Integrating the intrinsic conformational preferences of non-coded α-amino acids modified at the peptide bond into the NCAD database

Guillem Revilla-López, Francisco Rodríguez-Ropero, David Curcó, Juan Torras, M. Isabel Calaza, David Zanuy, Ana I. Jiménez, Carlos Cativiela, Ruth Nussinov, and Carlos Alemán

Abstract

Recently, we reported a database (NCAD, Non-Coded Amino acids Database; http://recerca.upc.edu/imem/index.htm) that was built to compile information about the intrinsic conformational preferences of non-proteinogenic residues determined by quantum mechanical calculations, as well as bibliographic information about their synthesis, physical and spectroscopic characterization, the experimentally-established conformational propensities, and applications (J. Phys. Chem. B 2010, 114, 7413). The database initially contained the information available for α-tetrasubstituted α-amino acids. In this work, we extend NCAD to three families of compounds, which can be used to engineer peptides and proteins incorporating modifications at the –NHCO– peptide bond. Such families are: N-substituted α-amino acids, thio-α-amino acids, and diamines and diacids used to build retropeptides. The conformational preferences of these compounds have been analyzed and described based on the information captured in the database. In addition, we provide an example of the utility of the database and of the compounds it compiles in protein and peptide engineering. Specifically, the symmetry of a sequence engineered to stabilize the 310-helix with respect to the α-helix has been broken without perturbing significantly the secondary structure through targeted replacements using the information contained in the database.

*Corresponding author: carlos.aleman@upc.edu.

Supporting Information
Dihedral angles and relative energies of the minimum energy conformations found for dc-Ala, dc-Aib, and dm-Aib at the B3LYP/6-31G(d) level.
Introduction

Apart from the genetically coded $\alpha$-amino acids found in proteins, there is a large number of amino acids grouped under the common name of non-proteinogenic or non-coded amino acids (nc-aa). Some of these nc-aa are present in nature, occurring in non-ribosomally produced peptides (e.g., antibiotics); however, most are synthesized by chemists. Among the latter, those used for protein engineering\(^1\)–\(^6\) and nanobiology\(^7\)–\(^12\) are particularly interesting, and their applications extend to many fields such as pharmacology (drug design),\(^13\),\(^14\) biotechnology (biosensors),\(^15\),\(^16\) and nanomedicine (systems for drug delivery and diagnosis through imaging).\(^17\),\(^18\)

In spite of the great significance and potential utility of nc-aa in modern biology, biomaterials engineering, and medicine, structural information is available only for a limited number of these. In the last two decades, the intrinsic conformational preferences of several tens of nc-aa have been investigated through sophisticated quantum mechanical calculations at the \textit{ab initio} or Density Functional Theory (DFT) levels. Specifically, these reliable theoretical methods have been used to sample the potential energy hypersurface of small peptide systems incorporating such nc-aa in the absence of external forces. At the same time, crystallographic and spectroscopic structural data on peptide sequences incorporating nc-aa are becoming available for an increasing number of cases. Comparison of such experimental information with that derived from high-level theoretical calculations helps establish the influence of packing forces, solvent effects, and the chemical environment on the conformational propensities of nc-aa.

In practice, the use of nc-aa with well-characterized conformational properties is frequently limited because the relevant information is highly dispersed. Quantum mechanical calculations describing the intrinsic conformational preferences of these compounds are typically reported in physical chemistry journals, while their synthesis and crystalline structures developed by organic and peptide chemists are published in journals specialized in these fields. In contrast, many applications of nc-aa are tested by researchers working on protein engineering, medicine, or materials science. In order to integrate the dispersed contributions about nc-aa, which is expected to facilitate their practical application in many research fields, a database of nc-aa was recently built.\(^19\)

NCAD (Non-Coded Amino acids Database) is a database conceived and created to identify the nc-aa that are compatible with a given structural motif, which is the key requirement for the application of these compounds in the life and materials sciences. The database, which is publicly accessible through the web page, http://recerca.upc.edu/imem/index.htm, integrates structural and energetic descriptors of those nc-aa whose intrinsic conformational properties have been previously studied using \textit{ab initio} or DFT quantum mechanical calculations and reported in the literature. In our recent work\(^19\) we described the technical aspects of the database (structure, contents, informatics features and relationships among the different descriptors) and presented the first family of nc-aa integrated into it. Specifically, we discussed the relevance of $\alpha$-tetrasubstituted $\alpha$-amino acids in the design of peptides with well-defined conformational properties and described the 29 $\alpha$-tetrasubstituted $\alpha$-amino acids whose intrinsic conformational propensities have been determined to date using \textit{ab initio} or DFT quantum mechanical methods according to a literature search.\(^19\)

In this work, we extend NCAD to three new families of compounds: (i) $N$-substituted $\alpha$-amino acids; (ii) thio-$\alpha$-amino acids; and (iii) diamines and diacids surrogates of $\alpha$-amino acids that are used to build pseudopeptides containing one (or more) retropeptide link(s), that is, a peptide bond in reversed direction to that of its immediate neighbors. The members of three such series can be used to engineer peptides and proteins incorporating...
modifications at the conventional –NHCO– peptide bond. Below, the work is organized as follows. First, a brief description of the contents and structure of NCAD is provided. After this, the three families of amino acids integrated in the database are presented. Next, the utility of the database is illustrated by engineering helical models using the information of the diamines and diacids included in NCAD. Finally, the conclusions of the work are outlined.

**NCAD: Integrating the intrinsic conformational preferences of non-coded amino acids**

NCAD integrates information about nc-aa whose intrinsic conformational preferences have been studied by applying *ab initio* or DFT quantum mechanical calculations to the smallest peptide systems which contains them (that is, the nc-aa bearing the amino and carboxylic acid termini as an amide, typically capped with the acetyl and methylamide groups). The information contained in NCAD is divided into three categories:

1. Minimum energy conformations. All the minima found for each compound contained in the database are described in detail. Dihedral angles, atomic coordinates and relative energies obtained at the highest level of theory among those used to study each nc-aa are stored through inter-connected descriptors. The backbone conformations of the different energy minima contained in NCAD have been identified using the nomenclature proposed by Perzcel *et al.*,20 which distinguish nine different conformations in the potential energy surface $E=E(\phi,\psi)$ of $\alpha$-amino acids (*i.e.* $\gamma_D, \delta_D, \epsilon_D, \beta_{DL}, \epsilon_L, \delta_L, \gamma_L$). This information is essential to check the compatibility of the different energy minima with the secondary structural motifs typically found in peptides and proteins.

2. Bibliographic information describing the available experimental information. Specifically, the following items are reported for each amino acid: (i) synthesis and characterization of the free form or with the amino group adequately protected for use in peptide synthesis; and (ii) spectroscopic and structural characterization when incorporated into peptides, the atomic coordinates extracted from the X-ray crystal structures being also included in selected cases.

3. Applications. The most relevant publications and patents describing the use of nc-aa in related fields along with the biological properties and relevance in materials science.

The labels identifying the information stored in the database are connected through a relational scheme. This facilitates the search process, which can be performed using a single or several descriptors. NCAD is accessed through a user-friendly interface, which integrates a simple window system, thus facilitating the use by non-experts in informatics. The interface is divided into six folders, the first one being used to define the criteria applied to the search process. The available search criteria are: (i) the chemical nature of the side chain; (ii) the structural feature which distinguishes a particular family of nc-aa; (iii) the molecular formula; (iv) the type of backbone conformation according to the nomenclature established by Perzcel *et al.*,20,21 (v) the values of the $(\phi,\psi)$ dihedral angles, with the tolerance being specified by the user; (vi) the name of the researcher who performed the theoretical or experimental investigation; (vii) the experimental data reported in the literature relating to the synthesis and characterization; and (viii) the applications described in the literature. The amino acids fulfilling all the search criteria are displayed in a list. Once the user selects a given amino acid from this list, the information stored in the database is presented in the remaining five folders of the interface. The specific information provided in each folder was detailed in our previous work.19
Amino-acid surrogates leading to peptides with modified peptide bonds

In our previous work,\textsuperscript{19} we provided evidence for the usefulness of the $\alpha$-tetrasubstituted $\alpha$-amino acids stored in NCAD in the control of peptide conformation. Another means of modulating the conformational properties of peptides can come from the modification of the peptide bond itself. This change should perturb the intramolecular hydrogen bonding network, the electronic distribution, and the steric hindrance, thus affecting the conformational preferences of the peptide chain.

N-substituted $\alpha$-amino acids

In $N$-substituted amino acids, the hydrogen atom attached to nitrogen is replaced by a functional group. The substituents most frequently used are: alkyl ($N$-alkylation), hydroxyl ($N$-hydroxylation) or amino ($N$-amination). To the best of our knowledge, reliable theoretical studies at the \textit{ab initio} or DFT quantum mechanical levels on the intrinsic conformational preferences of $N$-substituted amino acids have been performed only for $N$-methylalanine,\textsuperscript{22} $N$-hydroxyalanine\textsuperscript{23} and $N$-aminoglycine\textsuperscript{24} (Table 1).

$N$-alkylation (in particular, $N$-methylation) has been largely used to change the pharmacological properties of peptides. Remarkable biological and pharmacological profiles have been observed for many $N$-methylated peptides.\textsuperscript{25} The conformational preferences of $N$-methylated amino acids are dictated by increased steric hindrance and lack of hydrogen-bond donor ability associated with the replacement of the hydrogen atom by a methyl group.\textsuperscript{26,27} Figure 1a presents the Ramachandran map corresponding to the ($\phi$, $\psi$) backbone dihedral angles of all the energy minima reported for the $N$-acetyl-$N'$-methylamide derivatives of the $N$-methylated $\alpha$-amino acid included in the database, namely $N$-methylalanine.\textsuperscript{22} The minima derived from \textit{ab initio} quantum mechanical calculations for the analogous derivatives of glycine and alanine\textsuperscript{28} (MeCO-Gly-NHMe and MeCO-Ala-NHMe, respectively) are also shown in Figure 1a as a reference for proteinogenic amino acids. As it can be seen, the effect of the replacement of the amide hydrogen of Ala by a methyl group on the low-energy minimum ($\beta_{DL}$) is very significant, with the minimum being annihilated. The disappearance of this minimum should be attributed to the repulsive interactions produced by the confrontation of the $N$-methyl group and the oxygen atom of C=O moiety.

The $N$-hydroxylation and $N$-amination of peptides have been used to impart conformational rigidity through the formation of intramolecular hydrogen bonds with neighboring C=O groups.\textsuperscript{29,30} Besides its hydrogen-bonding abilities, the $N$-hydroxy group has been proven able to chelate metal ions, which may be of use for the design of peptides which can bind specifically to proteins containing metals in the active site.\textsuperscript{29} The Ramachandran map in Figure 1b shows all the energy minima reported for the $N$-acetyl-$N'$-methylamide derivatives of $N$-hydroxyalanine\textsuperscript{23} and $N$-aminoglycine\textsuperscript{24} which are compared to those of Ala and Gly,\textsuperscript{28} respectively. As can be seen, replacement of the amide hydrogen by a hydroxyl or amino group entails drastic conformational changes. The most remarkable ones can be summarized as follows: (i) the lowest energy minimum of Ala ($\gamma_L$) evolves towards $\epsilon_L$ upon $N$-hydroxylation; (ii) the extended $\beta_{DL}$ conformation of Ala is not a minimum in $N$-hydroxyalanine; (iii) the steric interactions induced by the $N$-amino group produces a destabilization of the $\beta_{DL}$ conformation in $N$-aminoglycine; and (iv) $N$-amination of Gly leads to the characterization of new local minima located at the $\epsilon_L$, $\epsilon_D$, $\delta_D$, and $\delta_L$ regions of the Ramachandran map.
**Thio-α-amino acids**

Within the surrogates used to replace the amide bond of peptides, the thioamide functionality, in which the carbonyl oxygen is replaced by a sulfur atom, is amongst the most synthetically accessible. This modification not only affects the proteolytic stability and bioactivity of peptides; it also produces significant changes in the secondary structure relative to the parent peptides. Both amides and thioamides are planar and exhibit high energy barriers for rotation about the C–N bond. However, the length of the C=S bond in thioamides (~0.45 Å larger than the C=O linkage in amides) and the distinct stereoelectronic features of the sulfur atom (its van der Waals radius is 32% larger than that of oxygen and sulfur is less electronegative) affect the hydrogen-bonding abilities of the thioamide bond and translate into a reduced proton-acceptor character.

The influence of this amide-thioamide replacement on the intrinsic conformational properties has been studied only for Gly and Ala (Table 1). It has been found that the sulfur atom reduces drastically the conformational flexibility of thioglycine when compared to the coded amino acid. Thus, the only minimum energy conformation identified for thioglycine, which is two-fold degenerate, corresponds to a γ-turn motif (Figure 2). Therefore, the minima found for Gly at the βDL and α regions of the Ramachandran map are annihilated upon replacement of oxygen by sulfur. Figure 2 shows a Ramachandran map with the five minima found for thioalanine. The three minima detected for Ala were also identified as minimum energy conformations for thioalanine and exhibit similar relative energies. Thus, γL is the most stable conformation for the two amino acids while the relative energies of the γD and βDL structures differ in 0.6 and 1.4 kcal/mol, respectively. The relative energies of the two additional minima characterized for thioalanine, αL and δD, are above 4 kcal/mol. Accordingly, the changes produced in Ala by the presence of the sulfur atom are considerably less important than those observed for Gly.

**Diamides and diacids used to build retropeptide bonds**

In the last decades, a great deal of effort has been devoted to the chemistry of the so-called retropeptides or partially modified retropeptides, in which the direction of all or some peptide bonds is reversed, i.e. NH–CO instead of CO–NH (Scheme 1). When the stereochemistry of one or more amino acids of the reversed segment is inverted, the resulting pseudopeptide is termed retroinverso. 1,1-Dicarboxylic acids (malonic acid derivatives) are used to initiate the retropeptide sequence, while the direction of the peptide bond can be restored by incorporating an 1,1-diamine unit (Scheme 1).

Reversal of the peptide bond has been used for designing bioactive pseudopeptides, for engineering and mirroring protein secondary structural motifs, and for the construction of synthetic (bio)polymers mimicking the structures found in fibrous proteins. NCAD includes the conformational preferences of 5 diacids and 4 diamines (Table 2) that are surrogates of the following α-amino acids: Gly, Ala, Val, α-aminoisobutyric acid (Aib), and dehydroalanine (ΔAla). Considering that the φ,ψ dihedral angles in α-amino acids define the rotation about the Cα-N and Cα-C(O) bonds, respectively, the conformation of the diacids is given by ψ1,ψ2 and that of the diamines by φ1,φ2. Scheme 2 compares the backbone dihedral angles of the N-acetyl-N′-methylamide derivatives, N,N′-dimethylamide derivatives, and N,N′-diacetyl derivatives of the α-amino acids, diamines and diacids, respectively.

The Ramachandran map in Figure 3a shows the location of the minimum energy conformations found for the N,N′-dimethylamide derivatives of the diacid surrogates of Gly, Ala, Aib, Val, and ΔAla, hereafter denoted dc-Gly, dc-Ala, dc-Aib, dc-Val, and dc-ΔAla, respectively (Table 2). The minimum energy conformations of the dc-Ala and dc-
Aib derivatives were not reported in the literature and have been calculated in this work at the B3LYP/6-31G(d) level (the dihedral angles and relative energies are given in the Supporting Information). The resulting conformations have been used for engineering a 3_10-helical retropeptide (see below).

Due to molecular symmetry, structures characterized by dihedral angles $\psi_1,\psi_2; -\psi_1,-\psi_2; \psi_2,\psi_1; \text{and } -\psi_2,-\psi_1$ are energetically degenerate (Figure 3a). The lowest energy conformation was found at $\psi_1,\psi_2 \approx 45^\circ,120^\circ$ for all the diacids but dc-$\Delta$Ala, in which case, the global minimum appears at $\psi_1,\psi_2 \approx 13^\circ,142^\circ$. This small difference should be attributed to the hybridization state of the $\alpha$ carbon atom, namely sp$^2$ for dc-$\Delta$Ala and sp$^3$ for dc-Gly, dc-Ala, dc-Aib, and dc-Val. These non-symmetric conformations are very similar to those found in the crystal state for malonamide derivatives. On the other hand, Figure 3a shows that the minima of dc-$\Delta$Ala with relative energies larger than 2 kcal/mol (i.e. minima with $\psi_1,\psi_2= \pm 25.0^\circ, \pm 25.0^\circ$ and $\pm 126.0^\circ, \pm 126.0^\circ$) are located in the map diagonals. However, the lowest energy conformation remains asymmetric ($\psi_1 \neq \psi_2$), even though this diacid is chemically symmetric.

Figure 3b shows the energy minima found for the $N,N'$-diacetyl derivatives of the Gly, Ala, Aib and Val diamine surrogates, hereafter denoted dm-Gly, dm-Ala, dm-Aib, and dm-Val, respectively. The minimum energy conformations of the dm-Aib derivative have been calculated in this work at the B3LYP/6-31G(d) level (see the Supporting Information). For the reason mentioned above, structures with $\phi_1,\phi_2; -\phi_1,-\phi_2; \phi_2,\phi_1; \text{and } -\phi_2,-\phi_1$ are energetically equivalent, as reflected by the map symmetry. As can be seen, the lowest energy conformations of the diamines are located in the $\alpha$ region at $\phi_1,\phi_2 \approx 60^\circ,60^\circ$. In addition, dm-Aib shows a low energy minima destabilized by 1.0 kcal/mol at $\phi_1,\phi_2 = 49^\circ,173^\circ$. Minima with relative energies higher than 2 kcal/mol are essentially located in the map diagonal as well as in the regions labeled as $\epsilon$ and $\delta$. As expected, the $\gamma$ regions are scarcely populated for diamines and diacids derivatives (Figure 3), whereas they are found to be preferred by the $N$-acetyl-$N'$-methylamide derivatives of proteinogenic amino acids. This is because in the latter compounds, the $\gamma$ conformations are stabilized by an intramolecular hydrogen bond defining a seven-membered ring, whereas this interaction is not possible in diamines and diacids.

**In silico molecular engineering of a 3_10-helical motif using NCAD: A test case**

In this section, we present a computational design study aimed at constructing non-symmetric retropeptides exhibiting a 3_10-helix conformation by using information extracted from NCAD. This is not intended to provide a deep and rigorous investigation of the resulting retropeptides, but to illustrate the potential utility of the database in applications involving peptide and protein engineering through a test case.

It is well known that Aib-based peptides have a remarkable tendency to adopt helical conformations both in solution and in the solid state. Helices of the 3_10 type are overwhelmingly formed by Aib homopeptides, independently of the environmental conditions, whereas small peptides containing a mixture of Aib and proteinogenic $\alpha$-amino acids are found to accommodate either $\alpha$- or 3_10-helices depending mainly on the peptide size and the Aib content.

In an earlier work, we used retropeptide bonds to enhance the stability of the 3_10-helix with respect to the $\alpha$-helix in peptides containing Aib. Specifically, Aib-containing homoretropeptides (hereafter denoted rAib-$n$, where $n$ indicates the number of residues) cannot form the intramolecular hydrogen-bonding network that stabilizes the $\alpha$-helix ($i \rightleftharpoons i$...
and this results in unfavorable interactions between the C=O groups of residues \(i\) and \(i+4\). In contrast, the molecular architecture of \(r\text{Aib}-n\) (Scheme 3) in which diacid and diamine surrogates of Aib are alternated, \((dm\text{-Aib–dc\text{-Aib})}\text{n/2}\), is fully compatible with the hydrogen-bonding scheme of the 3\text{10}\text{-helix (}i\leftarrow i+3). These features are schematically depicted in Figure 4, whereas Figure 5 shows the 3\text{10}\text{-helix adopted by } r\text{Aib}-n.

\(r\text{Aib}-n\) are symmetric molecules, \textit{i.e.} the two ends of the peptide are identical, which may be a limitation in some recognition processes. Accordingly, we used the information stored in the NCAD database to search for nc-aa that, after targeted replacements, facilitate the recognition process but retaining as much as possible the stability of 3\text{10}\text{-helix. The procedure followed for this purpose, which is schematized in Figure 6, can be summarized as follows:}

\begin{enumerate}
  \item NCAD was used to search for the nc-aa able to adopt a helical conformation with \(\phi, \psi = +60^\circ \pm 25^\circ, +40^\circ \pm 25^\circ\) or \(\phi, \psi = -60^\circ \pm 25^\circ, -40^\circ \pm 25^\circ\). The number of nc-aa that satisfied this initial criterion was 24.
  \item From the 24 nc-aa identified in (i), NCAD allowed identify those in which the helix correspond to the lowest energy conformation. This criterion was satisfied by 6 nc-aa including the diamine analogues of Ala, Aib and Val (dm-Ala, dm-Aib and dm-Val in Table 2). Obviously, dm-Aib was eliminated from the list since it is contained in the parent retropeptide.
  \item From the 5 nc-aa provided by (ii), we looked for those with lower conformational flexibility. For this purpose, we used NCAD to exclude those nc-aa in which the energy gap between the helical global minimum and the most stable local minimum was lower than 2 kcal/mol. Unfortunately, no nc-aa was eliminated using this criterion since in all cases the gap was larger than 2 kcal/mol. However, the information provided by the database was considerably useful since it allowed ascertain that the lowest (2.4 kcal/mol) and highest (4.3 kcal/mol) value of this energy gap were for dm-Val and dm-Ala, respectively.
  \item Inspection to the chemical structures stored in NCAD indicated that 2 of 5 nc-aa selected in (ii) and (iii) involved positively charged side groups \textit{i.e.} \((1S,2R)-\text{ and } (1S,2S)-1\text{-amino-2-}(\text{guanidinomethyl})\text{-cyclopentanecarboxylic acids}\). These two nc-aa were eliminated in order to avoid undesirable repulsive electrostatic interactions, \textit{i.e.} \(r\text{Aib}-n\) does not involve negatively charged groups able to form attractive interactions with the positively charged groups. On the other hand, another nc-aa \textit{i.e.} \((S)-\text{N-hydroxy-2-aminopropionic acid}\] involves an N-hydroxylation that was expected to induce drastic distortions in the helix through alterations in the geometry of the hydrogen bonding network, being also removed from the list. Accordingly, dm-Ala and dm-Val were considered by NCAD as the more suitable candidates for the proposed substitutions.
\end{enumerate}

The information stored in NCAD indicates that two different strategies can be followed to design a non-symmetric retropeptide based on \(r\text{Aib}-n\) through a single mutation without decreasing significantly the stability of the 3\text{10}\text{-helix. The first strategy requires the replacement of } dm\text{-Aib by } dm\text{-Ala (i.e. removal of a methyl group in a diamide). The second involves the replacement of } dm\text{-Aib by the corresponding Val analogue. Although the helical conformation is energetically favored for the diamines of both Ala and Val, the larger conformational flexibility of the Val derivatives and the larger steric interactions expected for the isopropyl group in comparison to the methyl substituent suggest that the replacement of } dm\text{-Aib by the corresponding Ala analogue is more appropriate. Accordingly, the first strategy was selected for further investigation.}
We first examined the most appropriate position to carry out the targeted replacement. For this purpose, three mutants of a retropeptide containing 12 residues, rAib-12 [i.e. (dm-Aib – dc-Aib)₆], were considered, their sequences being:


The three retropeptides were built up in a 3₁₀-helix conformation and their molecular geometries were submitted to complete geometry optimizations using the B3LYP quantum mechanical method combined with the 6-31G(d) basis set. The stability of the three mutants under study was similar. Thus, the diamine-2 was more stable than diamines-1 and -3 by only 0.3 and 0.1 kcal/mol, respectively, indicating the lack of preference towards a given position for the mutation.

Comparison of the backbone dihedral angles obtained for the three optimized mutants to those derived from geometry optimization of rAib-12 at the same theoretical level indicates that the conformational distortion introduced by dm-Ala is almost negligible. Thus, the largest distortion was observed for diamine-1, where the φ₁ dihedral of dm-Ala differs in 16.2° from the corresponding angle in rAib-12. For the rest of the mutants, no deviation was found to exceed 10°. Similarly, the end-to-end distance remained almost identical for the mutants and the parent peptide, the largest difference being 0.08 Å for diamine-1. These results confirm that the helical stability of rAib-12 is retained in the three mutants investigated. In order to quantitatively evaluate the effect of the mutation in the energetic stability of the 3₁₀-helix, the following isodesmic reaction has been considered:

\[ r\text{Aib-12} + \text{CH}_4 \rightarrow \text{mutant} + \text{CH}_3\text{CH}_3 \]  

Thus, the loss of stability of the helix (Δᵦ) was defined according to:

\[ \Deltaᵦ = E(\text{mutant}) + E(\text{CH}_3\text{CH}_3) - E(r\text{Aib-12}) - E(\text{CH}_4) \]  

where E refers to the energy of the different species. The Δᵦ values obtained for diamines-1, -2 and -3 are 1.7, 1.4 and 1.5, respectively. Comparison between these Δᵦ values and the relative stabilities mentioned above for the three mutants indicates that breaking the molecular symmetry of rAib-12 by replacement of one dm-Aib by dm-Ala produces a destabilization of 1.4 kcal/mol, independently of the position.

Inspection of the dipole moments calculated for the mutants indicates that incorporation of dm-Ala produces a very small variation with respect to rAib-12 (µ= 0.225 Debyes), the largest variation being of only 0.092 Debyes (diamine-1). Accordingly, the small destabilization detected by the substitution should not be attributed to electrostatic changes in the helix. However, a detailed analysis of the intramolecular distances may help explain this feature. Specifically, inspection of rAib-12 indicates that the methyl groups of the dm-Aib residues form a weak interaction with the oxygen atom of a proximal O=C moiety, which in turn, is involved in an intramolecular hydrogen bond with a neighboring amide hydrogen. This van der Waals interaction is illustrated in Figure 7, which also indicates that the average C-H\cdots O=C distance is 2.75±0.01 Å. Evaluation of such a distance for the three mutants under study shows that removal of the methyl group induces a variation that becomes significantly larger at the mutated positions (Figure 7). Thus, the distance increases by 0.32±0.02 in the retropeptides mutated at the diamines, with respect to that found in rAib-12, and this may contribute to the loss of stability. It should be mentioned that the
B3LYP method does not treat appropriately the dispersion effects. Therefore, the results described in this section illustrate the use of the database and should be considered only from a qualitative point view.

Conclusions
In this work, NCAD has been extended through the incorporation of residues that may be used as building blocks for the construction of peptides with modified peptide bonds. The information related to N-substituted α-amino acids, thio-α-amino acids, and diamines and diacids used to build retropeptides described in the literature has been collected and integrated into the database. The conformational propensities of the new α-amino acid surrogates incorporated into NCAD have been compared to those of the corresponding coded residues. We also present an example illustrating the utility of NCAD and the usefulness of the new amino acid surrogates in peptide engineering. More specifically, the information contained in NCAD has been used to propose targeted replacements able to break the molecular symmetry of a retropeptide sequence that adopts the 3_10-helix.

Supplementary Material
Refer to Web version on PubMed Central for supplementary material.

Acknowledgments
Computer resources were generously provided by the Centre de Supercomputació de Catalunya (CESCA). Financial support from Ministerio de Ciencia e Innovación - FEDER (grants CTQ2007-62245, CTQ2008-00423-E/BQU, and CTQ2010-17436), Generalitat de Catalunya (research group 2009 SGR 925; XRQTC), ICREA Academia (prize for excellence in research to C.A.), and Gobierno de Aragón (research group E40) is gratefully acknowledged. This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the view of the policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References


Figure 1.
Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for the N-acetyl-N\(^{'-}\)-methylamide derivatives of the N-modified \(\alpha\)-amino acids included in NCAD (see Table 1): (a) \(N\)-methylalanine (gray squares); and (b) \(N\)-aminoglycine (black squares) and \(N\)-hydroxyalanine (gray squares). The minimum energy conformations of Gly (black triangles) and Ala (grey triangles) have been included for comparison. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.
Figure 2.
Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for the N-acetyl-N'-methylamide derivatives of the thio-α-amino acids included in NCAD (see Table 1): thioglycine (black squares) and thioalanine (gray squares). The minimum energy conformations of Gly (black triangles) and Ala (grey triangles) have been included for comparison. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.
Figure 3.
Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for (a) the \(N,N'\)-dimethylamide derivatives of the diacids and (b) the \(N,N'\)-diacetyl derivatives of the diamines surrogates of \(\alpha\)-amino acids included in NCAD (see Table 2). (a) Diacids: dc-Gly (black squares), dc-Ala (black circles), dc-Aib (gray triangles), dc-Val (gray diamonds) and dc-\(\Delta\)Ala (empty circles); (b) Diamines: dm-Gly (black squares), dm-Ala (black circles), dm-Aib (gray triangles) and dm-Val (gray diamonds). The nomenclature used for the different diacids and diamines is discussed in the text. Minima with relative energies lower and higher than 2.0 kcal/mol are represented by large and small symbols, respectively.
Figure 4.
Hydrogen bonding network for the α- and 3_{10}-helices (\(i\leftarrow i+4\) solid line and \(i\leftarrow i+3\) dashed line, respectively) of Aib-\(n\) and rAib-\(n\).
Figure 5.
Side and top views of the $3_{10}$-helix adopted by rAib-$n$. 
Figure 6.
Search performed with the NCAD database to select nc-aa able to break the symmetry in rAib-n without alter significantly the stability of the 3_{10}-helix (see text).

Proteins. Author manuscript; available in PMC 2012 June 1.
Figure 7.
Detail about the C-H⋯O=C interaction (arrows) in the 3_{10}-helix of rAib-12 and diamine-1. Averaged H⋯O distances are indicated.
Scheme 1.

Parent peptide

Retropeptide

diamine

diacid

Proteins. Author manuscript; available in PMC 2012 June 1.
Scheme 2.
Scheme 3.
Table 1

*N*-substituted α-amino acids and thio-α-amino acids stored in NCAD. The systematic name (and the common name), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid.

<table>
<thead>
<tr>
<th>Systematic name (common name)</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)-N-methyl-2-aminopropionic acid (N-methyl-L-alanine)</td>
<td><img src="image1" alt="Structure" /></td>
<td>22</td>
</tr>
<tr>
<td>(S)-N-hydroxy-2-aminopropionic acid (N-hydroxy-L-alanine)</td>
<td><img src="image2" alt="Structure" /></td>
<td>23</td>
</tr>
</tbody>
</table>

Proteins. Author manuscript; available in PMC 2012 June 1.
<table>
<thead>
<tr>
<th>Systematic name</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydrazinoacetic acid (N-aminoglycine)</td>
<td><img src="image1" alt="Structure Image" /></td>
<td>24</td>
</tr>
<tr>
<td>aminothioacetic acid (thioglycine)</td>
<td><img src="image2" alt="Structure Image" /></td>
<td>43</td>
</tr>
<tr>
<td>(5)-2-aminothiopropionic acid (L-thioalanine)</td>
<td><img src="image3" alt="Structure Image" /></td>
<td>43</td>
</tr>
</tbody>
</table>
Table 2

Diacids and diamines analogues of α-amino acids stored in NCAD. The systematic name (and the abbreviated name used in this work), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid.

<table>
<thead>
<tr>
<th>Systematic name (abbreviated name)</th>
<th>Structure</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>malonic acid (dc-Gly)</td>
<td><img src="structure1.png" alt="malonic acid structure" /></td>
<td>61</td>
</tr>
<tr>
<td>methylmalonic acid (dc-Ala)</td>
<td><img src="structure2.png" alt="methylmalonic acid structure" /></td>
<td>This work</td>
</tr>
<tr>
<td>dimethylmalonic acid (dc-Aib)</td>
<td><img src="structure3.png" alt="dimethylmalonic acid structure" /></td>
<td>This work</td>
</tr>
<tr>
<td>isopropylmalonic acid (dc-Val)</td>
<td><img src="structure4.png" alt="isopropylmalonic acid structure" /></td>
<td>62</td>
</tr>
<tr>
<td>methylenemalonic acid (dc-ΔAla)</td>
<td><img src="structure5.png" alt="methylenemalonic acid structure" /></td>
<td>64</td>
</tr>
<tr>
<td>Systematic name (abbreviated name)</td>
<td>Structure</td>
<td>Ref.</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>methylenediamine (dm-Gly)</td>
<td><img src="image1" alt="Structure image" /></td>
<td>60</td>
</tr>
<tr>
<td>1,1-ethylenediamine (dm-Ala)</td>
<td><img src="image2" alt="Structure image" /></td>
<td>60</td>
</tr>
<tr>
<td>2,2-propylenediamine(dm-Aib)</td>
<td><img src="image3" alt="Structure image" /></td>
<td>This work</td>
</tr>
<tr>
<td>1,1-isobutylenediamine(dm-Val)</td>
<td><img src="image4" alt="Structure image" /></td>
<td>62</td>
</tr>
</tbody>
</table>