-Gla	Trc	Glu	Glu	Cyo	961	−7	−505.1	−122.4	9.2
C8	AcAsp	D-Gla	Trc	Glu	Glu	Fab	979	−7	−501.7	−133.8	0.5

[a] N-terminal residue of the peptide inhibitor is at the P_6 position [12], for acronyms used for unusual amino acids see *Abbreviations*.
[b] Molecular weight of the inhibitor.
[c] Molecular charge of the inhibitor. Total molecular charge of the serine protease domain of the NS3/4A protein–cofactor complex is equal to $7e^-$.
[d] Relative solvation Gibbs free energy of the inhibitors taken with respect to the best binding inhibitor **I**15 of the training set (Table 1).
[e] Relative enzyme–inhibitor interaction energies taken with respect to the interaction energy of the best binding training set inhibitor **I**15 (Table 1).
[f] Inhibition constants of the designed peptidic inhibitors towards the NS3/4A serine protease domain were predicted using the QSAR correlation Eq. (d) in Table 2.

the enzyme located within 5 Å distance from the bound inhibitors. Within the same distance, only three acidic residues are positioned, chiefly forming ion pair bridges with the basic residues. The presence of five negatively charged residues in the inhibitor candidates in combination with the α-ketoacid in the P_1 position lead to significant improvement of the interaction stabilisation of the enzyme–inhibitor complex due to attractive electrostatic interactions with the cationic binding site of the enzyme, which enhances the specificity of the designed inhibitors towards the NS3 serine protease of HCV. This interaction stabilisation of the complexes is to a large extent counterbalanced by the solvent effects contribution, Eqs. (7) and (8) (Fig. 2), however, the attractive enzyme–inhibitor interactions are strong enough to lead to predicted high inhibitory potencies of the **C**1 to **C**8 analogues.

Of course, the actual inhibitory potencies of the designed HCV inhibitors have to be determined experimentally, however, as a result of our structure-based design and combinatorial optimisation approach the number of hexapeptides to be synthesised and screened can be reduced to a smaller number of hexapeptides containing acidic residues (Fig. 3).

Recently, Johansson et al. [23] have shown that inhibitory potencies of peptides towards the native bifunctional full-length NS3 protease-helicase/NTPase differ from the reported data determined only from the corresponding NS3 serine protease domain assay. Therefore, our results, which were derived from modelling of the NS3 serine protease domain with the NS4A cofactor, can be further improved once the crystal structure of the full-length enzyme–inhibitor complex is resolved. Another direction for improvement

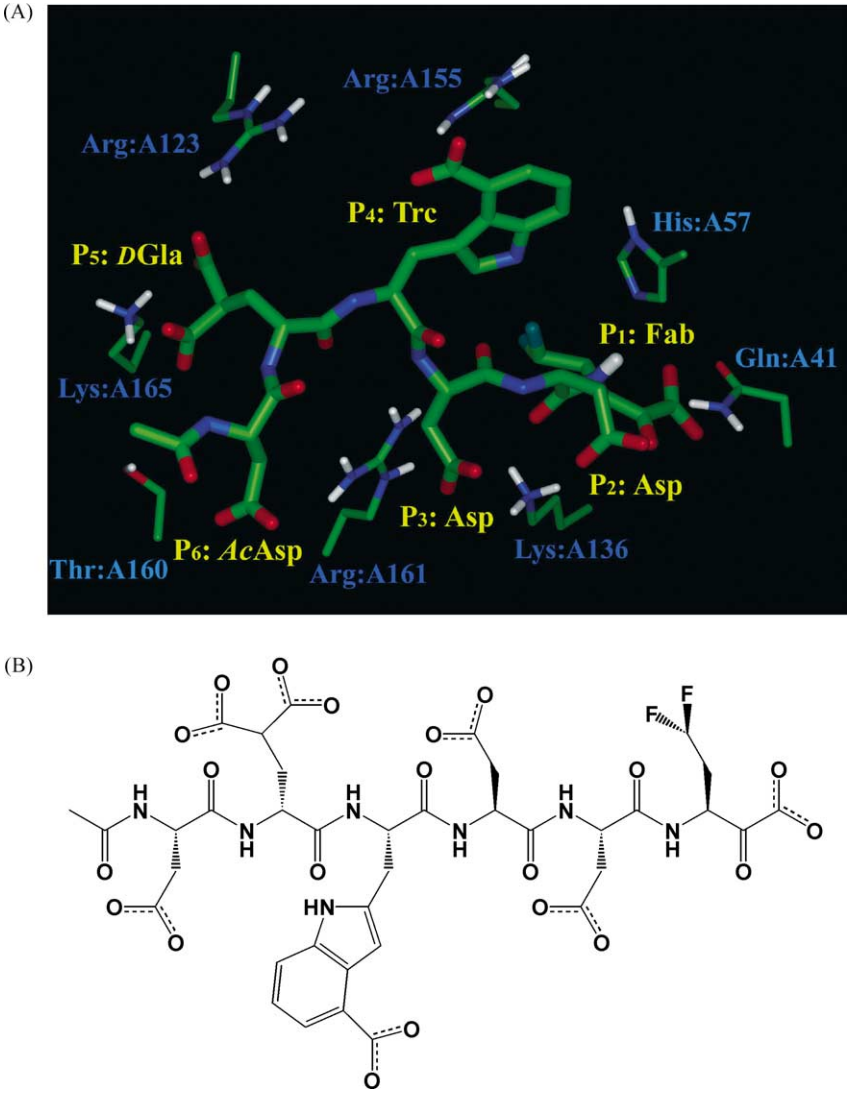

Fig. 3. (A) The anionic **C**2 inhibitor AcAsp–D-Gla–Trc–Asp–Asp–Fab (Table 4) designed by combinatorial optimisation with the lowest predicted inhibition constant of the of NS3/4A protease of HCV in sub-picomolar range shown in stick representation forms a number of ion-pairs with the cationic residues lining the active site of the serine protease domain (non-essential hydrogen atoms were omitted for better clarity). (B) Chemical structure of the **C**2 inhibitor.

of the substrate-based inhibitors lies in structure-based design of pseudo-petidic inhibitors that will span both the S and S′ subsites of the NS3 protease binding site with a non-cleavable bond positioned in the protease active site. Design and optimisation of such P, P′-peptidic inhibitors of the NS3 protease described previously by Ingallinella et al. [20,45] resulted in an increase in the inhibitors binding by more than three orders of magnitude. These observations open the possibility to design potent pseudo-peptides with lower molecular weight, more favourable $\log P_{o/w}$ and higher oral bioavailability.

References

[1] M. Houghton, Hepatitis C viruses, in: B.N. Fields, D.M. Knipe, P.M. Howley (Eds.), Fields' Virology, third ed., Lippincott-Raven, Philadelphia, PA, 1996, pp. 1035–1058.

[2] J. Avital, C. Hepatitis, Curr. Opin. Infect. Dis. 11 (1998) 293–299.

[3] J.H. Hoofnagle, A.M. di Bisceglie, The treatment of chronic viral hepatitis, N. Engl. J. Med. 336 (1997) 347–356.

[4] J.H. Hoofnagle, Therapy for acute hepatitis C, N. Engl. J. Med. 345 (2001) 1495–1497.

[5] A. Grakoui, D.W. McCourt, C. Wychowski, S.M. Feinstone, C.M. Rice, Characterization of the hepatitis C virus-encoded serine protease: determination of proteinase-dependent polyprotein cleavage sites, J. Virol. 67 (1993) 2832–2843.

[6] A. Lesk, W.D. Fordham, Conservation and variability in the structures of serine proteinases of the chymotrypsin family, J. Mol. Biol. 258 (1996) 501–537.

[7] P. Neddermann, L. Tomei, C. Steinkühler, P. Gallinari, A. Tramontano, R. De Francesco, The non-structural proteins of the hepatitis C virus: structure and function, Biol. Chem. 378 (1997) 469–476.

[8] R.A. Love, H.E. Parge, J.A. Wickersham, Z. Hostomsky, N. Habuka, E.W. Moomaw, T. Adachi, Z. Hostomska, The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site, Cell 87 (1996) 331–342.

[9] J.L. Kim, K.A. Morgenstern, C. Lin, T. Fox, M.D. Dwyer, J.A. Landro, S.P. Chambers, W. Markland, C.A. Lepre, E.T. O'Malley, S.L. Harbeson, C.M. Rice, M.A. Murcko, P.R. Caron, J.A. Thomson, Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide, Cell 87 (1996) 343–355. PDB entry code 1A1R.

[10] D.O. Cicero, G. Barbato, U. Koch, P. Ingallinella, E. Bianchi, M.C. Nardi, C. Steinkühler, R. Cortese, V.G. Matassa, R. De Francesco, A. Pessi, R. Bazzo, Structural characterization of the interactions of optimised product inhibitors with the N-terminal proteinase domain of the hepatitis C virus (HCV) NS3 protein by NMR and modelling studies, J. Mol. Biol. 289 (1999) 385–396.

[11] G. Barbato, D.O. Cicero, F. Cordier, F. Narjes, B. Gerlach, S. Sambucini, S. Grzesiek, V.G. Matassa, R. De Francesco, R. Bazzo, Inhibitor binding induces active site stabilisation of the HCV NS3 protein serine protease domain, EMBO J. 19 (2000) 1195–1206.

[12] The nomenclature designates the substrate cleavage sites P_6–P_5–P_4–P_3–P_2–$P_1 \cdots P_{1'}$–$P_{2'}$–$P_{3'}$–$P_{4'}$, etc. with the scissile bond between P_1 and $P_{1'}$ and the C-terminus of the substrate on the prime site. The corresponding binding sites on the protease are indicated as S_6–S_5–S_4–S_3–S_2–$S_1 \cdots S_{1'}$–$S_{2'}$–$S_{3'}$–$S_{4'}$, etc., according to: I. Schecter, A. Berger, On the size of the active site in proteases, Biochem. Biophys. Res. Commun. 27 (1967) 157–162.

[13] R. Zhang, J. Durkin, W.T. Windsor, C. McNemar, L. Ramanathan, H.V. Le, Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides, J. Virol. 71 (1997) 6208–6213.

[14] D. Fattori, A. Urbani, M. Brunetti, R. Ingenito, A. Passi, K. Prendergast, F. Narjes, V.G. Matassa, R. De Francesco, C. Steinkühler, Probing the active site of the hepatitis C virus serine protease by fluorescence resonance energy transfer, J. Biol. Chem. 275 (2000) 15106–15113.

[15] P. Ingallinella, S. Altamura, E. Bianci, M. Taliani, R. Ingenito, R. Cortese, R. De Francesco, C. Steinkühler, A. Pessi, Potent peptide inhibitors of human hepatitis C virus protease are obtained by optimising the cleavage products, Biochemistry 37 (1998) 8906–8914.

[16] C. Steinkühler, A. Urbani, L. Tomei, G. Biasol, M. Sardana, E. Bianchi, A. Pessi, R. De Francesco, Activity of purified hepatitis C virus protease NS3 on peptide substrates, J. Virol. 70 (1996) 6694–6700.

[17] A. Urbani, E. Bianchi, F. Narjes, A. Tramontano, R. De Francesco, C. Steinkühler, A. Pessi, Substrate specificity of the hepatitis C virus serine protease NS3, J. Biol. Chem. 272 (1997) 9204–9209.

[18] S.R. LaPlante, D.R. Cameron, N. Aubry, S. Lefebvre, G. Kukolj, R. Maurice, D. Thibeault, D. Lamarre, M. Llinàs-Brunet, Solution structure of substrate-based ligands when bound to hepatitis C virus NS3 protease domain, J. Biol. Chem. 274 (1999) 18618–18624.

[19] C. Steinkühler, G. Biasiol, M. Brunetti, A. Urbani, U. Koch, R. Cortese, A. Passi, R. De Francesco, Product inhibition of the hepatitis C virus NS3 protease, Biochemistry 37 (1998) 8899–8905.

[20] P. Ingallinella, E. Bianci, R. Ingenito, U. Koch, C. Steinkühler, S. Altamura, A. Pessi, Optimisation of the P′-region of peptide inhibitors of Hepatitis C virus NS3/4A protease, Biochemistry 39 (2000) 12898–12906.

[21] M. Llinàs-Brunet, M. Bailey, R. Deziel, G. Fazal, V. Gorys, S. Goulet, T. Halmos, R. Maurice, M. Poirier, M.-A. Poupart, J. Rancourt, D. Thibeault, D. Wernic, D. Lamarre, Studies on the C-terminal of hexapeptide inhibitors of the hepatitis C virus serine protease, Bioorg. Med. Chem. Lett. 8 (1998) 2719–2724.

[22] M. Llinàs-Brunet, M. Bailey, G. Fazal, E. Ghiro, V. Gorys, S. Goulet, T. Halmos, R. Maurice, M. Poirier, M.-A. Poupart, J. Rancourt, D. Thibeault, D. Wernic, D. Lamarre, Highly potent and selective peptide-based inhibitors of the hepatitis C virus serine protease: towards smaller inhibitors, Bioorg. Med. Chem. Lett. 10 (2000) 2267–2270.

[23] A. Johansson, I. Hubatsch, E. Åkerblom, G. Lindeberg, S. Winiwarter, U.H. Danielson, A. Hallberg, Inhibition of hepatitis C virus NS3 protease activity by product-based peptides is dependent on helicase domain, Bioorg. Med. Chem. Lett. 11 (2001) 203–206.

[24] R.M. Dundson, J.R. Greening, P.S. Jones, S. Jordan, F.X. Wilson, Solidphase synthesis of aminoboronic acids: potent inhibitors of the hepatitis C virus NS3 protease, Bioorg. Med. Chem. Lett. 10 (2000) 1577–1579.

[25] F. Narjes, M. Brunetti, S. Coalrusso, B. Gerlach, U. Koch, G. Biasiol, D. Fattori, R. De Francesco, V.G. Matassa, C. Steinkühler, Alpha-ketoacids are potent slow binding inhibitors of the hepatitis C virus NS3 protease, Biochemistry 39 (2000) 1849–1861.

[26] C.S. Cassidy, J. Lin, P.A. Frey, A new concept for the mechanism of action of chymotrypsin: the role of the low-barrier hydrogen bond, Biochemistry 36 (1997) 4576–4584.

[27] K. Sudo, Y. Matsumoto, M. Matsushima, M. Fujiwara, K. Konno, K. Shimotohno, S. Shigeta, T. Yokota, Novel hepatitis C virus protease inhibitors: thiazolidine derivatives, Biochem. Biophys. Res. Commun. 18 (1997) 643–647.

[28] W.T. Sing, C.L. Lee, S.L. Yeo, S.P. Lim, M.M. Sim, Arylalkylidene rhodamine with bulky and hydrophobic functional groups as selective HCV NS3 protease inhibitor, Bioorg. Med. Chem. Lett. 11 (2001) 91–94.

[29] K.-S. Yeung, N.A. Meanwell, Z. Qiu, D. Hernandez, S. Zhang, F. McPhee, S. Weiheimer, J.M. Clark, J.W. Janc, Structure–activity relationship studies of a bisbenzimidazole-based, Zn^{2+}-dependent inhibitor of HCV NS3 serine protease, Bioorg. Med. Chem. Lett. 11 (2001) 2355–2359.

[30] S. Di Marco, M. Rizzi, C. Volpari, M.A. Walsh, F. Narjes, S. Colarusso, R. De Francesco, V.G. Matassa, M. Sollazzo, Inhibition of the hepatitis C virus NS3/4A protease. The crystal structures of two protease-inhibitor complexes, J. Biol. Chem. 275 (2000) 7152–7157 (PDB entry codes 1DY8 and 1DY9).
[31] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242 (http://www.pdb.org/).
[32] Insight II 2000 and Discover 2.98 Molecular Modelling Software, 2000, Accelrys Inc., San Diego, CA.
[33] J.R. Maple, M.-J. Hwang, T.P. Stockfish, U. Dinur, M. Waldman, C.S. Ewing, A.T. Hagler, Derivation of class II force fields. 1. Methodology and quantum force field for the alkyl functional group and alkane molecules, J. Comput. Chem. 15 (1994) 162–182.
[34] V. Frecer, S. Miertus, Interactions of ligands with macromolecules: rational design of specific inhibitors of aspartic protease of HIV-1, Macromol. Chem. Phys. 203 (2002) 1650–1657.
[35] D. Bashford, D.A. Case, Generalized Born models of macromolecular solvation effects, Annu. Rev. Phys. Chem. 51 (2000) 129–152.
[36] J. Tomasi, M. Persico, Molecular interactions in solutions: an overview of methods based on continuous distribution of the solvent, Chem. Rev. 94 (1994) 2027–2094.
[37] S. Miertus, E. Scrocco, J. Tomasi, Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects, Chem. Phys. 55 (1981) 117–129.
[38] V. Frecer, S. Miertus, Polarizable continuum model of solvation for biopolymers, Int. J. Quant. Chem. 42 (1992) 1449–1468.
[39] B.T. Thole, Molecular polarizabilities calculated with a modified dipole interaction, Chem. Phys. 59 (1981) 341–350.
[40] V. Frecer, B. Ho, J.L. Ding, Interpretation of biological activity data of bacterial endotoxins by simple molecular models of mechanism of action, Eur. J. Biochem. 267 (2000) 837–852.
[41] U. Koch, G. Biasol, M. Brunetti, D. Fattori, M. Pallaoro, C. Steinkühler, Role of charged residues in the catalytic mechanism of hepatitis C virus NS3 protease: electrostatic precollision guidance and transition-state stabilisation, Biochemistry 40 (2001) 631–640.
[42] S.K. Burley, G.A. Petsko, Weakly polar interactions in proteins, Adv. Protein. Chem. 39 (1988) 125–189.
[43] J.L. Markley, W. Westler, Protonation state dependence of hydrogen bond strengths and exchange rates in a serine protease catalytic triad: bovine chymotrypsinogen, Biochemistry 35 (1996) 11092–11097.
[44] M.A. Navia, B.M. McKeever, J.P. Springer, T.Y. Lin, H.R. Williams, E.M. Fluder, C.P. Dorn, K. Hoogsteen, Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84 Å resolution, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 7–12.
[45] P. Ingallinella, D. Fattori, S. Altamura, C. Steinkühler, U. Koch, D.O. Cicero, R. Bazzo, R. Cortese, E. Bianchi, A. Pessi, Prime site binding inhibitors of serine protease: NS3/4A of hepatitis C virus, Biochemistry 41 (2002) 5483–5492.