

A Minimalistic 3D Pharmacophore Model for

Cyclopentapeptide CXCR4 Antagonists

Jon Våbenø, Gregory V. Nikiforovich, and Garland R. Marshall

Correspondence to: Garland R. Marshall; e-mail: garland@pcg.wustl.edu

Center for Computational Biology, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri

Abstract: Due to its involvement in HIV entry, the chemokine receptor CXCR4 is an attractive target for anti-retroviral drugs. Despite the large number of CXCR4 inhibitors studied, the 3D pharmacophore for binding to CXCR4 remains elusive, mainly as a result of conformational flexibility inherent in the identified ligands. In the present study, an exhaustive systematic exploration of the conformational space for a series of analogs of FC131, a cyclopentapeptide CXCR4 antagonist, has been performed. By comparing the resulting low-energy conformations using different sets of atoms, specific conformational features common only to the high/medium affinity compounds were identified. These features included the spatial arrangement of three pharmacophoric side chains as well as the orientation of a specific backbone amide bond. Together these features represent a minimalistic 3D pharmacophore model for binding of the cyclopentapeptide antagonists to CXCR4. The model enables rationalization of the experimental affinity data for this class of compounds as well as for the peptidomimetic KRH-1636.

A Minimalistic 3D Pharmacophore Model for Cyclopentapeptide CXCR4

Antagonists

Jon Våbenø, Gregory V. Nikiforovich, and Garland R. Marshall

Keywords: HIV; CXCR4 antagonist; cyclopentapeptide; FC131; pharmacophore

INTRODUCTION

The most commonly transmitted HIV-1 strains are the R5 (M-tropic) strains, which mainly target macrophages.^{1,2} These strains evolve to T-cell targeting (X4; T-tropic) strains, which are associated with a rapid decline in CD4⁺ T-cell levels and development of clinical AIDS.³⁻⁹ The G-protein coupled receptor (GPCR) CXCR4, originally identified as LESTR/fusin, has been shown to function as the co-receptor in the process of entry of T-tropic HIV-1 strains into CD4⁺ T-cells.^{10,11} This receptor therefore represents an interesting biological target for HIV therapeutics, specifically for CXCR4 antagonists, to prevent HIV-positive individuals from developing AIDS.

CXCR4 belongs to the chemokine receptor family, and its natural ligand is the 67-residue peptide CXCL12, also known as SDF-1.^{12,13} A large number of CXCR4 antagonists have been reported in the literature, one of the most potent being the cyclic pentapeptide FC131 (*c*(Gly¹-D-Tyr²-Arg³-Arg⁴-Nal⁵), Figure 1A; Nal is 2-naphthylalanine) which was discovered by an orthogonal design of two cyclopentapeptide (CPP) libraries combining the different permutations and stereoisomers of Arg, Arg, Nal, Tyr, and Gly.¹⁴ The library design was based on the identification of Arg², Nal³, Tyr⁵, and Arg¹⁴ as the most important residues for the anti-HIV activity of another peptide antagonist of CXCR4, the 14-residue T140 (Arg¹-Arg²-Nal³-*c*(Cys⁴-Tyr⁵-Arg⁶-Lys⁷-D-Lys⁸-Pro⁹-Tyr¹⁰-Arg¹¹-Cit¹²-Cys¹³)-Arg¹⁴).^{15,16} More recently, a series of Arg³ substituted FC131 analogs using neutral (Ala, D-Ala, NMe-Ala, D-NMe-Ala, Pro, and D-Pro), basic (Dab, Orn, Lys, guanidino-Dab, guanidino-Lys, and

cis/trans-4-guanidino-Pro), hydrophilic (Asn, Gln) and acidic (Glu) residues was described, demonstrating a subtle pattern for substitutions in the Xaa³ position of FC131.¹⁷ More conservative substitutions in the Xaa², Xaa⁴, and Xaa⁵ positions have also been performed.¹⁸ Moreover, the incorporation of amide bond isosteres and alternative cyclization strategies have been described for more drug-like FC131 analogs.^{17,19} However, FC131 still remains the CPP-based CXCR4 antagonist with highest affinity.

As for most other human GPCRs, the 3D structure of the CXCR4 receptor is unknown; therefore, ligand-based design is crucial for the development of CXCR4 antagonists. Since FC131 can be considered the lead compound for other peptide/peptidomimetic antagonists of CXCR4, it is desired to establish a 3D pharmacophore model for FC131 receptor interactions. SAR studies on FC131 analogs have shown that the presence of the four side chains (Tyr², Arg³, Arg⁴, Nal⁵) results in the compounds with highest affinity to CXCR4.^{17,18} However, substitution studies have shown that Arg³, as the only of these four residues, can be replaced by Ala (Ala³FC131, compound **1**, Table I) without a dramatic loss in affinity,^{17,18} i.e. Arg (or Arg-mimetics) in the Xaa³ position is not an absolute prerequisite for binding. For FC131, the guanidino group of Arg³ contributes with a positive charge and/or H-bond donor properties, having a favorable effect on affinity. However, substitution of Arg³ with Lys³ (containing an ϵ -amino group) results in a compound with lower affinity than the “neutral” Ala³ analog, which demonstrates the delicate nature of Xaa³ interaction with the receptor. The stereochemistry of the chiral residues in FC131 also plays a very important role: of the 16 possible stereoisomers of FC131, only three, namely FC131 itself, D-Arg³FC131 and D-Arg³-D-Nal⁵FC131 (compounds **9-11**, Table I), possess nanomolar affinities to CXCR4

(approximately corresponding to $EC_{50} < 1 \mu\text{M}$).¹⁴ In contrast, the retro-inverso (RI) analogs of **9-11** (compounds **12-14**, respectively, Table I, Figure 1B) showed low anti-HIV activity ($EC_{50} > 5 \mu\text{M}$), implying low affinity to CXCR4.²⁰ Interestingly, the three RI-analogs with highest activity ($EC_{50} < 5 \mu\text{M}$) were derived from parent FC131 stereoisomers with low activity ($EC_{50} > 5 \mu\text{M}$).²⁰

Based on these observations, our main focus has been the identification of a minimalistic 3D pharmacophore model for the CXCR4 antagonist FC131 involving only the three most important side chains, namely Tyr², Arg⁴, and Nal⁵ (the “three-point” model), i.e. ignoring Arg³. (Accordingly, the “four-point” model involves the Arg³ side chain as well.) In order to achieve this, we have performed extensive molecular mechanics-based investigations of 11 compounds of the type $c(\text{Gly}^1\text{-D-Tyr}^2\text{-L/D-Xaa}^3\text{-L-Arg}^4\text{-L/D-Nal}^5)$ (compounds **1-11**, Table I). By generating models of increasing complexity (taking different parts of the backbone into account), we were able to identify a set of critical conformational features that collectively constitute the desired 3D pharmacophore model. The modeling results for the retro-inverso compounds **12-15** were used for validation of the obtained model. It is our belief that this model will be of importance for the further development of CXCR4 antagonists suitable for anti-retroviral therapy.

METHODS

Calculations were run on a Dell Dimension 4600 workstation and/or on a Dell SC400 cluster. The Babel program²¹ was used for conversion between the PDB and MacroModel file formats, followed by manual editing when necessary.

Conformational Sampling and Energy Minimization

Due to the different nature (peptides and non-peptides) and conformational flexibility (number of rotatable bonds) of the compounds included in the study, different strategies for conformational sampling were used. The CPPs (**1-15**) were subjected to different systematic sampling procedures, whereas stochastic sampling was used for the non-peptide compound KRH-1636 (**16**).

Systematic Conformational Sampling for Compounds 1-8. In order to reduce the combinatorial complexity, systematic conformational sampling for **1-8** (Table I) was done in a sequential manner. Initially only the compound backbones were sampled, where Gly, L/D-Ala, and L/D-Pro replaced the actual residues. Only backbones satisfying the ring closure criterion were subjected to energy minimization and used as backbone templates for compounds **1-8**. The actual side chains were then added, followed by generation of all relevant side chain rotamers for each backbone.

Generation of Backbone Templates for Compounds 1-8. An analysis of experimentally determined CPP structures has shown that the compounds almost

exclusively adopt combinations of (ϕ, ψ) dihedral angles that are close to the allowed regions for linear peptides in the Ramachandran plot, as shown in figure 6 of ref. 22.

Hence, starting conformations for the peptide backbone were generated systematically by including all the six allowed combinations of (ϕ, ψ) dihedral angles for the non-Gly and non-Pro residues, namely $(-140^\circ, 140^\circ)$, $(-140^\circ, 80^\circ)$, $(-75^\circ, 140^\circ)$, $(-75^\circ, 80^\circ)$, $(-60^\circ, -60^\circ)$, and $(60^\circ, 60^\circ)$, i.e. the local minima β , pII' , pII , γ' , α_R and α_L in the Ramachandran plot were all covered. For Pro residues, three (ϕ, ψ) combinations were used: $(-75^\circ, 140^\circ)$, $(-75^\circ, 80^\circ)$, and $(-75^\circ, -60^\circ)$. For the residues with D-configuration, the same (ϕ, ψ) combinations with opposite signs of angles were used. For the Gly residue, the β , pII , γ' , α_R and α_L minima, and the minima symmetrical to β , pII , and γ' were used, giving a total of eight local minima. Only *trans* amide bonds ($\omega = 180^\circ$) were considered, except when Xaa^3 was L-/D-Pro or L-/D-NMe-Ala, where the possibility of *cis* orientation ($\omega = 0^\circ$) for the D-Tyr²-Xaa³ amide bond was included. Only the combinations of the ϕ , ψ , and ω angles resulting in a ring closure were selected as starting conformations for further energy minimization, i.e. conformations with a distance $\leq 4 \text{ \AA}$ between the C^α atom of Gly¹ and the dummy atom corresponding to the Me atom in the Nal⁵-NHMe construct. Redundant backbone conformations were subsequently removed. If one (or more) of the backbone torsion angles differed by more than 40° from the corresponding angle of a previous conformation, the backbone was considered unique (non-redundant).

Generation of Starting Conformations for Compounds 1-8. The appropriate side chains were added to the energy minimized unique backbones to give the CPP analogs (**1-8**). Starting conformations for compounds **1-8** were generated systematically for each backbone by including all the relevant side chain rotamers (i.e., 60° , -60° , and 180°) for

the following dihedral angles: χ_1 for Tyr²; χ_3 for 4-guanidino-Pro³; χ_1 , χ_2 , χ_3 , and χ_4 for Arg⁴; and χ_1 for Nal⁵. For the χ_2 angle of Nal⁵, the values of 90° and -90° were sampled.

The χ_2 and χ_3 angles of Tyr² were not considered in the sampling; their starting values were 90° and 180°, respectively. For the *cis*- and *trans*-guanidino-Pro residues, both puckerings of the proline ring (up and down) were included to allow for the two different orientations of the attached guanidino moiety.

Systematic Conformational Sampling for Compounds 9-15. Due to the presence of the highly flexible Arg³ side chain in compounds **9-15**, the conformational sampling procedure described for **1-8** (using backbones as templates) would result in too many conformations to be computationally feasible (up to *ca.* 8 million starting structures). Instead, the conformations obtained for the corresponding Ala³ analogs of **9-15** after energy minimization with the ECEPP/2 force field (see below) were used as templates onto which Arg³ was added, followed by generation of all Arg³ rotamers (81 starting points). These templates were compounds **1**, **3**, and the “virtual” compounds D-Ala³-D-Nal⁵FC131, *c*(Gly¹-D-Nal⁵-D-Arg⁴-D-Ala³-Tyr²), *c*(Gly¹-D-Nal⁵-D-Arg⁴-Ala³-Tyr²), *c*(Gly¹-Nal⁵-D-Arg⁴-Ala³-Tyr²), and *c*(Gly¹-Nal⁵-D-Arg⁴-D-Ala³-D-Tyr²), respectively.

Stochastic Conformational Sampling for Compound 16. The non-peptide compound **16** was subjected to a stochastic conformational search of 100 000 steps using the Monte Carlo search algorithm as implemented in the commercially available MacroModel program package v7.2.²³ All rotatable bonds, including the two amide bonds, were sampled.

Energy Minimization. The guanidino groups of **1-16** and the amino group of **16** were all built with positive charge. All energy calculations were performed *in vacuo* with a dielectric constant (ϵ) of 80.0. There are considerable uncertainties involved in mimicking the heterogeneous transmembrane protein environment of the CXCR4 receptor, and this treatment was chosen to dampen the strong electrostatic interaction between charged groups, thereby allowing exploration of a wider set of low-energy conformations. Energy minimization was routinely performed with the ECEPP/2 force field, which employs rigid valence geometry that drastically reduces the computational time required for each minimization run, and/or the OPLS-AA force field.

ECEPP/2 Force Field. The ECEPP/2 force field^{24,25} with all additions was implemented in the same in-house programs as extensively used previously, see e.g. ref. 26. To close the CPP backbone ring, parabolic closing potentials with U_0 of 100 kcal/mol were added to interactions between the C^α and C' atoms of Gly¹ and the dummy atoms representing the Me and C' atoms in the Nal⁵-NHMe construct, respectively. The criterion to exit minimization was a gradient of $0.001 \text{ kcal} \times \text{mol}^{-1} \times \text{\AA}^{-1}$.

OPLS-AA Force Field. The OPLS-AA force field²⁷ was used as implemented in MacroModel v7.2.²³ The Truncated Newton Conjugate Gradient (TNCG) algorithm²⁸ was used for energy minimization, with a maximum number of iterations of 5000 and a convergence threshold of $0.0024 \text{ kcal} \times \text{mol}^{-1} \times \text{\AA}^{-1}$ ($0.01 \text{ kJ} \times \text{mol}^{-1} \times \text{\AA}^{-1}$). Non-bonded cut-offs of 4 \AA for H-bond interactions, 7 \AA for van der Waals interactions and 12 \AA for electrostatics were used. Charges were taken from the force field.

Energy Minimization for Backbone Templates of Compounds 1-8. Starting conformations for the backbones (satisfying the ring closing criterion, see above) were energy minimized using the ECEPP/2 force field only. All resulting conformations were kept, and thereafter tested for redundancy.

Energy Minimization for Starting Conformations of Compounds 1-15. For the starting conformations of compounds **1-15**, initial energy minimization was done with the ECEPP/2 force field, and low-energy conformations were selected using an energy cut-off of 10 kcal/mol. The resulting conformations of compounds **1-15** were subsequently subjected to energy minimization using the OPLS-AA force field. Duplicate conformations were removed in the latter minimization procedure based on comparison of spatial positions of heavy atoms as implemented in the Maestro interface (v4.1.012) to MacroModel. The symmetric C^δ and C^ε atoms of D-Tyr and the symmetric Nⁿ atoms of Arg were excluded from the heavy atom comparison.

Energy Minimization for Compound 16. Only the OPLS-AA force field was used for energy minimization of **16**.

Structural Comparison of Compounds

An in house program was used for inter-compound RMSD comparison between selected atoms. In all cases, an RMSD value of ≤ 1.0 Å was employed. Initially, only conformations within 3 kcal/mol of the lowest energy conformation after energy minimization with OPLS-AA were included in the structural comparison. For the retro-inverso compounds (**12-15**), however, the energy cut-off was subsequently increased to 4 kcal/mol.

Removal of Artificial Conformations for Compound 5

The lowest energy conformations for the *N*-methylated compound **5** resulting from the conformational search after energy minimization with OPLS-AA contained an unrealistic backbone conformation. More specifically, a positive value of the ϕ angle (approx. 56°) and a negative value of the ψ angle (approx. -92°) was observed for the NMe-Ala residue, a combination that is sterically forbidden for amino acid residues with L-configuration according to the Ramachandran plot. Therefore, these conformations were regarded as computational artifacts and were removed from further investigations. This was done by clustering all conformations of **5** based on the heavy atoms in the backbone and all heavy atoms directly attached (25 atoms overall) using an $\text{RMSD} \leq 1.0 \text{ \AA}$. The conformations contained in the cluster representing the artificial backbone conformation were subsequently discarded.

RESULTS AND DISCUSSION

Compound Selection

Totally, 16 compounds were included in the computational investigations (compounds **1-16**, Table I), 15 of which were cyclic pentapeptides (**1-15**). For all the compounds included in the study, anti-HIV activity has been correlated with affinity to CXCR4. For ease of discussion, we therefore define high affinity as $IC_{50} \leq 100$ nM or $EC_{50} \leq 1$ μ M, medium affinity as 100 nM $< IC_{50} \leq 1000$ nM or 1 μ M $< EC_{50} \leq 5$ μ M, and low affinity as $IC_{50} > 1000$ nM or $EC_{50} > 5$ μ M in order to enable comparison between compounds with different kinds of available experimental data. Compounds **1-11**, with the general sequence $c(\text{Gly}^1\text{-D-Tyr}^2\text{-L/D-Xaa}^3\text{-Arg}^4\text{-L/D-Nal}^5)$, were used for generation of the three-point pharmacophore model. These compounds represent high affinity (compounds **1, 2, and 7-11**), medium affinity (compounds **3-5**), and low affinity (compound **6**). The RI-compounds **12-14** (low affinity) and **15** (medium affinity) were used for validation of the final three-point model. Compounds **7-16** were used for the subsequent development of a four-point pharmacophore model based on the features identified for the final three-point model.

Conformational Search

The results of the exhaustive conformational searches for compounds **1-8** and **9-16** are summarized in Tables II and III, respectively. For compound **5**, low-energy conformations containing unrealistic backbone torsions were removed in order to

represent the reference energy adequately (see Methods). Note that the initial sampling of compounds **9-15** was performed on the corresponding Ala³-analogs (see Methods).

Identification of the Three-Point Pharmacophore Model

The retro-inverso concept is based on the assumption that a similar side chain topology can be attained by simultaneously inverting the stereochemistry of chiral residues and reversing the peptide sequence. For the CPP CXCR4 inhibitors, application of the retro-inverso concept has been shown to be unsuccessful, i.e. the RI-analogs **12-14** (derived from the high affinity diastereomers **9-11**) displayed lower anti-HIV activity than certain RI-analogs derived from parent FC131 diastereomers with low affinity²⁰ (see also ref. 29). Compound **15** ($EC_{50} = 1.7 \mu\text{M}$), which is the RI-analog of L-Tyr²-D-Nal⁵FC131 ($EC_{50} = 11 \mu\text{M}$), displayed the highest activity of all the 16 RI-compounds.²⁰ Even if calculations have revealed inherent differences between a RI-analog and its parent compound that stem from the different connectivity (i.e. bond lengths) and the transposition of (ϕ, ψ) angles,³⁰ these differences will be negligible in most cases. The most obvious structural difference between a peptide and its RI-analog is the exchange of spatial positions between the carbonyl- and the NH-groups in the backbone. Due to this exchange, there are two main reasons why a RI-analog may display poorer biological effect than its parent compound. First, the reversed amide bonds may alter the intramolecular hydrogen-bonding pattern (backbone-backbone or backbone-side chain interactions) possibly preventing adaptation of the conformation corresponding to the bioactive conformation of the parent peptide (i.e. the similar side chain arrangement). Second, one or more of the amide bonds in the parent compound may be involved in

peptide-receptor interaction; the reversed amide bond will therefore be less compatible with the binding mode of the parent compound. Structure determination of compounds **1** and **9** by NMR did not reveal specific intramolecular hydrogen bonds within the backbone that can be ascribed to the characteristic backbone turn conformations.^{14,17} Hence, our working hypothesis was that the main reason for failure of the retro-inverso concept for CPP CXCR4 inhibitors is involvement of the peptide backbone in peptide-receptor interaction. Therefore, in addition to sharing a common arrangement of the three “pharmacophoric” side chains, it was expected that the low-energy conformations of the high/medium affinity compounds **1-5** and **7-11** would have certain common features in the backbone as well. Accordingly, some or all of these conformational characteristics would be absent for the low affinity compound **6**. Since the CPPs contain five amide bonds, all of which were candidates for receptor interaction, the amide bond(s) involved in this interaction had to be identified in an iterative fashion. Compound **1** was chosen as the reference compound for the three-point pharmacophore model due to its high affinity, the relatively small number of low-energy conformations, and its direct analogy to FC131.

Structural Comparison not Including the Backbone. Initially, a simple three-point pharmacophore model including only the Tyr², Arg⁴, and Nal⁵ side chains was considered, i.e. not including any backbone amide bonds. The 53 low-energy conformations obtained for compound **1** were compared to those of compounds **2-11** using eight atoms, namely C^ζ of D-Tyr², C^ζ of Arg⁴, C^{ζ1} of Nal⁵ (representing the three pharmacophoric points), and the five C^α atoms (representing the overall molecular

volume). By using this approach, three conformations of compound **1** were found that matched all the other compounds (**2-11**), i.e. the low affinity compound **6** was not excluded. The fact that it was not possible to discriminate between the compounds based on this simplistic approach further supported our hypothesis that one or more amide bond had to be involved in peptide-receptor interaction.

Structural Comparison Including both the Xaa³-Arg⁴ and Arg⁴-Nal⁵ Amide Bonds. By replacing the Nal⁵-Gly¹ fragment of compound **10** (IC₅₀ = 8 nM) with a γ -Nal residue (IC₅₀ = 54 nM), it has been shown that the Nal⁵-Gly¹ amide bond is not required for high affinity.¹⁷ Replacement of the Arg⁴-Nal⁵ amide bond of compound **9** with an (*E*)-alkene isostere resulted in a 33-fold reduction in activity, which was interpreted as either involvement of this amide bond in receptor interaction, or alternatively as an unfavorable effect of the increased hydrophobicity.¹⁹ Based on the 3-4 fold difference in affinity for compounds **1** and **3** (Xaa³ epimers), for which NMR data showed an opposite orientation of the D-Tyr²-Xaa³ amide bond plane, it was suggested that this amide bond could be important for activity.¹⁷ The same reasoning would apply for compounds **9** and **10** (2-fold difference in affinity) since these are also Xaa³ epimers. However, since the differences in affinity for the Xaa³ epimers are small, it is not likely that the D-Tyr²-Xaa³ amide bond is directly involved in receptor interaction. Based on the high affinity of the peptidomimetic CXCR4 antagonist KRH-1636 (compound **16**, Figure 1C), which presumably mimics the Arg³-Arg⁴-Nal⁵ fragment of FC131, we considered the amide bonds between Xaa³-Arg⁴ and Arg⁴-Nal⁵ to be the most likely candidates for backbone receptor interaction. Therefore, the carbonyl oxygens and the H^N atoms of these two

amide bonds were included in the structural comparison to give 12 atoms overall. By applying this more restrictive approach, three conformations of **1** were found that matched all compounds except **6** (low affinity) and **11** (high affinity); this result also did not fit the desired pattern.

Structural Comparison Including the Arg⁴-Nal⁵ Amide Bond. When including only the Arg⁴-Nal⁵ amide bond (10 atoms overall), we again failed to exclude the low affinity compound **6**. Conversely, in this case three conformations of **1** were found that matched all compounds except the high affinity compound **11**.

Structural Comparison Including the Xaa³-Arg⁴ Amide Bond. By including only the Xaa³-Arg⁴ amide bond (10 atoms overall) in the structural comparison, we successfully obtained the desired pattern: three conformations of the reference compound **1** were found that matched all the high and medium affinity compounds (**2-5** and **7-11**), but not compound **6** (low affinity). In terms of the CPP backbone, these three reference conformations (termed t1-t3) were virtually identical. The only difference between t1 and t3 was χ_2 of Nal⁵, i.e. flipping of the naphthyl ring system, whereas the orientation of Arg⁴ for t2 was somewhat different from that of t1/t3 (Figure 2). The torsion angles for t1 are given in Table IV; the torsion angles for t2 and t3 were similar, except a χ_2 angle of $\sim 115^\circ$ for Nal⁵ (t2 and t3) and slightly different χ_i values for Arg⁴ (t2). At this point, the spatial arrangement of the D-Tyr², Arg⁴ and Nal⁵ side chains and the specific orientation of the Xaa³-Arg⁴ amide bond plane – as defined by the reference conformations t1-t3 –

was taken to represent the three-point pharmacophore model. Thus, both $\text{Nal}^5 \chi_2$ rotamers were retained as possible alternatives.

Discussion – The Three-Point Pharmacophore Model

Since it was not possible to discriminate between the high/medium and low affinity compounds based solely on the three side chains, important information had to be contained in the conformational preference around the identified $\text{Xaa}^3\text{-Arg}^4$ backbone fragment. With the exception of compound **11**, all the high/medium affinity compounds displayed conformations with the same backbone geometry as t1-t3 (Table IV, Figure 2).

In the further discussion, we assume that this backbone conformation is optimal for binding to CXCR4.

Regarding compounds **1-5** (having only three pharmacophoric side chains), it is seen that the (ϕ, ψ) angles for Xaa^3 correspond to $(-60^\circ, -60^\circ)$, a region of the Ramachandran plot that is allowed for both L- and D-configured residues (with the exception of D-Pro). However, this combination of (ϕ, ψ) angles, corresponding to a right-handed α -helix, is energetically slightly preferred for L-configured residues since the side chain of a D-configured residue experiences steric strain from the neighboring carbonyl groups. Thus, the preference for compound **1** (L-Ala³) over **3** (D-Ala³) can be explained by the conformational ensemble effect since **1** will have a somewhat higher population of the optimal backbone conformation. The easiest way for compound **3** to relieve the strain is by flipping the D-Tyr²-D-Ala³ amide bond plane, which results in a backbone conformation corresponding to the deduced NMR structure for **3**.¹⁷ In contrast, flipping of the D-Ala³-Arg⁴ amide bond plane will result in (ϕ, ψ) angles of approximately $(-140^\circ,$

80°) and (100°, -50°) for D-Ala³ and L-Arg⁴, respectively, i.e. angles that are incompatible with the side chain stereochemistry of **3**. For the NMe-Ala³ epimers **2** and **5**, the *N*-methylation will switch the conformational preference for (ϕ, ψ) angles of (-60°, -60°) in favor of the D-configured residue (**2**) due to the pronounced steric clash between the *N*-methyl group and the α -methyl group of **5**. In contrast to **3**, compound **5** does not have the option of flipping the D-Tyr²-NMe-Ala³ amide bond plane, since this would result in (ϕ, ψ) angles that are incompatible with L-Xaa³ stereochemistry. Hence, it is expected that the optimal backbone conformation is poorly populated for **5**.

The fact that compound **4** (Pro³) displays a somewhat lower affinity than **1**, even if **4** is able to adopt conformations virtually identical to t1/t2/t3, again suggests a contribution from the conformational ensemble effect. This is rationalized by the fact that compound **4**, due to the presence of the proline residue, is prone to adaptation of backbone conformations containing a *cis* amide bond, which would be unfavorable for affinity (see below). An additional explanation for the difference between **1** and **4** is that a certain degree of flexibility around ϕ_3 is beneficial for optimal accommodation of the ligand within the binding site.

In contrast, the incorporation of D-Pro in Xaa³ (i.e. the low affinity compound **6**) is incompatible with (ϕ, ψ) angles of (-60°, -60°) due to the restraints imposed by the proline ring. The low affinity compound **6** therefore has to adopt a different backbone conformation in order to resemble the 3D pharmacophore. Inspection of the conformations of **6** with a similar side chain arrangement as t1/t2/t3 revealed substantial differences in the backbone region connecting D-Tyr² and D-Pro³ due to adaptation of *cis* amide bond geometry. This conformational change propagates to the Xaa³-Arg⁴ amide

bond, which is flipped by approx. -90° compared to the reference conformations of compound **1** (Figure 3A). Another consequence of the *cis* amide bond geometry is occupation of a different volume than seen for **1-5**. Moreover, it results in a distorted orientation of the D-Tyr² C^α-C^β vector, which could result in sub-optimal interaction between the phenol moiety and its binding partner(s) in the receptor. The reduced affinity of **6** is therefore assumed to be a sum of several unfavorable contributions.

For the high affinity compounds **7-11** it is evident that the additional Xaa³ guanidino group, which is not considered in the three-point model, contributes favorably to affinity. This means that the three-point model cannot account for the overall affinity of these compounds. However, since the applicable features of the model should still be valid, some general lines can be drawn. In the same way as **1** (L-Ala³) and **3** (D-Ala³), the high affinity compounds **9** (L-Arg³) and **10** (D-Arg³) are Xaa³ epimers. This suggests that the slight (2-fold) difference in affinity between **9** and **10** could be explained by analogy to compounds **1** and **3** (see above). Similarly, the 2.5-fold difference in affinity between **9** and **7/8** (containing a substituted Pro residue) could be explained by analogy to compounds **1** and **4**.

Compound **11**, having a D-configured Nal⁵ residue, did not share the backbone conformation seen for the other high/medium affinity compounds. This is due to the fact that the (ϕ, ψ) angles for Nal⁵ described by t1-t3 (-107° , 81° ; Table IV) are exclusive for L-configured residues. This means that the amide bond planes between Arg⁴-Nal⁵ and/or Nal⁵-Gly¹ have to be flipped in order to arrive at (ϕ, ψ) angles compatible with a D-configured Nal⁵. For all the low-energy conformations of **11** compatible with the three-point model, both of these amide bond planes were flipped by $\sim 180^\circ$ (Figure 3B). The

orientation of the Arg⁴-Nal⁵ amide bond seen for **11** was also found for compatible conformations of **2**, **4-5**, **7-8**, **9**, and **10**, i.e. all the high/medium affinity compounds except **1** and **3**. Since compound **1** was used as reference compound for the model, we conclude that the candidate Arg⁴-Nal⁵ amide bond does not seem to be involved in receptor interaction. Another result of the D-configured Nal⁵ is that the naphthyl group of **11** is not coplanar with the naphthyl groups of the other high/medium affinity compounds. This difference is likely to be the main reason for the small (2-fold) reduction in affinity for **11** compared to its Nal⁵-epimer **10**.

In conclusion, we were able to rationalize the experimental affinity/activity data for compounds **1-11** based on the features defined by the model.

Validation – Comparison with Retro-Inverso Analogs. In order to validate the three-point pharmacophore model, conformations t1-t3 were subsequently compared to the low-energy conformations of the RI-compounds **12-14** (low affinity) and **15** (medium affinity) using the same 10 atoms. When using the original energy cut-off of 3 kcal/mol, none of the RI-compounds matched the three-point model. However, when the cut-off was increased to 4 kcal/mol, three matching conformations of the medium affinity compound **15** were found. Interestingly, in addition to a similar side chain arrangement, these conformations of **15** displayed an orientation of all the five amide bond planes similar to t1-t3, even if the C=O and NH groups were shifted. Hence, we concluded that the three-point model was applicable also to the retro-inverso CPP analogs. It should be emphasized that none of the 16 RI-analogs displayed high activity (defined as EC₅₀ < 1 μM), i.e. the amino acid sequence for e.g. compound **9** is preferred over that of e.g. **15**.

The fact that the matching conformations of compound **15** were relatively high in energy (between 3 and 4 kcal/mol) could obviously be a contributing factor in this respect.

When the Xaa³-Arg⁴ amide bond was excluded from the comparison, i.e. including only the three side chains (8 atoms), several conformations matching t1-t3 were found also for **12-14**. This means that that the main reason for the low affinities of compounds **12-14**, with a side chain stereochemistry more compatible with the three-point model than compound **15**, is a different backbone preference in the Xaa³-Arg⁴ fragment.

Validation – Compatibility with Xaa², Xaa⁴, and Xaa⁵ Substituted FC131 Analogs.

Recently, Tamamura et al. published affinity data for a series of FC131 analogs with substitutions in the Xaa², Xaa⁴, and Xaa⁵ positions.¹⁸ Interestingly, substitution of D-Tyr² with a conformationally constrained analog, D-Tic(7-OH), resulted in a compound with an IC₅₀ of 160 nM, here classified as the one with medium affinity. Direct incorporation of D-Tic(7-OH) into the reference conformation t1 results in an incorrect orientation of the phenol group due to a constrained χ_1 -angle of $\sim 0^\circ$ as opposed to a value of $\sim 180^\circ$ for t1 (not shown). However, when the preceding amide bond (Gly¹-D-Tic(7-OH)²) was modeled with *cis* geometry, which is commonly seen for N-substituted residues in CPPs,²² the phenyl groups of the D-Tic(7-OH)² analog and t1 overlap reasonably well (Figure 6A).

In contrast, substitution of Nal⁵ with Tpi, a constrained Trp-analog (Trp⁵ substitution itself resulting in a high affinity compound), gave a low affinity compound (no detectable binding up to 1000 nM). Incorporation of Tpi into t1 does not result in a

good overlap of the indole ring of Tpi with the naphthyl ring of Nal, regardless of the geometry (*trans* or *cis*) of the Arg⁴-Tpi⁵ amide bond (Figure 6B).

Moreover, substitution of Arg⁴ with a series of basic residues (Arg/Lys-mimetics) revealed that the Xaa⁴ position is very sensitive even to very conservative substitutions. Interestingly, replacement of Arg⁴ with guanidino-Dab (one CH₂-group shorter than Arg) resulted in a low affinity compound, whereas replacement with guanidino-Lys (one CH₂-group longer than Arg) resulted in a high affinity compound. This is in sharp contrast to Arg³ substitution, where all Arg/Lys-mimetics tested resulted in high affinity compounds.¹⁷ The experimental results for the Xaa⁴ substitutions suggest that the side chain of Arg⁴ binds in an extended conformation, which is in accordance with the three-point pharmacophore model of the present study.

Validation – Consistency with NMR Data. By simulated annealing/molecular dynamics (SA/MD) based on ¹H NMR data, Fujii et al. have proposed a 3D structure for **1**.¹⁷ Visual inspection revealed that the common backbone conformation for conformations t1-t3 of compound **1** (representing the three-point model) that we have obtained by independent energy calculations is compatible with the backbone of this semi-experimental structure. In contrast, the side chain orientation of t1-t3 is not the same as that suggested by Fujii et al., the most evident difference being the χ_1 angle of Nal⁵. However, the side chain orientations of the 3D structure suggested by Fujii et al. were not based on actual experimental data, which were confined to the backbone-backbone NOEs and the H ^{α} -H^N coupling constants.

Validation – Development of a Four-Point Pharmacophore Model. As an additional validation step, further investigations were performed to see if an independent four-point pharmacophore model could be developed based on the features identified for the three-point model. The four-point model was based on compounds **7-15**, containing the fourth pharmacophoric side chain (the Xaa³ guanidino group), using **9** as the reference compound. Hence, the 219 low-energy conformations of **9** were compared to compounds **7-8** and **10-15**, this time using 11 atoms: C^ζ of Tyr², the C atom of the Xaa³ guanidino group, C^ζ of Arg⁴, C^{ζ1} of Nal⁵, the carbonyl oxygen and H^N in the Xaa³-Arg⁴ amide bond, and the five C^α atoms. At this increased level of comparison, we identified two conformations of **9** that matched all the high/medium affinity compounds (**7-8**, **10-11**, and **15**), and none of the low affinity compounds (**12-14**). Moreover, comparison of **9** with the peptidomimetic **16** using 7 atoms (Figure 1) resulted in the identification of several conformations of **16** that matched the same two conformations of **9**. These two conformations of **9** (termed f1 and f2) are superimposed in Figure 4, which clearly illustrates their geometrical similarity. Conformation f1 displayed a backbone similar to the common backbone conformation of t1-t3 (Table IV), whereas the Arg⁴-Nal⁵ amide bond plane of f2 was flipped by ~180° relative to that of f1, resulting in (φ,ψ) angles of (-143°, 119°) and (59°, 77°) for Arg⁴ and Nal⁵, respectively. Interestingly, conformations f1 and f2 were both among the 16 conformations of **9** that were compatible with the three-point model (represented by t1-t3). Furthermore, the ¹H NMR data for **9** in DMSO reported by Fujii et al.¹⁴ showed cross-peaks between Gly¹ H^N/D-Tyr² H^N, Arg³ H^N/Arg⁴ H^N, and Arg⁴ H^N/Nal⁵ H^N, indicating that these pairs of amide H-atoms were oriented in the same direction. The NMR data also showed that Nal⁵ H^α/Gly¹ H^N and D-Tyr² H^α/Arg³

H^N were in proximity across the peptide bonds. These experimental data are consistent with the backbone of the reference conformation f1. Thus, the selected side chain and backbone features collectively represented by f1 and f2 describe a plausible four-point model.

The fact that several conformations of the peptidomimetic **16** matched f1 and f2 supports the view that compound **16** mimics the Arg³-Arg⁴-Nal⁵ fragment of the CPPs and binds to the same CXCR4 site. For the matching conformations of **16**, the naphthyl group (contained in the 1-naphthylglycine mimicking fragment) was not coplanar with the naphthyl group of L-Nal⁵ for f1/f2, but instead arranged similarly to the naphthyl group of D-Nal⁵ for compound **11** (Figure 5). Interestingly, the affinities of compound **11** and **16** are almost identical (16 and 13 nM, respectively).

For compounds **7** and **8** [*c*(Gly¹-D-Tyr²-*trans/cis*-4-guanidino-Pro³-Arg⁴-Nal⁵)], the fourth pharmacophoric side chain group (the Pro³ 4-guanidino group) is to some extent pre-defined by the backbone due to the limited flexibility of the proline ring. Superimposition of conformation f1 of **9** and compatible conformations of compounds **7-8** revealed a very similar positioning of the Xaa³ guanidino group for these three high affinity compounds. The good overlap of this guanidino group for **7** (*trans*) and **8** (*cis*) is a result of down-puckering of the proline ring for **7**, and up-puckering for **8**. However, even if conformations matching f1/f2 within the RMSD threshold of 1.0 Å were obtained for all high/medium affinity compounds containing the Xaa³ guanidino group, relatively large differences in the positioning of this guanidino group could be observed. Obviously, this difference was most pronounced for the compounds containing the Pro³ 4-guanidino group (**7** and **8**) relative to the compounds containing a D-configured Arg³ residue (**10**

and 11). Moreover, substitution of the FC131 Arg³ residue with Arg analogs of different chain length, i.e. guanidino-Dab and guanidino-Lys, have demonstrated that the guanidino group contributed favorably to affinity (relative to Ala³ substitution) in both cases.¹⁷ In other words, there seem to be several possibilities for positioning of the Xaa³ guanidino group that results in beneficial interaction with the receptor, which means that exact mapping of the fourth pharmacophoric side chain (the Xaa³ guanidino group) is associated with considerable uncertainty. This leads us to suggest that the binding partner for Arg³ (probably an Asp/Glu residue) is flexible enough to accommodate different binding modes. Consequently, we are reluctant to draw very strong conclusions about the position of this group from the four-point model. However, the fact that the independently developed four-point model is compatible with the minimalistic three-point model further supports the general applicability of the three-point model that we consider to be the main finding of this study.

Obviously, as this pharmacophore model has been developed without explicit consideration of ligand-receptor interactions, minor deviations from the model could be expected in the real binding situation due to e.g. induced fit. These deviations are likely to be more pronounced for the side chains than for the backbone.

Comparison with Previously Published Model. A QSAR model for the CPP CXCR4 antagonists was recently published by Bhonsle et al.³¹ The training set consisted of the 16 stereoisomers of FC131, whereas the test set was composed of 9 compounds with different sequences of the same five residues. Based on CoMFA/PLS analysis of the 25 lowest energy conformations of the training set, it was concluded that the positioning of

the two positively charged Arg residues, the hydroxyl group of Tyr, and the aromatic groups of Tyr and Nal were the important features for activity. The backbone was not identified as an important structural feature. Explicit 3D information for the final model is not available, but visual inspection of the representative conformations shown in the paper reveals that also the side chain arrangement is different from the three-point model identified in the present study, especially with respect to the positioning of Arg⁴.

In the model developed by Bhonsle et al., a single backbone conformation derived from NMR data for FC131 was used as template for both the training and test sets; the coordinates of this backbone template are not available, however, visual inspection reveals that this backbone is not the same as that suggested earlier by Fujii et al. based on NMR data.¹⁴ Obviously, since all the 25 compounds were modeled with the same backbone, it would not be possible to identify the backbone as an important feature for activity. Moreover, based on the Ramachandran plot of allowed (ϕ, ψ) angles of the peptide backbone, it is not reasonable that all 16 stereoisomers of FC131 (i.e. all combinations of four L- or D-configured residues) would be able to adopt the exact same backbone conformation (with the exception that all (ϕ, ψ) angles for the chiral residues corresponded to the ($60^\circ, 60^\circ$) or ($-60^\circ, -60^\circ$) minima). In the present study, this irregularity is exemplified by compound **11**, in which two amide bonds were flipped relative to the reference conformations t1-t3 of compound **1** (see above) in order for the (ϕ, ψ) angles of D-Nal⁵ to comply with the Ramachandran plot. Also, the strict use of this single backbone template would generally result in inadequate conformational sampling and in most cases lead to a misrepresentation of the global energy minimum.

CONCLUSIONS

Based on an extensive exploration of the conformational space for a series of highly flexible CPP CXCR4 antagonists, we have successfully identified a minimalistic 3D pharmacophore model (the three-point model) that is consistent with available experimental data. The model accounts also for the activities of retro-inverso analogs of parent CPPs, which indicates that it should be of general applicability. This knowledge will be valuable for guiding the rational design of new CXCR4 antagonists, i.e. transfer of the identified features from the relatively flexible CPP template to more rigid peptidomimetic templates with drug-like properties that are better suited for preorganization of the important functionalities of the pharmacophoric groups.

Acknowledgement. J.V. acknowledges financial support from the Research Council of Norway (167113/V40). G.V.N. and G.R.M. acknowledge support from the National Institutes of Health (GM068460). We would like to thank JW Feng for excellent assistance in resolving computational challenges, and Prof. N. Fujii for sharing of data prior to publication.

REFERENCES

1. Schuitemaker, H.; Koot, M.; Kootstra, N. A.; Dercksen, M. W.; de Goede, R. E.; van Steenwijk, R. P.; Lange, J. M.; Schattenkerk, J. K.; Miedema, F.; Tersmette, M. *J Virol* 1992, 66, 1354-1360.
2. Zhu, T.; Mo, H.; Wang, N.; Nam, D. S.; Cao, Y.; Koup, R. A.; Ho, D. D. *Science* 1993, 261, 1179-1181.
3. Tersmette, M.; de Goede, R. E.; Al, B. J.; Winkel, I. N.; Gruters, R. A.; Cuypers, H. T.; Huisman, H. G.; Miedema, F. *J Virol* 1988, 62, 2026–2032.
4. Tersmette, M.; Lange, J. M.; de Goede, R. E.; de Wolf, F.; Eeftink-Schattenkerk, J. K.; Schellekens, P. T.; Coutinho, R. A.; Huisman, J. G.; Goudsmit, J.; F., M. *Lancet* 1989, 1, 983-985.
5. Tersmette, M.; Gruters, R. A.; de Wolf, F.; de Goede, R. E.; Lange, J. M.; Schellekens, P. T.; Goudsmit, J.; Huisman, H. G.; Miedema, F. *J Virol* 1989, 63, 2118–2125.
6. Richman, D. D.; Bozzette, S. A. *J Infect Dis* 1994, 169, 968-974.
7. Connor, R. I.; Ho, D. D. *J Virol* 1994, 68, 4400-4408.
8. Connor, R. I.; Sheridan, K. E.; Ceradini, D.; Choe, S.; Landau, N. R. *J Exp Med* 1997, 185, 621-628.
9. Zhang, L.; He, T.; Huang, Y.; Chen, Z.; Guo, Y.; Wu, S.; Kunstman, K. J.; Brown, R. C.; Phair, J. P.; Neumann, A. U.; Ho, D. D.; Wolinsky, S. M. *J Virol* 1998, 72, 9307-9312.

10. Feng, Y.; Broder, C. C.; Kennedy, P. E.; Berger, E. A. *Science* 1996, 272, 872-877.
11. Berson, J. F.; Long, D.; Doranz, B. J.; Rucker, J.; Jirik, F. R.; Doms, R. W. *J Virol* 1996, 70, 6288-6295.
12. Bleul, C. C.; Farzan, M.; Choe, H.; Parolin, C.; Clark-Lewis, I.; Sodroski, J.; Springer, T. A. *Nature* 1996, 382, 829-833.
13. Oberlin, E.; Amara, A.; Bachelier, F.; Bessia, C.; Virelizier, J.-L.; Arenzana-Seisdedos, F.; Schwartz, O.; Heard, J.-M.; Clark-Lewis, I.; Legler, D. F.; Loetscher, M.; Baggiolini, M.; Moser, B. *Nature* 1996, 382, 833-835.
14. Fujii, N.; Oishi, S.; Hiramatsu, K.; Araki, T.; Ueda, S.; Tamamura, H.; Otaka, A.; Kusano, S.; Terakubo, S.; Nakashima, H.; Broach, J. A.; Trent, J. O.; Wang, Z. X.; Peiper, S. C. *Angew Chem Int Ed Engl* 2003, 42, 3251-3253.
15. Tamamura, H.; Xu, Y.; Hattori, T.; Zhang, X.; Arakaki, R.; Kanbara, K.; Omagari, A.; Otaka, A.; Ibuka, T.; Yamamoto, N.; Nakashima, H.; Fujii, N. *Biochem Biophys Res Commun* 1998, 253, 877-882.
16. Tamamura, H.; Omagari, A.; Oishi, S.; Kanamoto, T.; Yamamoto, N.; Peiper, S. C.; Nakashima, H.; Otaka, A.; Fujii, N. *Bioorg Med Chem Lett* 2000, 10, 2633-2637.
17. Tamamura, H.; Araki, T.; Ueda, S.; Wang, Z.; Oishi, S.; Esaka, A.; Trent, J. O.; Nakashima, H.; Yamamoto, N.; Peiper, S. C.; Otaka, A.; Fujii, N. *J Med Chem* 2005, 48, 3280-3289.

18. Tamamura, H.; Esaka, A.; Ogawa, T.; Araki, T.; Ueda, S.; Wang, Z.; Trent, J. O.; Tsutsumi, H.; Masuno, H.; Nakashima, H.; Yamamoto, N.; Peiper, S. C.; Otaka, A.; Fujii, N. *Org Biomol Chem* 2005, 3, 4392-4394.
19. Tamamura, H.; Hiramatsu, K.; Ueda, S.; Wang, Z.; Kusano, S.; Terakubo, S.; Trent, J. O.; Peiper, S. C.; Yamamoto, N.; Nakashima, H.; Otaka, A.; Fujii, N. *J Med Chem* 2005, 48, 380-391.
20. Tamamura, H.; Mizumoto, M.; Hiramatsu, K.; Kusano, S.; Terakubo, S.; Yamamoto, N.; Trent, J. O.; Wang, Z.; Peiper, S. C.; Nakashima, H.; Otaka, A.; Fujii, N. *Org Biomol Chem* 2004, 2, 1255-1257.
21. Walters, P.; Stahl, M. Dolata Research Group, Department of Chemistry, University of Arizona, Tucson, AZ 85721 (<http://www.smog.com/chem/babel>)
22. Viles, J. H.; Mitchell, J. B.; Gough, S. L.; Doyle, P. M.; Harris, C. J.; Sadler, P. J.; Thornton, J. M. *Eur J Biochem* 1996, 242, 352-362.
23. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J Comput Chem* 1990, 11, 440-467.
24. Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J Phys Chem* 1978, 82, 2609-2616.
25. Nemethy, G.; Pottle, M. S.; Scheraga, H. A. *J Phys Chem* 1983, 87, 1883-1887.
26. Nikiforovich, G. V. *Int J Pept Protein Res* 1994, 44, 513-531.
27. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J Am Chem Soc* 1996, 118, 11225-11236.

28. Ponder, J. W.; Richards, F. M. *Journal of Computational Chemistry* 1987, 8, 1016-1024.
29. Chorev, M. *Biopolymers* 2005, 80, 67-84.
30. Freidinger, R. M.; Veber, D. F. *J Am Chem Soc* 1979, 101, 6129–6131.
31. Bhonsle, J. B.; Wang, Z.-X.; Tamamura, H.; Fujii, N.; Peiper, S. C.; Trent, J. O. *QSAR Comb Sci* 2005.
32. Ichiyama, K.; Yokoyama-Kumakura, S.; Tanaka, Y.; Tanaka, R.; Hirose, K.; Bannai, K.; Edamatsu, T.; Yanaka, M.; Niitani, Y.; Miyano-Kurosaki, N.; Takaku, H.; Koyanagi, Y.; Yamamoto, N. *Proc Natl Acad Sci U S A* 2003, 100, 4185-4190.

TABLES

TABLE I

Compounds included in the study with affinity/activity data and affinity classification.

Compound	Name	Sequence	IC ₅₀ (nM)	EC ₅₀ (μM)	Affinity Classification	Ref.
1	Ala ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -Ala ³ -Arg ⁴ -Nal ⁵)	63	N/A	High	17
2	D-NMe-Ala ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -D-NMe-Ala ³ -Arg ⁴ -Nal ⁵)	42	N/A	High	17
3	D-Ala ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -D-Ala ³ -Arg ⁴ -Nal ⁵)	230	N/A	Medium	17
4	Pro ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -Pro ³ -Arg ⁴ -Nal ⁵)	420	N/A	Medium	17
5	NMe-Ala ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -NMe-Ala ³ -Arg ⁴ -Nal ⁵)	490	N/A	Medium	17
6	D-Pro ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -D-Pro ³ -Arg ⁴ -Nal ⁵)	1600	N/A	Low	17
7	<i>trans</i> -4-guanidino-Pro ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² - <i>trans</i> -4-guanidino-Pro ³ -Arg ⁴ -Nal ⁵)	10	N/A	High	17
8	<i>cis</i> -4-guanidino-Pro ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² - <i>cis</i> -4-guanidino-Pro ³ -Arg ⁴ -Nal ⁵)	10	N/A	High	17
9	FC131	<i>c</i> (Gly ¹ -D-Tyr ² -Arg ³ -Arg ⁴ -Nal ⁵)	4	0.088	High	14,20

10	D-Arg ³ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -D-Arg ³ -Arg ⁴ -Nal ⁵)	8	0.34	High	14,20
11	D-Arg ³ -D-Nal ⁵ FC131	<i>c</i> (Gly ¹ -D-Tyr ² -D-Arg ³ -Arg ⁴ -D-Nal ⁵)	16	1.0	High	14,20
12	retro-inverso FC131	<i>c</i> (Gly ¹ -D-Nal ⁵ -D-Arg ⁴ -D-Arg ³ -Tyr ²)	N/A	9.6	Low	20
13	retro-inverso D-Arg ³ FC131	<i>c</i> (Gly ¹ -D-Nal ⁵ -D-Arg ⁴ -Arg ³ -Tyr ²)	N/A	75	Low	20
14	retro-inverso D-Arg ³ -D-Nal ⁵ FC131	<i>c</i> (Gly ¹ -Nal ⁵ -D-Arg ⁴ -Arg ³ -Tyr ²)	N/A	>100	Low	20
15	retro-inverso L-Tyr ² -D-Nal ⁵ FC131	<i>c</i> (Gly ¹ -Nal ⁵ -D-Arg ⁴ -D-Arg ³ -D-Tyr ²)	N/A	1.7	Medium	20
16	KRH-1636	Arg-Arg-Nal mimetic	13	0.019	High	32

TABLE IIResults of the systematic conformational search for compounds **1-8**.

Compound	Backbones considered	Ring closure ≤ 4 Å	Unique backbones	Total ^a	Energy Minimization	
					ECEPP/2 ^b	OPLS-AA ^c
1	10368	408	71	103518	1748	53
2	20736	1965	498	726084	6305	150
3	10368	376	55	80190	7523	86
4	10368	515	124	180792	3751	81
5	20736	909	299	435942	3526	139 ^d
6	10368	648	156	227448	5998	250
7	10368	515	124	1084752	11138	194
8	10368	515	124	1084752	7133	176

^a Total number of starting conformations; equal to number of unique backbone conformations times number of side chain rotamers (1458 for compounds **1-6** and 8748 (including the two puckerings of proline) for compounds **7-8**).

^b Energy cutoff = 10 kcal/mol.

^c Energy cutoff = 3 kcal/mol; duplicate conformations removed.

^d After removal of artificial low-energy conformations, see Methods.

TABLE IIIResults of the conformational search for compounds **9-16**.

Compound	Templates ^a	Total ^b	Energy Minimization	
			ECEPP/2 ^c	OPLS-AA
9	1748	141588	5919	219 ^d
10	7523	609363	6747	79 ^d
11	5923	479763	7790	115 ^d
12	6528	528768	7064	255 ^d (454 ^e)
13	4889	396009	6431	156 ^d (394 ^e)
14	7464	604584	8791	62 ^d (201 ^e)
15	5852	474012	8484	36 ^d (117 ^e)
16^f	-	-	-	1142 ^d

^a Number of low-energy conformations of Ala³-analog (energy cutoff = 10 kcal/mol) after energy minimization with ECEPP/2; see Methods.

^b Number of templates (second column) times 81 (number of Arg³ rotamers).

^c Energy cutoff = 10 kcal/mol.

^d Energy cutoff = 3 kcal/mol; duplicate conformations removed.

^e Energy cutoff = 4 kcal/mol; duplicate conformations removed.

^f Sampled by Monte Carlo method; see Methods.

TABLE IV

Torsion angles (°) for conformation t1 of compound **1**, representing the three-point pharmacophore model.

Residue	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Gly ¹	76	65	177	-	-	-	-
D-Tyr ²	138	-103	180	177	105	180	-
Ala ³	-68	-48	177	-	-	-	-
Arg ⁴	-138	-59	-178	-175	174	178	140
Nal ⁵	-107	81	-179	-62	-64	-	-

FIGURE LEGENDS

FIGURE 1

Structures of selected compounds: (A) the cyclopentapeptide FC131 (**9**); (B) general structure for the retro-inverso analogs (**12-15**); (C) the peptidomimetic KRH-1636 (**16**). The seven atoms used for the structural comparison of compound **9** and compound **16** are numbered from 1-7.

FIGURE 2

Superimposition of the three reference conformations (t1, t2, and t3) of compound **1**, together representing the three-point pharmacophore model.

FIGURE 3

Superimposition of the reference conformation t1 of compound **1** (green; representing the three-point model) and: (A) a non-matching conformation of the low affinity compound **6** (red); (B) a matching conformation of compound **11**, showing the different orientation of the Arg⁴-Nal⁵ and Nal⁵-Gly¹ amide bond planes and the difference in naphthyl ring plane.

FIGURE 4

Superimposition of the two reference conformations (f1 and f2) of compound **9**, together representing a plausible four-point pharmacophore model.

FIGURE 5

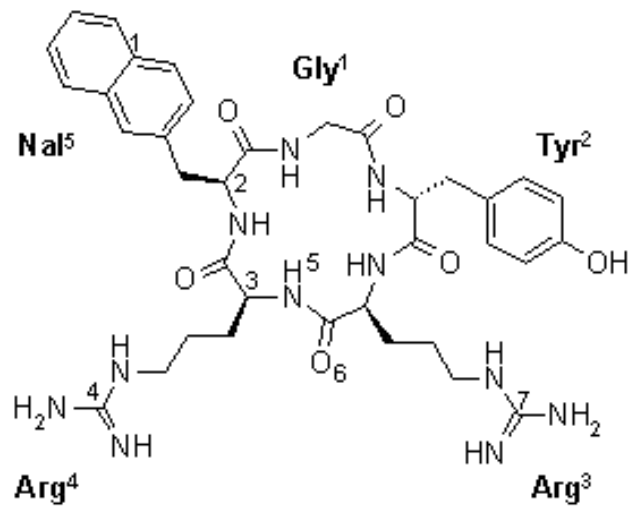
Superimposition of conformations of compounds **11** and **16** (green) that both matched the four-point pharmacophore model represented by conformations f1 and f2 (Figure 4), showing good overlap of the naphthyl groups, the positively charged groups, and the Arg-Arg amide bond for these two compounds.

FIGURE 6

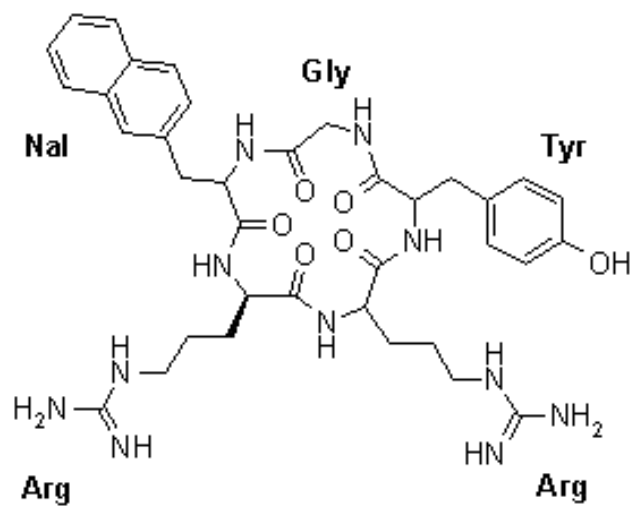
Superimposition of the reference conformation t1 of compound **1** (green; representing the three-point model) and: (A) t1 with incorporation of D-Tic(7-OH)² and *cis* geometry of the Gly¹-D-Tic(7-OH)² amide bond, demonstrating a reasonable overlap of the phenyl groups in Xaa²; (B) t1 with incorporation of Tpi⁵ and *trans* (purple) or *cis* (red) geometry of the Arg⁴-Tpi⁵ amide bond, demonstrating a poor overlap of the naphthyl/indole rings in both cases.



A



B



C

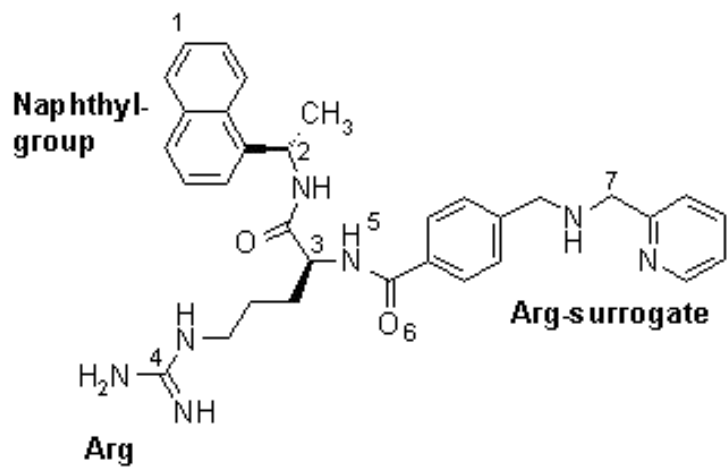


Figure 1



t

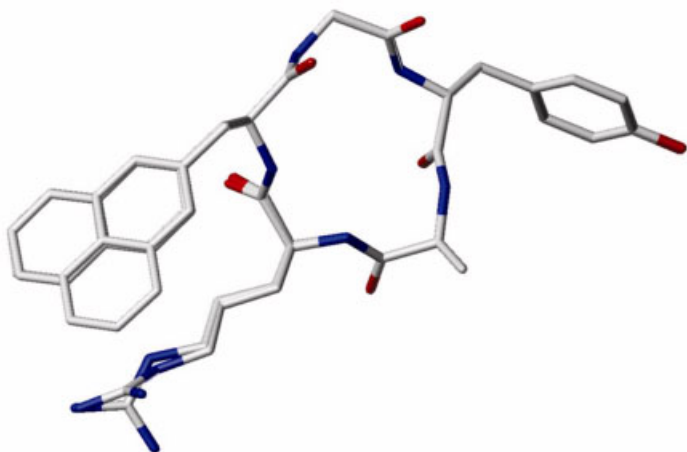


Figure 2

A



B

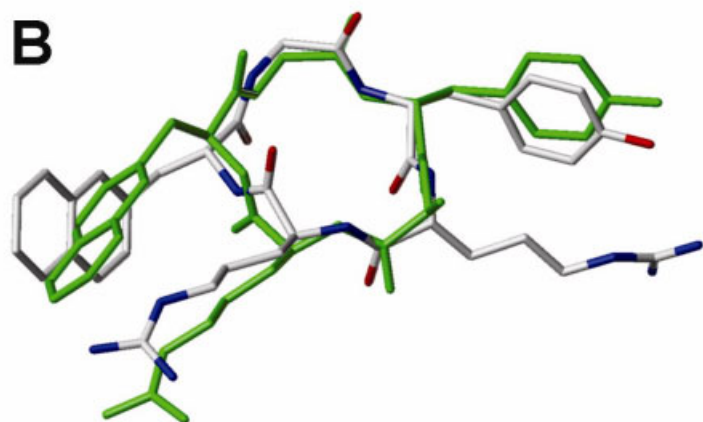


Figure 3

A

t

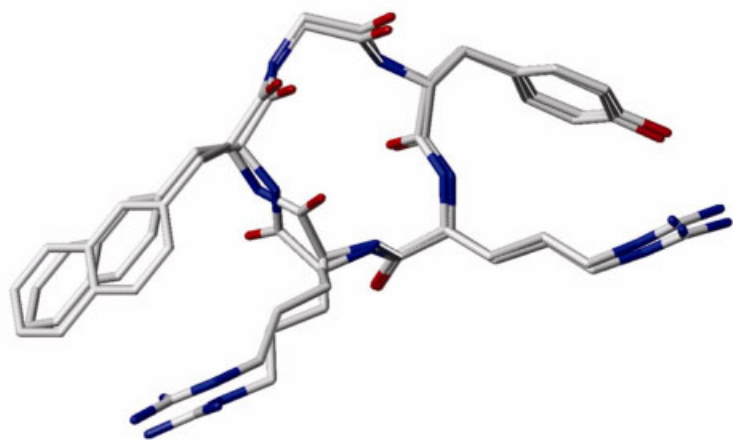


Figure 4

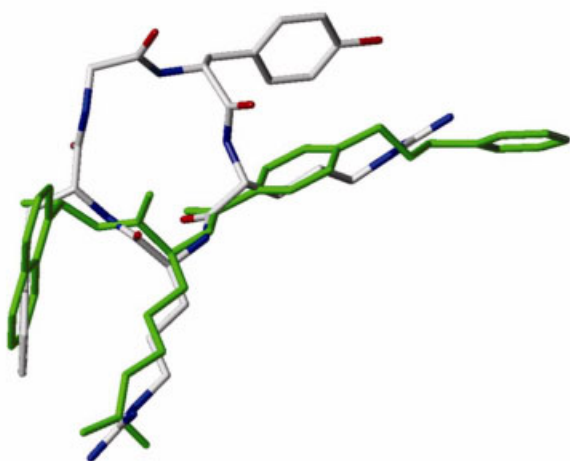
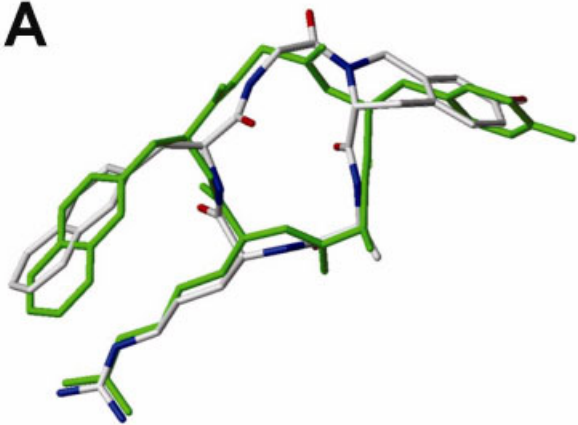


Figure 5

Accept

↑

A



B

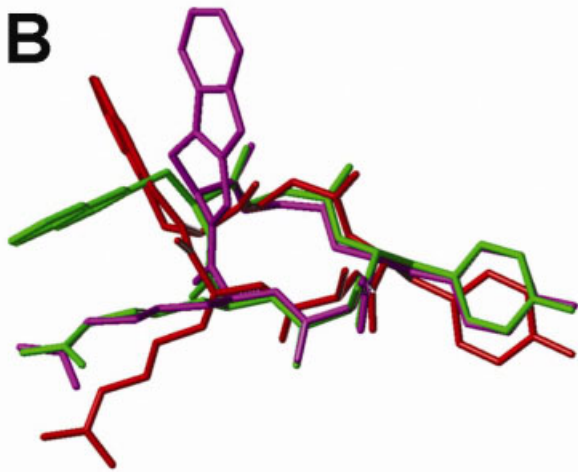


Figure 6

Accepted