

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/44585536>

# NCAD, a Database Integrating the Intrinsic Conformational Preferences of Non-Coded Amino Acids

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2010

Impact Factor: 3.3 · DOI: 10.1021/jp102092m · Source: PubMed

CITATIONS

6

READS

30

## 11 AUTHORS, INCLUDING:



**Guillem Revilla-lopez**

Universität Regensburg

31 PUBLICATIONS 142 CITATIONS

SEE PROFILE



**Juan Torras**

Polytechnic University of Catalonia

53 PUBLICATIONS 461 CITATIONS

SEE PROFILE



**David Zanuy**

Polytechnic University of Catalonia

109 PUBLICATIONS 1,766 CITATIONS

SEE PROFILE



**Carlos Alemán**

Polytechnic University of Catalonia

462 PUBLICATIONS 6,028 CITATIONS

SEE PROFILE

# NCAD, a Database Integrating the Intrinsic Conformational Preferences of Non-Coded Amino Acids

Guillem Revilla-López,<sup>†</sup> Juan Torras,<sup>‡</sup> David Curcó,<sup>§</sup> Jordi Casanovas,<sup>||</sup> M. Isabel Calaza,<sup>⊥</sup> David Zanuy,<sup>†</sup> Ana I. Jiménez,<sup>⊥</sup> Carlos Cativiela,<sup>⊥</sup> Ruth Nussinov,<sup>#,▽</sup> Piotr Grodzinski,<sup>○</sup> and Carlos Alemán<sup>\*,†,◆</sup>

*Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain, Departament d'Enginyeria Química, EUETII, Universitat Politècnica de Catalunya, Pça Rei 15, Igualada 08700, Spain, Departament d'Enginyeria Química, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona E-08028, Spain, Departament de Química, Escola Politècnica Superior, Universitat de Lleida, c/Jaume II n°69, Lleida E-25001, Spain, Departamento de Química Orgánica, Instituto de Ciencia de Materiales de Aragón, Universidad de Zaragoza - CSIC, 50009 Zaragoza, Spain, Basic Science Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI, Frederick, Maryland 21702, Department of Human Genetics, Sackler Medical School, Tel Aviv University, Tel Aviv 69978, Israel, Alliance for Nanotechnology in Cancer, National Cancer Institute, Bethesda, Maryland 20892, and Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Campus Sud, Edifici C', C/Pasqual i Vila s/n, Barcelona E-08028, Spain*

*Received: March 8, 2010; Revised Manuscript Received: April 8, 2010*

Peptides and proteins find an ever-increasing number of applications in the biomedical and materials engineering fields. The use of non-proteinogenic amino acids endowed with diverse physicochemical and structural features opens the possibility to design proteins and peptides with novel properties and functions. Moreover, non-proteinogenic residues are particularly useful to control the three-dimensional arrangement of peptidic chains, which is a crucial issue for most applications. However, information regarding such amino acids—also called non-coded, non-canonical, or non-standard—is usually scattered among publications specialized in quite diverse fields as well as in patents. Making all these data useful to the scientific community requires new tools and a framework for their assembly and coherent organization. We have successfully compiled, organized, and built a database (NCAD, Non-Coded Amino acids Database) containing 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, conformational propensities established experimentally, and applications. The architecture of the database is presented in this work together with the first family of non-coded residues included, namely,  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids. Furthermore, the NCAD usefulness is demonstrated through a test-case application example.

## Introduction

The repertoire of amino acids currently available for application in life and materials sciences is rapidly expanding. Besides the 20 genetically coded  $\alpha$ -amino acids contained in proteins, more than 700 different amino acids have been found in Nature<sup>1</sup> (most of them, also  $\alpha$ -amino acids), and many others have been synthesized by organic chemists.<sup>1a,2</sup> These residues are joined under the common name of *non-proteinogenic* or *non-coded* amino acids (nc-aa). Although they are rigorously excluded from ribosomally synthesized native peptides and proteins, several naturally occurring peptides (produced non-ribosomally) have been found to contain nc-aa.<sup>2i,3</sup> Additionally, they are increas-

ingly used to improve the pharmacological profile of natural peptides endowed with biological activity<sup>4</sup> (to confer resistance against enzymatic degradation, enhance membrane permeability, or increase selectivity and affinity for a particular receptor). Nc-aa have also led to the development of efficient drugs of non-peptidic nature.<sup>5</sup> Moreover, recent advances in biotechnology have paved the way for protein engineering with nc-aa.<sup>6</sup> Thus, proteins containing residues with fluorescent,<sup>7</sup> redox-active,<sup>8</sup> or photosensitive groups<sup>7c,9</sup> or other side chains endowed with specific chemical reactivity<sup>10</sup> or spectroscopic properties<sup>11</sup> may serve as biosensors and spectroscopic or biophysical probes, as well as to construct systems for drug delivery and diagnosis through imaging useful in medicine.<sup>6–11</sup> Nc-aa have also found application in nanobiology to promote the self-assembly of nanostructures.<sup>12</sup> Other remarkable applications of nc-aa include the development of bioinspired synthetic organic polymers that emulate the shape and properties of natural peptides and proteins.<sup>13</sup>

In spite of the great significance and potential utility of nc-aa in modern biology, biomaterials engineering, and medicine, structural information is only available for a limited number of them. Specifically, the intrinsic conformational preferences of nc-aa, which are typically obtained by sampling the potential

\* Corresponding author. E-mail: carlos.aleman@upc.edu.

<sup>†</sup> E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya.

<sup>‡</sup> EUETII, Universitat Politècnica de Catalunya.

<sup>§</sup> Universitat de Barcelona.

<sup>||</sup> Universitat de Lleida.

<sup>⊥</sup> Universidad de Zaragoza - CSIC.

<sup>#</sup> Center for Cancer Research Nanobiology Program.

<sup>▽</sup> Tel Aviv University.

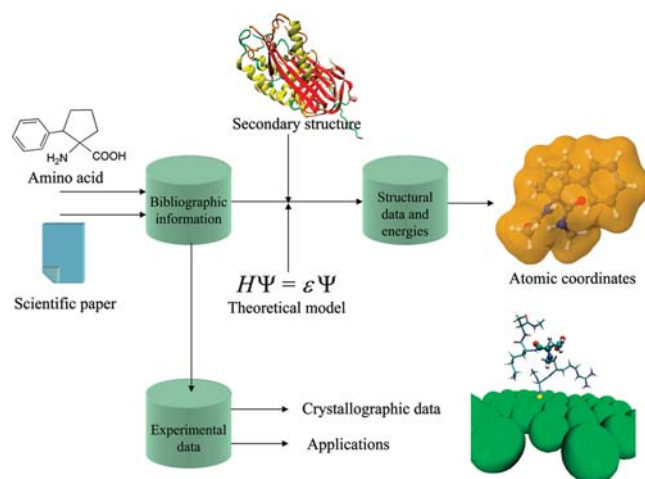
<sup>○</sup> National Cancer Institute.

<sup>◆</sup> Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya.

energy hypersurface of small peptide systems incorporating them through sophisticated quantum mechanical calculations at the *ab initio* or Density Functional Theory (DFT) levels in the absence of external forces, have been reported for only several tenths of these compounds. Knowledge of such intrinsic conformational preferences is however essential to understand the features that distinguish these nc-aa from the coded ones and, therefore, for a complete and satisfactory exploitation of their capabilities. Structural data yielded by crystallographic or spectroscopic studies on peptide sequences incorporating nc-aa are also 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 the external packing forces, the 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. Indeed, the quantum mechanical calculations describing the intrinsic conformational preferences of these compounds are typically reported in physical chemistry journals, while their synthesis appears in organic chemistry publications. Crystallographic and spectroscopic studies of small peptide sequences containing nc-aa are usually developed by organic and peptide chemists and 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. This highly dispersed information together with the wide and diverse potential applications of nc-aa made us realize that the design of a simple informatics tool integrating the contributions supplied by the different research fields would facilitate the use of these compounds in many practical applications. To this aim, we have built a database of nc-aa containing the fundamental conformational descriptors and relevant bibliographic information about experimental studies and already developed practical applications. This user-friendly online facility is intended as an aid to the development of new applications by combining the structural descriptors with thermodynamics criteria for a proper selection of nc-aa compatible with the user-defined requirements.

In this work, we present NCAD (Non-Coded Amino acids Database), 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 life and materials sciences. The database integrates structural and energetic descriptors of those nc-aa whose intrinsic conformational properties have been previously studied using *ab initio* or DFT quantum mechanical calculations and reported in the literature. The information about NCAD presented in the following sections is organized as explained next. First, the most relevant aspects of the database are described, such as structure, contents, technical features, and relationships allowed among the different descriptors. Next,  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids, being the first family of nc-aa integrated into the database, are presented. The conformational preferences of these compounds, which have been the subject of extensive study by our research groups, are examined and discussed using the overall information captured in the database. Finally, the utility and applicability of the database is illustrated with a simple example. Specifically, conformationally restricted analogues of the neuropeptide Met-enkephalin are designed through targeted replacements with nc-aa, using the information incorporated in the database.



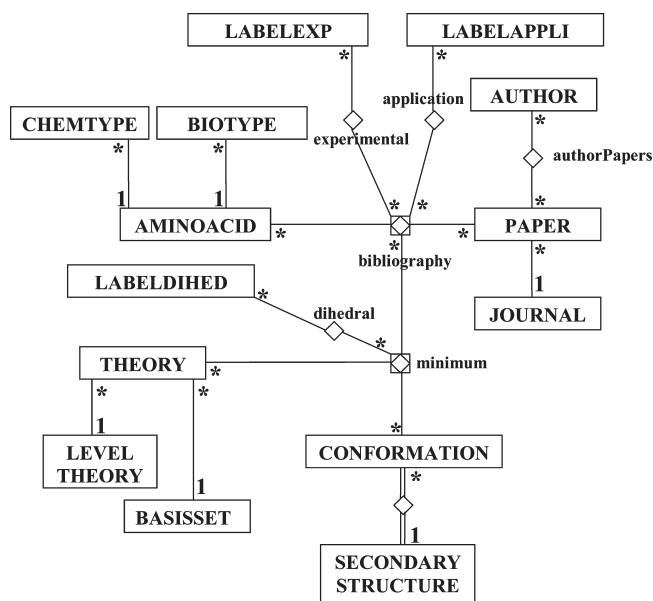
**Figure 1.** Schematic graphic algorithm showing the structure of the NCAD database.

### NCAD: Contents, Tools, and Technical Features

Figure 1 portrays the NCAD contents. The information contained in the database for each nc-aa is the following:

(a) A complete description (dihedral angles, three-dimensional structure, relative energy, etc.) of all the minimum energy conformations found for each amino acid using *ab initio* or DFT quantum mechanical calculations. It should be noted that such theoretical studies are carried out on small peptide systems of general formula RCO-Xaa-NHR' (Xaa = amino acid; R and R' = Me or H). When several theoretical studies were available in the literature for a particular amino acid, the information was extracted from that using the highest level of calculation. Conformational studies based on calculations less accurate than *ab initio* or DFT methods, i.e., molecular mechanics based on classical force fields, semiempirical procedures, etc., have not been considered for NCAD. It is worth noting that only in a few cases are the atomic coordinates of the minimum energy conformations supplied by the authors as Supporting Information of the original article. Accordingly, to incorporate such coordinates into the database for all nc-aa, all minimum energy conformations for which coordinates were not available have been recalculated using the same theoretical level as in the original study. Although this is an arduous and extremely time-consuming task, we considered it essential for the usefulness of the database. Throughout the theoretical works describing the conformational propensities of nc-aa, different nomenclature systems are used to term the energy minima located. We have unified them so that only the nomenclature introduced by Perczel et al.<sup>14</sup> is used in NCAD to identify the backbone conformation of the different minima characterized for each amino acid. According to it, nine different conformations can be distinguished in the potential energy surface  $E = E(\phi, \psi)$  of  $\alpha$ -amino acids,<sup>14</sup> namely,  $\gamma_D$ ,  $\delta_D$ ,  $\alpha_D$ ,  $\epsilon_D$ ,  $\beta_{DL}$ ,  $\epsilon_L$ ,  $\alpha_L$ ,  $\delta_L$ , and  $\gamma_L$ . This information is particularly useful to check the compatibility of the minima energetically accessible to a particular nc-aa with the secondary structure motifs typically found in peptides and proteins (helices, sheets, turns, etc.).

(b) Bibliographic information related to experimental data, when available. Selected references describing the synthesis and characterization of each particular amino acid in its free form, as a salt, or with the amino group adequately protected for use in peptide synthesis are given. Also, references providing experimental information about the conformational preferences of the amino acid are included in NCAD. These concern mainly



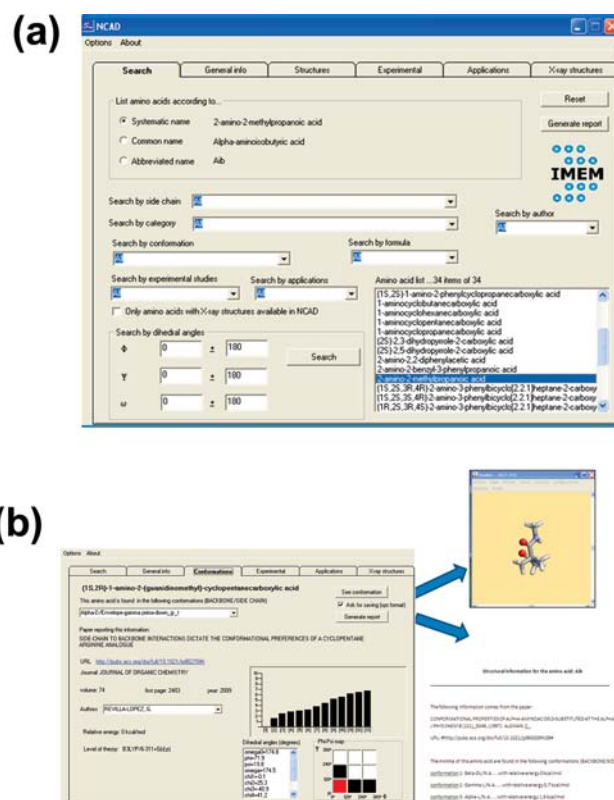
**Figure 2.** Chen's diagram showing the relationships among the entities in NCAD.

structural studies of peptides incorporating nc-aa using either spectroscopic or crystallographic techniques. Furthermore, in selected cases, atomic coordinates extracted from X-ray crystal structures have also been included in the database. Such three-dimensional structures allow a direct comparison with the minimum energy conformations predicted by theoretical calculations. Upon request and permission of the authors, we plan to include in NCAD examples of crystal structures of peptide sequences incorporating different nc-aa. This does not intend to be comprehensive since we are aware that the crystal structures of peptides and proteins containing nc-aa are already stored in specialized databases, like the Cambridge Structural Database<sup>15</sup> (CSD) and the Protein Data Bank<sup>16</sup> (PDB).

(c) Applications reported for each particular nc-aa divided into two main categories, namely, those related to the biological properties of the amino acid (or compounds incorporating it) and applications in materials science. The most relevant publications in each field, either papers in scientific journals or patents, are given.

As commented above, NCAD has been conceived to contain all those nc-aa whose intrinsic conformational propensities have been determined through *ab initio* or DFT quantum mechanical calculations. Such amino acids may be chiral or not. In the former case, two enantiomeric forms do exist for a given amino acid. Obviously, the conformational preferences of only one enantiomer have been calculated and are included in the database since those of the enantiomeric species may be obtained by simply changing the signs of the dihedral angles. The bibliographic information concerning synthesis, spectroscopic characterization, experimental structural properties, and applications included in the database covers the two enantiomers as well as the racemic form, when applicable.

Figure 2 shows the entity–relationship diagram<sup>17</sup> (ERD) of NCAD. This diagram summarizes the conceptual representation of the information stored in the database, as well as their inter-relational scheme. The entities are shown in squares; the relationships are lines between boxes; and their multiplicities are displayed using numbers and stars following ERD standards. All the experimental and theoretical information in the database has been extracted from relevant literature sources, and accordingly, all data available in NCAD for a given amino acid are



**Figure 3.** Interface system used to interact with the NCAD database. (a) Window used to perform the search. (b) Window showing details about the theoretical study at the highest quantum mechanical level reported in the literature for the nc-aa selected; the Rasmol 3D-view of a selected minimum energy conformation and the report are also displayed (both are generated through internal calls of the interface upon the user's request).

connected with the bibliographic references. Labels (*labelexp*, *labelappli* entities) allow the identification of the applications described for each amino acid as well as the experimental data. The database also stores the atomic coordinates of the minimum energy conformations (*minimum*) predicted by a particular quantum mechanical method (*theory*), which have been extracted from the literature (*bibliography*), and its correspondence with secondary structure motifs (*conformation-secondary structure*). Furthermore, internal coordinates, such as dihedral angles (*dihedrals*) for both the backbone and the side chain, can be easily obtained from the database for each minimum energy conformation of a given amino acid.

NCAD can be easily used by non-experts in informatics since the interface designed applies a simple windows system. This system provides access to six folders. The first one (Figure 3a) is used to define the criteria (one or several, i.e., .AND. or .OR.) applied to the search. These may be chosen among the following: (i) the chemical nature of the side chain, i.e., aromatic, aliphatic, charged, or polar (non-charged); (ii) the category, which corresponds to the structural feature distinguishing a particular family of nc-aa, e.g.,  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids, dehydro  $\alpha$ -amino acids, N-substituted  $\alpha$ -amino acids,  $\omega$ -amino acids, etc.; (iii) the molecular formula of the amino acid; (iv) the type of backbone conformation located among the energy minima, according to the nomenclature established by Perczel et al.,<sup>14</sup> i.e.,  $\gamma_D$ ,  $\delta_D$ ,  $\alpha_D$ ,  $\epsilon_D$ ,  $\beta_{DL}$ ,  $\epsilon_L$ ,  $\alpha_L$ ,  $\delta_L$ , and  $\gamma_L$ ; (v) the  $(\phi, \psi)$  backbone dihedral angles corresponding



to such energy minima, with both the values and the tolerance being specified by the user; (vi) the name of the researcher who performed the theoretical or experimental investigation (author); (vii) the experimental studies reported in the literature relative to synthesis and characterization or to conformational properties; and (ix) the applications (in medicine or materials science) described.

The amino acids fulfilling all the selected criteria are displayed in a list (Figure 3a) (by default, when no criterion is specified, all residues included in NCAD are listed), and a report with the search criteria and results may be generated. The user can select three different nomenclatures to indicate the amino acid names: systematic, common, or abbreviated (usually resembling the three-letter code used for proteinogenic amino acids), e.g., 2-amino-2-methylpropanoic acid,  $\alpha$ -aminoisobutyric acid, or Aib, respectively. It should be noted that the common and abbreviated names are not available for all residues.

Selection of an amino acid from this list allows access to all the information stored in the database for this particular residue, which is presented in the remaining five folders. The "general info" folder provides the chemical description of the amino acid, including a graphical scheme of its chemical structure and the types of experimental studies and applications available. These data together with the corresponding bibliographic references stored in the database for the selected amino acid may be exported to a report. The "conformations" folder (Figure 3b) displays information extracted from the theoretical study performed at the highest quantum mechanical level reported in the literature. Specifically, it includes the bibliographic source, the number of minima characterized, the list of dihedral angles and energy of each minimum, the position of each minima within the Ramachandran map, and the quantum mechanical method used in the theoretical study. A report containing this information may be generated. Furthermore, this folder allows a graphical inspection of each minimum energy conformation with the molecular visualization software Rasmol.<sup>18</sup> This is made possible by accessing the atomic coordinates stored in the database through an internal call of the interface. The atomic coordinates of each minimum energy conformation may be also extracted. The next two folders provide bibliographic data about experimental studies (synthesis and characterization, conformational propensities) and reported applications in medicine and materials science. Such bibliographic information is not intended to be comprehensive but rather present some significant publications in the field. Thus, these folders supply essential information reported for the amino acid considered, but a systematic literature search is recommended if this amino acid is finally selected by the user for practical applications. Finally, the last folder allows access to the X-ray crystal structures of peptide sequences containing the selected amino acid that are available in the database. Three-dimensional visualization of these structures is made through the Rasmol software.<sup>18</sup>

### $\alpha$ -Tetrasubstituted $\alpha$ -Amino Acids

Because of our interest in  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids and their relevance in the design of peptides with well-defined conformational properties,<sup>19</sup> this is the first family of nc-aa integrated into NCAD. Table 1 lists the 29  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids whose intrinsic conformational propensities have been determined to date using *ab initio* or DFT quantum mechanical methods<sup>20–37</sup> according to a literature search. This list is expected to be dynamic and thus be updated when studies on new compounds in this category are reported. Other families of nc-aa will be integrated into NCAD in the near future.

$\alpha$ -Tetrasubstituted amino acids are characterized by an  $\alpha$ -carbon atom bearing four substituents (different from hydrogen) and hence are also called quaternary amino acids. They are usually divided into two subtypes, depending on whether the  $\alpha$  carbon is involved in a cyclic structure or not. Notably, the former has been studied much more frequently, the number of entries in the database being 24 and 5, respectively. Regarding chirality, 9 residues in Table 1 present two identical substituents at the  $\alpha$  carbon and are therefore achiral, namely, Aib, Ac<sub>3</sub>c, Dpg, Dbg, Ac<sub>4</sub>c, Ac<sub>5</sub>c, Adt, Ac<sub>6</sub>c, and Toac. The remaining 20 amino acids present at least one chiral center. In the latter case, only the low energy minima of one enantiomer (usually L) have been considered for incorporation into NCAD, those of the other enantiomer being easily deducible by simply changing the sign of both  $(\phi, \psi)$  angles. When two or more chiral centers are present in the molecule, the stereochemistry is indicated through the *R/S* instead of the *L/D* nomenclature (the *S* configuration at the  $\alpha$  carbon usually corresponding to L).

Figure 4 presents Ramachandran maps corresponding to the  $(\phi, \psi)$  backbone dihedral angles of all the energy minima reported for the 29  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids included in the database.<sup>20–37</sup> Those chiral in nature (Figure 4a) are represented together with the minima calculated by *ab initio* quantum mechanical methods for the *N*-acetyl-*N'*-methylamide derivative of L-alanine<sup>38</sup> (MeCO-Ala-NHMe), as a reference for proteinogenic amino acids. The non-chiral  $\alpha$ -tetrasubstituted residues (Figure 4b) are compared with the glycine minima (calculated *ab initio* on MeCO-Gly-NHMe<sup>38</sup>). The latter map is centrosymmetric since two points  $(\phi, \psi)$  and  $(-\phi, -\psi)$  are equivalent for achiral amino acids.

As can be seen, tetrasubstitution at the  $\alpha$  carbon affects the location of the minimum energy conformations typically found for proteinogenic residues. Specifically, the  $\gamma_L$  conformation (intramolecularly hydrogen bonded seven-membered ring, also denoted C<sub>7</sub> conformation), which was found<sup>38</sup> at  $(\phi, \psi) = (-86, 79)$  and  $(-86, 72)$  for Ala and Gly, respectively, tends to evolve toward lower absolute  $\psi$  values upon replacement of the  $\alpha$ -hydrogen atom by a substituent. The distortion of the  $\psi$  dihedral angle shown by quaternary  $\alpha$ -amino acids is maximal for 1-aminocyclopropanecarboxylic acid (Ac<sub>3</sub>c, Table 1), which presents a  $\gamma_L$  minimum at  $(\phi, \psi) = (-77, 34)$ ,<sup>21</sup> as a consequence of the peculiar stereoelectronic properties of the three-membered ring.<sup>39</sup>

The  $\psi$  backbone dihedral angle characterized for the  $\alpha$ -helix conformation also deviates to lower absolute values when the  $\alpha$  carbon is tetrasubstituted. Thus, the  $\alpha_L$  conformation of Ala was located at  $(\phi, \psi) = (-61, -41)$ <sup>38</sup> (as was for Gly<sup>38</sup>), whereas that of its  $\alpha$ -methylated counterpart,  $\alpha$ -aminoisobutyric acid (Aib, Table 1), is found at  $(-65, -31)$ .<sup>20</sup> A similar trend, or even more pronounced, is observed for most amino acids in Table 1, either chiral or not. Thus, for example, Ac<sub>5</sub>c, Ac<sub>6</sub>c, and L-c<sub>3</sub>Dip exhibit  $\alpha_L$  minima with  $(\phi, \psi)$  values  $(-73, -15)$ ,<sup>27</sup>  $(-70, -20)$ ,<sup>31</sup> and  $(-80, -20)$ ,<sup>22</sup> respectively. This apparently minor change brings about important consequences, and indeed, Aib and other quaternary  $\alpha$ -amino acids are known to stabilize the  $3_{10}$ -helix<sup>19</sup> rather than the  $\alpha$ -helix typically found for proteinogenic amino acids. The slight difference in the backbone dihedral angles translates into significant variations in the parameters associated to each helical structure, including the hydrogen bonding pattern, which involves residues *i* and *i*+3 in the  $3_{10}$ -helix and residues *i* and *i*+4 in the  $\alpha$ -helix.<sup>40</sup> Quaternary  $\alpha$ -amino acids can therefore be used to design helical structures exhibiting hydrogen-bonding schemes and geometries different from those formed by standard amino acids.

TABLE 1:  $\alpha$ -Tetrasubstituted  $\alpha$ -Amino Acids Stored in NCAD<sup>a</sup>

Abbreviated name	Systematic name (common name)	Structure	Ref.	Abbreviated name	Systematic name (common name)	Structure	Ref.
Aib	2-amino-2-methylpropanoic acid ( $\alpha$ -aminoisobutyric acid)		20	Ac <sub>6</sub> c	1-aminocyclohexanecarboxylic acid		31
Ac <sub>3</sub> c	1-aminocyclopropanecarboxylic acid		21	(1 <i>S</i> ,2 <i>S</i> )c <sub>6</sub> Phe	(1 <i>S</i> ,2 <i>S</i> )-1-amino-2-phenylcyclohexanecarboxylic acid		32
(1 <i>S</i> ,2 <i>S</i> )c <sub>3</sub> Phe	(1 <i>S</i> ,2 <i>S</i> )-1-amino-2-phenylcyclopropanecarboxylic acid		21	(1 <i>S</i> ,2 <i>R</i> )c <sub>6</sub> Phe	(1 <i>S</i> ,2 <i>R</i> )-1-amino-2-phenylcyclohexanecarboxylic acid		32
(1 <i>S</i> ,2 <i>R</i> )c <sub>3</sub> Phe	(1 <i>S</i> ,2 <i>R</i> )-1-amino-2-phenylcyclopropanecarboxylic acid		21	Toac	4-amino-2,2,6,6-tetramethylpiperidine-1-oxyl-4-carboxylic acid		33
L-c <sub>3</sub> Dip	( <i>S</i> )-1-amino-2,2-diphenylcyclopropanecarboxylic acid		22	—	(1 <i>S</i> ,2 <i>R</i> ,4 <i>R</i> )-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid		34
(2 <i>S</i> ,3 <i>S</i> )c <sub>3</sub> diPhe	(2 <i>S</i> ,3 <i>S</i> )-1-amino-2,3-diphenylcyclopropanecarboxylic acid		23	—	(1 <i>R</i> ,2 <i>R</i> ,4 <i>S</i> )-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid		34
Dpg	2-amino-2,2-diphenylacetic acid (diphenylglycine)		24	—	(1 <i>S</i> ,2 <i>S</i> ,3 <i>R</i> ,4 <i>R</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
Dbg	2-amino-2-benzyl-3-phenylpropanoic acid (dibenzylglycine)		25	—	(1 <i>S</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>R</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
Ac <sub>4</sub> c	1-aminocyclobutanecarboxylic acid		26	—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>R</i> ,4 <i>S</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
Ac <sub>5</sub> c	1-aminocyclopentanecarboxylic acid		27	—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>S</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
Adt	4-amino-1,2-dithiolane-4-carboxylic acid		28	—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>S</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
(1 <i>S</i> ,2 <i>S</i> )c <sub>5</sub> Phe	(1 <i>S</i> ,2 <i>S</i> )-1-amino-2-phenylcyclopentanecarboxylic acid		29	—	(1 <i>R</i> ,2 <i>S</i> ,3 <i>S</i> ,4 <i>S</i> )-2-amino-3-phenylbicyclo[2.2.1]heptane-2-carboxylic acid		34
(1 <i>S</i> ,2 <i>R</i> )c <sub>5</sub> Phe	(1 <i>S</i> ,2 <i>R</i> )-1-amino-2-phenylcyclopentanecarboxylic acid		29	L-( $\alpha$ Me)Pro	(2 <i>S</i> )-2-methylpyrrolidine-2-carboxylic acid [( $\alpha$ -methyl)proline]		35
(1 <i>S</i> ,2 <i>S</i> )c <sub>5</sub> Arg	(1 <i>S</i> ,2 <i>S</i> )-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid		30	L-( $\alpha$ Ph)Pro	(2 <i>R</i> )-2-phenylpyrrolidine-2-carboxylic acid [( $\alpha$ -phenyl)proline]		35
(1 <i>S</i> ,2 <i>R</i> )c <sub>5</sub> Arg	(1 <i>S</i> ,2 <i>R</i> )-1-amino-2-(guanidinomethyl)-cyclopentanecarboxylic acid		30	—	(1 <i>S</i> ,2 <i>R</i> ,3 <i>R</i> ,5 <i>R</i> ,6 <i>R</i> ,7 <i>S</i> ,8 <i>R</i> ,9 <i>R</i> ,10 <i>R</i> )-8-aminopentacyclo[5.4.0.0 <sup>(2,6)</sup> .0 <sup>(3,10)</sup> .0 <sup>(5,9)</sup> ]-undecane-8-carboxylic acid		36
				—	( <i>S</i> )-4-aminopentacyclo[6.3.0.0 <sup>(2,6)</sup> .0 <sup>(3,10)</sup> .0 <sup>(5,9)</sup> ]-undecane-4-carboxylic acid		37

<sup>a</sup> The systematic name (and the common and abbreviated names,<sup>b</sup> when available), the chemical structure, and the reference reporting the quantum mechanical study are given for each amino acid. <sup>b</sup> In the abbreviated name, the configuration is defined by the L/D nomenclature when only the  $\alpha$  carbon is chiral (L usually corresponding to *S*). When two or more chiral centers are present, the *R/S* stereochemical descriptors are used instead.

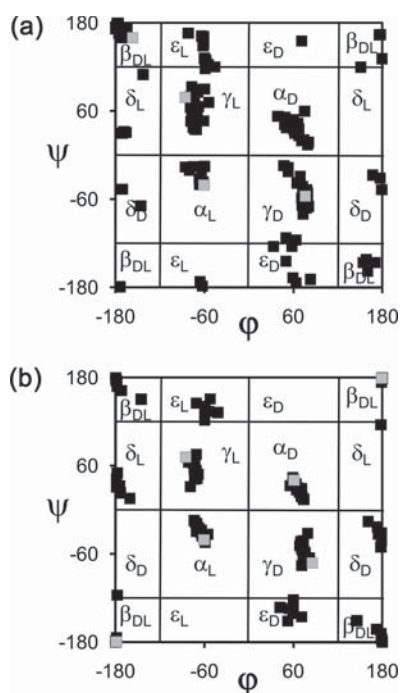
Another effect of  $\alpha$ -tetrasubstitution is the stabilization of minimum energy conformations that are not usually detected experimentally for coded amino acids because they are energetically unfavored. This feature is clearly observed when only the minima found for quaternary amino acids within a relative energy interval of 2.0 kcal/mol, i.e., those energetically accessible, are considered and compared with Ala or Gly (see Figures 5a and 5b for chiral and achiral residues, respectively). Specifically, the  $\epsilon_D$  and  $\epsilon_L$  regions of the  $(\phi, \psi)$  map are visited by some  $\alpha$ -tetrasubstituted amino acids, like Aib<sup>20</sup> and (1*S*,2*R*)c<sub>3</sub>Phe,<sup>21</sup> but not by Ala<sup>38</sup> or Gly.<sup>38</sup> Furthermore, the  $\alpha_L$  conformation is energetically accessible to several of the nc-aa considered, whereas the Ala<sup>38</sup>  $\alpha_L$  minimum is unfavored by nearly 4.0 kcal/mol with respect to the global minimum. The same holds true for Gly.<sup>38</sup> Moreover, minima in the  $\alpha_D$  region are found under the threshold of 2.0 kcal/mol for several amino acids in Table 1, among which are not only those of achiral nature, and therefore presenting an identical propensity to accommodate  $\alpha_L$  and  $\alpha_D$  conformations, but also some exhibiting an L configuration, like (1*S*,2*S*)c<sub>5</sub>Arg.<sup>30</sup> Therefore, tetrasubstitution at the  $\alpha$  carbon may result in the stabilization of

conformations typically accessed by D amino acids even if the L chirality is maintained.

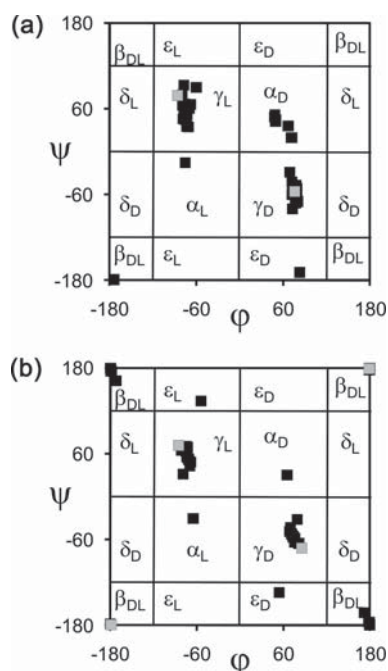
The features commented on above illustrate the great conformational impact that may derive from the presence of  $\alpha$ -tetrasubstituted amino acids in a peptide chain, which has been demonstrated in numerous investigations.<sup>19,40</sup> Nowadays, the usefulness of this family of nc-aa in the control of peptide conformation is beyond doubt, but the practical application to the design of biologically relevant molecules or peptide-based materials is only in its earliest stages. NCAD is intended to mean a significant contribution to its development.

### In Silico Molecular Engineering for Targeted Replacements in Met-Enkephalin Using NCAD: A Test Case

We provide in this section a computational design study aimed at constructing mutants of Met-Enkephalin using information extracted from NCAD. This is not intended to provide a deep and rigorous investigation of the resulting peptides but to illustrate the potential utility of the database in applications involving peptide and protein engineering through a test case.



**Figure 4.** Ramachandran maps showing all the minimum energy conformations predicted by quantum mechanical calculations for the  $\alpha$ -tetrasubstituted  $\alpha$ -amino acids included in NCAD (see Table 1): (a) chiral residues (black squares) compared with Ala (gray squares); (b) achiral residues (black squares) compared with Gly (gray squares).



**Figure 5.** As in Figure 4, but only minimum energy conformations with a relative energy below 2.0 kcal/mol are considered.

The Met-enkephalin neurotransmitter is a pentapeptide (Tyr-Gly-Gly-Phe-Met) that interacts with opioid receptors.<sup>41</sup> Opioid systems play a significant role in pain medication and opiate dependence, regulate dopamine release, and are thought to be important for drug-induced reward. Interestingly, no unique native structure has been found for Met-enkephalin, whose conformational flexibility has been examined using a wide variety of experimental techniques (e.g., X-ray crystallography, NMR, circular dichroism, and infrared, ultraviolet, and fluorescence spectroscopies).<sup>42</sup> Accordingly, this short peptide is

able to adopt different conformations depending on the biological context, which explains its ability to interact with opioid receptors of the  $\mu$ ,  $\kappa$ , and  $\delta$  types.<sup>41</sup> The multiple conformational states accessible to Met-enkephalin are therefore at the basis of its poor selectivity of action. A number of theoretical studies using both conventional and advanced simulation methods based on molecular dynamics (MD) and Monte Carlo (MC) algorithms have been devoted to explore the energy landscape of Met-enkephalin and have shown its ill-defined conformational state.<sup>43–45</sup>

We have examined the information stored in NCAD to propose targeted replacements able to stabilize one of the conformations experimentally detected for Met-enkephalin. Specifically, we considered the conformation identified by NMR for Met-enkephalin in fast-tumbling bicelles.<sup>46</sup> The atomic coordinates of this conformation were extracted from the Protein Data Bank<sup>16</sup> (PDB code: 1PLW).

The peptide was placed in the center of a cubic simulation box ( $a = 44.6$  Å) filled with 2916 explicit water molecules, which were represented using the TIP3 model.<sup>47</sup> The *N*- and *C*-termini were described using the positively charged ammonium and the negatively charged carboxylate groups, respectively. All the MD simulations were performed using the NAMD computer program.<sup>48</sup> To equilibrate the density of the simulation box, different consecutive rounds of short MD runs were performed. Thus, 0.5 ns of NVT-MD at 350 K was used to homogeneously distribute the solvent in the box. Next, 0.5 ns of NVT-MD at 298 K (thermal equilibration) and 0.5 ns of NPT-MD at 298 K (density relaxation) were run. After this, the peptide conformation was equilibrated by running a 3 ns trajectory of NVT-MD at 298 K. The last snapshot of this trajectory was used to determine the  $(\phi, \psi)$  dihedral angles of the three central residues of Met-enkephalin (Gly-Gly-Phe), which define the most flexible region of the peptide. The  $(\phi, \psi)$  values thus obtained were (83,77) for Gly2, (61,−81) for Gly3, and (−66,−38) for Phe4, which correspond to the  $\alpha_D$ ,  $\gamma_D$ , and  $\alpha_L$  conformations, respectively, according to Perczel's nomenclature.<sup>14</sup>

These structural parameters have been used to design targeted substitutions in Met-enkephalin aimed at reducing the flexibility of the peptide backbone so that the particular conformation selected is stabilized. This molecular design pursues the validation of the performance and usefulness of the NCAD database. The following criteria were considered in the design: (i) the side chain functionality of the nc-aa selected should be similar to that of the natural residue to be replaced, which is essential to preserve the biological functions of the peptide and (ii) the intrinsic conformational properties of the selected nc-aa must be consistent with the conformation adopted by the targeted residue in the parent peptide. On the basis of these selection criteria and the structural parameters provided by MD for the Gly-Gly-Phe peptide fragment, three different mutations were proposed using the information in NCAD:

(a) Replacement of Gly2 or Gly3 by  $\alpha$ -aminoisobutyric acid (Aib, see Table 1). This amino acid bears two methyl groups at the  $\alpha$  position (instead of the two  $\alpha$ -hydrogen atoms in Gly) and presents low-energy minima<sup>20</sup> located at the  $\alpha_L$  and  $\gamma_L$  regions of the Ramachandran map, with  $(\phi, \psi)$  angles (−65,−31) and (−76,58) and relative energies 1.7 and 0.0 kcal/mol, respectively, at the MP2/6-31G(d)//HF/6-31G(d) level. Given the non-chiral nature of Aib, these minimum energy conformations are equivalent to the enantiomeric  $\alpha_D$  and  $\gamma_D$ , characterized by  $(\phi, \psi)$  dihedrals (65,31) and (76,−58), respectively. The latter values fit very well with those provided by MD for the Gly2



[(83,77)] and Gly3 [(61,−81)] residues of Met-enkephalin. The mutants resulting from these substitutions are denoted G2 and G3.

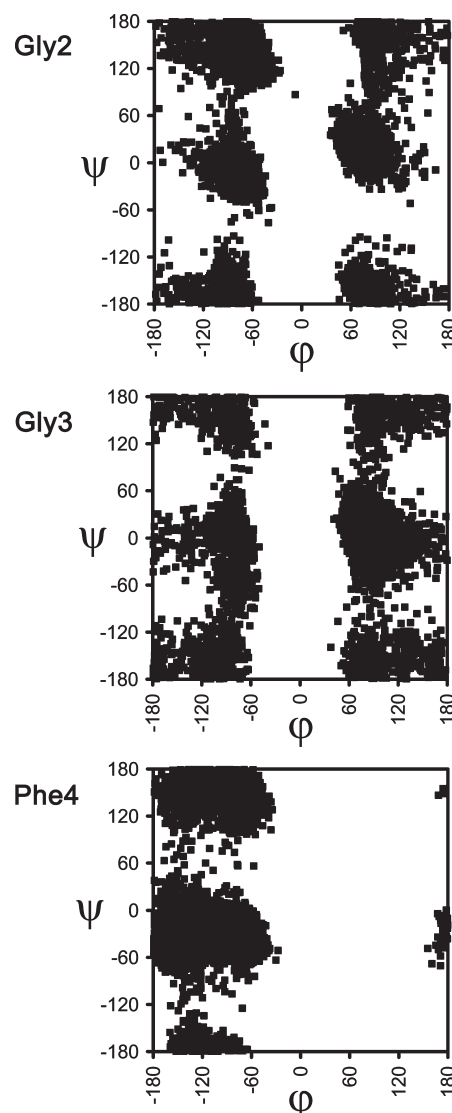
(b) Replacement of Phe4 by (2*R*,3*R*)-1-amino-2,3-diphenylcyclopropanecarboxylic acid, known in the abbreviated form as (2*R*,3*R*)c<sub>3</sub>diPhe. This cyclic  $\alpha$ -tetrasubstituted residue contains a cyclopropane ring bearing two vicinal phenyl substituents in a *trans* relative disposition [see the (2*S*,3*S*) enantiomer in Table 1]. It retains the side chain of Phe and shows an energy minimum of the  $\alpha_L$  type characterized by ( $\phi$ , $\psi$ ) values (−67,−27),<sup>23</sup> which closely resemble those obtained for Phe4 in Met-enkephalin in the MD trajectory described above, (−66,−38). Moreover, (2*R*,3*R*)c<sub>3</sub>diPhe has been shown experimentally to exhibit a high propensity to adopt  $\alpha_L$  conformations due to the interaction between the two rigidly held aromatic groups and the peptide backbone.<sup>49</sup> The mutant resulting from the replacement of Phe4 by (2*R*,3*R*)c<sub>3</sub>diPhe in Met-enkephalin is denoted F4.

MD trajectories of 6 ns were performed for the wild-type peptide and the three mutants designed. The starting geometries for the mutants were constructed using the coordinates stored in the database for the minimum energy conformation selected for each nc-aa. The protocols applied for thermal equilibration and density relaxation were identical to those described above. All the force-field parameters for the MD simulations were extracted from the AMBER libraries,<sup>50</sup> with the exception of the electrostatic charges for the nc-aa, which were taken from the papers quoted in the database.

Figure 6 shows the accumulated Ramachandran plots obtained through 6 ns of simulation corresponding to the Gly2, Gly3, and Phe4 residues for the wild-type peptide. As can be seen, the two Gly residues sample all the accessible regions of the ( $\phi$ , $\psi$ ) space, which is consistent with the large conformational flexibility observed for Met-enkephalin in aqueous solution by both experimental<sup>42a,b,d–f</sup> and theoretical techniques.<sup>43–45</sup> The Phe4 residue also exhibits a great flexibility, although it only explores conformations in the left-half part of the map, as expected from its L configuration (in comparison, for the non-chiral Gly residue, the left and right parts of the map are energetically equivalent). Therefore, all three residues are actually found to adopt ( $\phi$ , $\psi$ ) values in all the regions that are energetically accessible to them. In other words, no selectivity is detected toward a particular structure among those allowed to L (in the case of Phe) or achiral (in the case of Gly) coded residues. It is worth noting that the sampling of the conformational space shown in Figure 6 is in excellent agreement with that obtained for the same peptide using replica exchange MD.<sup>45</sup>

Figure 7 shows the ( $\phi$ , $\psi$ ) distributions obtained for the nc-aa included in mutants G2, G3, and F4 of Met-enkephalin. Table 2 displays the population of the different backbone conformations, which is expressed in terms of number of visits to a certain region of the Ramachandran map during the MD trajectories. These data provide evidence for the great conformational selectivity induced by the  $\alpha$ -tetrasubstituted amino acids used. In particular, for the G2 mutant, the Aib residue (Gly2 substitute) falls exclusively in the  $\alpha_D$  region. Accordingly, the substitution designed using NCAD successfully restricts the conformational flexibility of the Gly2 residue and, moreover, preserves the selected arrangement. Thus, the replacement of Gly2 by Aib results in the stabilization of the conformation adopted by the coded residue in the starting geometry.

Replacement of Gly3 by Aib also has a conformational confinement effect although less intense. Indeed, only the  $\alpha_D$ ,



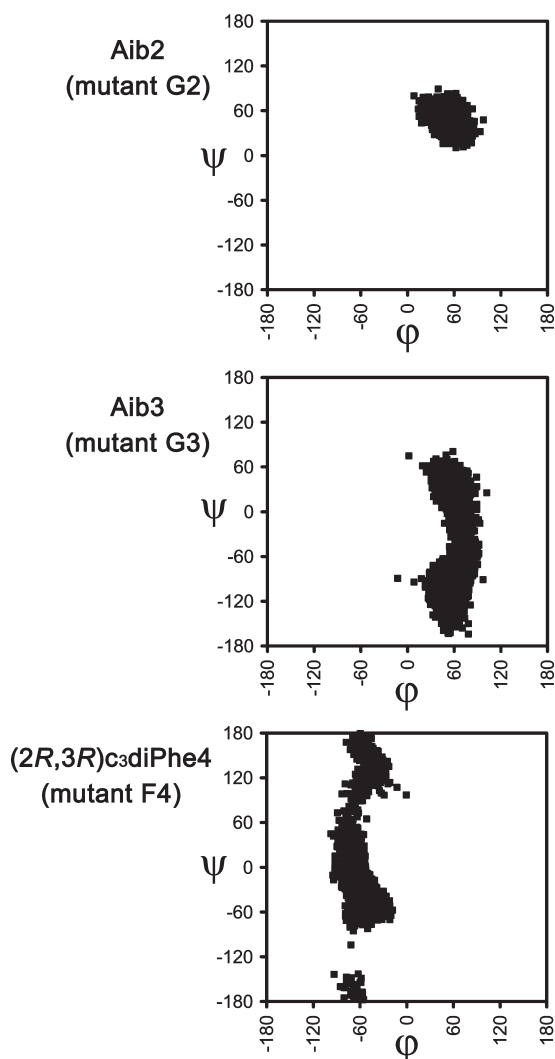
**Figure 6.** Backbone ( $\phi$ , $\psi$ ) dihedral angle distributions obtained for the Gly2, Gly3, and Phe4 residues in wild-type Met-enkephalin.

$\gamma_D$ , and  $\varepsilon_D$  structures are adopted by the Aib residue in the G3 mutant (Figure 7), their respective populations being 29.5%, 66.1%, and 4.4% (Table 2). In contrast, Gly3 in the unmodified peptide was found to visit all nine regions used by Perczel et al.<sup>14</sup> to describe the conformational space. We also note that the two regions most populated by Gly3 in the wild-type system, namely,  $\alpha_D$  and  $\gamma_D$ , are kept by Aib in the G3 mutant. Thus, although the Gly-Aib exchange proved less restrictive in G3 than in G2, the results obtained for the G3 mutant illustrate again that properly selected nc-aa can be successfully used to reduce the conformational freedom of biological (macro)molecules through the stabilization of selected secondary structure motifs.

To further evidence the magnitude of the conformational restrictions introduced by the nc-aa in this study, Table 3 displays the ratio between the populations of the conformations found for the quaternary  $\alpha$ -amino acids and those of the corresponding coded residues in the wild-type peptide. For instance, mutant G3 increases the population obtained for the  $\gamma_D$  conformation of Gly3 in Met-enkephalin by a factor of 2.2.

In the F4 mutant, the restricted geometry of (2*R*,3*R*)c<sub>3</sub>diPhe affects mainly the  $\phi$  angle, which remains fixed in the neighborhood of −60° during the whole MD trajectory (Figure 7). In





**Figure 7.** Backbone ( $\phi, \psi$ ) dihedral angle distributions obtained for the Aib2, Aib3, and (2R,3R) $c_3$ diPhe4 residues in the G2, G3, and F4 mutants of Met-enkephalin.

**TABLE 2: Relative Shares of Populations (%) Found for Selected Residues in Wild-Type Met-Enkephalin (Gly2, Gly3, and Phe4) and for the  $\alpha$ -Tetrasubstituted  $\alpha$ -Amino Acids Included in the G2, G3, and F4 Mutants<sup>a</sup>**

	$\alpha_D$	$\epsilon_D$	$\gamma_D$	$\delta_D$	$\beta_{DL}$	$\delta_L$	$\gamma_L$	$\epsilon_L$	$\alpha_L$
Gly2	25.1	8.9	3.1	0.5	5.0	1.0	10.1	34.0	12.3
Gly3	29.2	7.3	30.3	3.6	7.0	1.9	5.7	7.3	7.7
Phe4	0.0	0.0	0.0	19.1	18.4	3.5	2.9	21.7	34.4
Aib2 (G2)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aib3 (G3)	29.5	4.4	66.1	0.0	0.0	0.0	0.0	0.0	0.0
(2R,3R) $c_3$ diPhe4 (F4)	0.0	0.0	0.0	0.0	0.0	0.0	13.1	4.7	82.2

<sup>a</sup> The nine types of backbone conformation defined in ref 14 have been considered.

comparison, the Phe4 residue in Met-enkephalin was found to cover all the possible negative values of  $\phi$ :  $-180^\circ < \phi < -50^\circ$  (Figure 6). As a result, the  $\alpha_L$  conformation, which corresponds to that adopted by Phe4 in the original arrangement selected, significantly increases its population from the wild-type peptide (34.4%) to the F4 mutant (82.2%) and becomes the most visited structure during the simulation of the latter peptide. The selectivity ratio for the F4 mutant is 4.5 and 2.4 for the  $\gamma_L$  and  $\alpha_L$  regions, respectively (Table 3). Accordingly, the nc-aa used

**TABLE 3: Ratio between the Populations of Conformations<sup>a</sup> Found for the  $\alpha$ -Tetrasubstituted  $\alpha$ -Amino Acids in the G2, G3, and F4 Mutants and the Corresponding Proteinogenic Residues in Wild-Type Met-Enkephalin**

	$\alpha_D$	$\epsilon_D$	$\gamma_D$	$\delta_D$	$\beta_{DL}$	$\delta_L$	$\gamma_L$	$\epsilon_L$	$\alpha_L$
Aib2 (G2)	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aib3 (G3)	1.0	0.6	2.2	0.0	0.0	0.0	0.0	0.0	0.0
(2R,3R) $c_3$ diPhe4 (F4)	— <sup>b</sup>	— <sup>b</sup>	— <sup>b</sup>	0.0	0.0	0.0	4.5	0.2	2.4

<sup>a</sup> The nine types of backbone conformation defined in ref 14 have been considered. <sup>b</sup> This conformation was not adopted by the  $\alpha$ -tetrasubstituted residue.

as a Phe4 substitute also stabilizes the desired conformation and further proves the utility and potential applicability of the new database.

## Conclusions

NACD (Non-Coded Amino acids Database) is an easy-to-use research tool that integrates the intrinsic conformational preferences of non-coded amino acids to facilitate their use in the design of peptide- or protein-based compounds useful in different fields of life and materials sciences. Our aim is to provide the outcome of structural studies of nc-aa, including their properties and chemical characteristics, obtained by scientists in the organic and physical chemistry disciplines to those focusing on biological, bioengineering, and medicinal applications, bringing the fields together. The information provided in NCAD, which is not available in other databases, is also expected to be a useful resource for protein designers, modelers, and experts in bioinformatics. The present paper presents the structure and informatics architecture of the database, the interface that connects the database with the users, the description of the first family of nc-aa integrated into the database ( $\alpha$ -tetrasubstituted  $\alpha$ -amino acids), and a test case showing the applicability of the database. Researchers can now easily select the most appropriate residue from a collection of nc-aa with well-established conformational properties and introduce it in a peptide sequence or any other (macro)molecule. Additional families of non-proteinogenic amino acids will be included in NCAD in due course.

**Acknowledgment.** 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 and CTQ2008-00423-E/BQU; Ramón y Cajal contract for D.Z.), Gobierno de Aragón (research group E40), and Generalitat de Catalunya (research group 2009 SGR 925; XRQTC; ICREA Academia prize for excellence in research to C.A.) 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 and Notes

- (1) (a) *Amino Acids, Peptides and Proteins in Organic Chemistry: Origins and Synthesis of Amino Acids*; Hughes, A. B., Ed.; Wiley-VCH:

- Weinheim, 2009; Vol. 1. (b) Barrett, G. C.; Elmore, D. T. *Amino Acids and Peptides*; Cambridge University Press: Cambridge, 1998. (c) Mauger, A. B. *J. Nat. Prod.* **1996**, *59*, 1205. (d) Wagner, I.; Musso, H. *Angew. Chem., Int. Ed. Engl.* **1983**, *22*, 816.
- (2) (a) *Asymmetric Synthesis and Application of  $\alpha$ -Amino Acids*; Soloshonok, V. A.; Izawa, K., Eds.; American Chemical Society: Washington DC, 2009. (b) Cativiela, C.; Ordóñez, M. *Tetrahedron: Asymmetry* **2009**, *20*, 1. (c) Michaux, J.; Niel, G.; Campagne, J.-M. *Chem. Soc. Rev.* **2009**, *38*, 2093. (d) Ordóñez, M.; Rojas-Cabrera, H.; Cativiela, C. *Tetrahedron* **2009**, *65*, 17. (e) Cativiela, C.; Díaz-de-Villegas, M. D. *Tetrahedron: Asymmetry* **2007**, *18*, 569. (f) Ordóñez, M.; Cativiela, C. *Tetrahedron: Asymmetry* **2007**, *18*, 3. (g) Bonauer, C.; Walenzyk, T.; König, B. *Synthesis* **2006**, 1. (h) Aurelio, L.; Brownlee, R. T. C.; Hughes, A. B. *Chem. Rev.* **2004**, *104*, 5823. (i) Lelais, G.; Seebach, D. *Biopolymers (Pept. Sci.)* **2004**, *76*, 206.
- (3) (a) *Peptaibiotics: Fungal Peptides Containing  $\alpha$ -Dialkyl  $\alpha$ -Amino Acids*; Toniolo, C.; Brückner, H., Eds.; Wiley-VCH: Weinheim, 2009. (b) *Handbook of Biologically Active Peptides*; Kastin, A. J., Ed.; Academic Press: London, 2006. (c) Degenkolb, T.; Bruckner, H. *Chem. Biodiversity* **2008**, *5*, 1817. (d) Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, *20*, 275.
- (4) (a) *Peptide and Protein Design for Biopharmaceutical Applications*; Jensen, K. J., Ed.; John Wiley & Sons: Chichester, 2009. (b) Nestor, J. J. *Curr. Med. Chem.* **2009**, *16*, 4399. (c) Horne, W. S.; Gellman, S. H. *Acc. Chem. Res.* **2008**, *41*, 1399. (d) Seebach, D.; Gardiner, J. *Acc. Chem. Res.* **2008**, *41*, 1366. (e) Chatterjee, J.; Gilon, C.; Hoffman, A.; Kessler, H. *Acc. Chem. Res.* **2008**, *41*, 1331. (f) Cowell, S. M.; Lee, Y. S.; Cain, J. P.; Hruby, V. J. *Curr. Med. Chem.* **2004**, *11*, 2785. (g) Sagan, S.; Karoyan, P.; Lequin, O.; Chassaing, G.; Lavielle, S. *Curr. Med. Chem.* **2004**, *11*, 2799. (h) Adessi, C.; Soto, C. *Curr. Med. Chem.* **2002**, *9*, 963.
- (5) (a) *Pharmaceutical Substances: Syntheses, Patents, Applications*; Kleemann, A.; Engel, J.; Kutscher, B.; Reichert, D., Eds.; Thieme: Stuttgart, 2009. (b) Ersmark, K.; Del Valle, J. R.; Hanessian, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 1202. (c) Menard, J.; Patchett, A. A. *Adv. Protein Chem.* **2001**, *56*, 13.
- (6) (a) Voloshchuk, N.; Montclare, J. K. *Mol. Biosyst.* **2010**, *6*, 65. (b) Wu, X.; Schultz, P. G. *J. Am. Chem. Soc.* **2009**, *131*, 12497. (c) Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4356. (d) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2005**, *44*, 34. (e) Hendrickson, T. L.; De Crécy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* **2004**, *73*, 147. (f) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.
- (7) (a) Lee, H. S.; Guo, J.; Lemke, E.; Dimla, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2009**, *131*, 12921. (b) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9785. (c) Vázquez, M. E.; Nitz, M.; Stehn, J.; Yaffe, M. B.; Imperiali, B. *J. Am. Chem. Soc.* **2003**, *125*, 10150. (d) Murakami, H.; Hoshida, T.; Ashizuka, Y.; Hashimoto, K.; Sisido, M. *Macromolecules* **2000**, *1*, 118.
- (8) (a) Alfonta, L.; Zhang, Z.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 14662. (b) Shinohara, H.; Kusaka, T.; Yokota, E.; Monden, R.; Sisido, M. *Sens. Actuators B* **2000**, *65*, 144.
- (9) (a) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. *Nat. Chem. Biol.* **2007**, *3*, 769. (b) Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 388. (c) Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokohama, S. *Nat. Methods* **2005**, *2*, 201. (d) Rothman, D. M.; Vázquez, M. E.; Vogel, E. M.; Imperiali, B. *Org. Lett.* **2002**, *4*, 2865.
- (10) (a) Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L. *Nat. Neurosci.* **2007**, *10*, 1063. (b) Kick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19.
- (11) (a) Cellitti, S. E.; Jones, D. H.; Lagpacan, L.; Hao, X.; Zhang, Q.; Hu, H.; Brittain, S. M.; Brinker, A.; Caldwell, J.; Bursulaya, B.; Spraggon, G.; Brock, A.; Ryu, Y.; Uno, T.; Schultz, P. G.; Geierstanger, B. H. *J. Am. Chem. Soc.* **2008**, *130*, 9268. (b) Reid, P. J.; Loftus, C.; Beeson, C. C. *Biochemistry* **2003**, *42*, 2441.
- (12) (a) Zanuy, D.; Ballano, G.; Jiménez, A. I.; Casanovas, J.; Haspel, N.; Cativiela, C.; Curcú, D.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* **2009**, *49*, 1623. (b) Cejas, M. A.; Kinney, W. A.; Chen, C.; Vinter, J. G.; Almond, H. R., Jr.; Balss, K. M.; Maryanoff, C. A.; Schmidt, U.; Breslav, M.; Mahan, A.; Lacy, E.; Maryanoff, B. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8513. (c) Zanuy, D.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Phys. Chem. B* **2007**, *111*, 3236. (d) Yang, Z. M.; Liang, G. L.; Ma, M. L.; Gao, Y.; Xu, B. *Small* **2007**, *3*, 558. (e) Crisma, M.; Toniolo, C.; Royo, S.; Jiménez, A. I.; Cativiela, C. *Org. Lett.* **2006**, *8*, 6091. (f) Behenna, H. A.; Donners, J. J. J. M.; Gordon, A. C.; Stupp, S. I. *J. Am. Chem. Soc.* **2005**, *127*, 1193. (g) Brea, R. J.; Amorín, M.; Castedo, L.; Granja, J. R. *Angew. Chem., Int. Ed.* **2005**, *44*, 5710. (h) Hartgerink, J. D. *Curr. Opin. Chem. Biol.* **2004**, *8*, 604. (i) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. *Science* **2001**, *294*, 1684. (j) Clark, T. D.; Buehler, L. K.; Ghadiri, M. R. *J. Am. Chem. Soc.* **1998**, *120*, 651.
- (13) (a) Lee, M.-R.; Stahl, S. S.; Gellman, S. H.; Masters, K. S. *J. Am. Chem. Soc.* **2009**, *131*, 16779. (b) Connor, R. E.; Tirrell, D. A. *Polym. Rev.* **2007**, *47*, 9. (c) Börner, H. G.; Schlaad, H. *Soft Matter* **2007**, *3*, 394. (d) Tu, R. S.; Tirrell, M. *Adv. Drug Delivery Rev.* **2004**, *56*, 1537. (e) Kirshenbaum, K.; Zuckermann, R. N.; Dill, K. A. *Curr. Opin. Struct. Biol.* **1999**, *9*, 530. (f) Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuckermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4303. (g) Wang, X.; Huq, I.; Rana, T. M. *J. Am. Chem. Soc.* **1997**, *119*, 6444. (h) Yoshikawa, E.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1994**, *27*, 5471. (i) Cho, C. Y.; Moran, E. J.; Cherry, S. R.; Stephans, J. C.; Fodor, S. P. A.; Adams, C. L.; Sundaram, A.; Jacobs, J. W.; Schultz, P. G. *Science* **1993**, *261*, 1303.
- (14) (a) Perczel, A.; Angyán, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256. (b) Hudáky, I.; Kiss, R.; Perczel, A. *J. Mol. Struct.* **2004**, *675*, 177.
- (15) Allen, F. H. *Acta Crystallogr. B* **2002**, *58*, 380.
- (16) (a) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235. (b) Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. *Acta Crystallogr. D* **1998**, *54*, 1078.
- (17) Chen, P.-S. *ACM Trans. Database Syst.* **1976**, *1*, 9.
- (18) Sayle, R. A.; Milner-White, E. J. *Trends Biochem. Sci.* **1995**, *20*, 374.
- (19) (a) Toniolo, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q. B. *Synlett* **2006**, 1295. (b) Venkatraman, J.; Shankaramma, S. C.; Balaran, P. *Chem. Rev.* **2001**, *101*, 3131. (c) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C. *Biopolymers (Pept. Sci.)* **2001**, *60*, 396.
- (20) Alemán, C. *J. Phys. Chem. B* **1997**, *101*, 5046.
- (21) Alemán, C.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Casanovas, J. *J. Phys. Chem. B* **2002**, *106*, 11849.
- (22) Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Phys. Chem. B* **2006**, *110*, 5762.
- (23) Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Pérez, J. J.; Alemán, C. *J. Org. Chem.* **2003**, *68*, 7088.
- (24) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *J. Org. Chem.* **2007**, *72*, 2174.
- (25) Casanovas, J.; Nussinov, R.; Alemán, C. *J. Org. Chem.* **2008**, *73*, 4205.
- (26) Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. *Chem. Phys. Lett.* **2006**, *429*, 558.
- (27) Alemán, C.; Zanuy, D.; Casanovas, J.; Cativiela, C.; Nussinov, R. *J. Phys. Chem. B* **2006**, *110*, 21264.
- (28) Aschi, M.; Lucente, G.; Mazza, F.; Mollica, A.; Morera, E.; Nalli, M.; Paradisi, M. P. *Org. Biomol. Chem.* **2003**, *1*, 1980.
- (29) Casanovas, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* **2008**, *73*, 644.
- (30) Revilla-López, G.; Torras, J.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C. *J. Org. Chem.* **2009**, *74*, 2403.
- (31) Rodríguez-Ropero, F.; Zanuy, D.; Casanovas, J.; Nussinov, R.; Alemán, C. *J. Chem. Inf. Model.* **2008**, *48*, 333.
- (32) Alemán, C.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Casanovas, J. *J. Org. Chem.* **2009**, *74*, 7834.
- (33) D'Amore, M.; Improtà, R.; Barone, V. *J. Phys. Chem. A* **2003**, *107*, 6264.
- (34) Cordoní, A.; Gomez-Catalan, J.; Jimenez, A. I.; Cativiela, C.; Perez, J. J. *J. Pept. Sci.* **2002**, *8*, 253.
- (35) Flores-Ortega, A.; Jiménez, A. I.; Cativiela, C.; Nussinov, R.; Alemán, C.; Casanovas, J. *J. Org. Chem.* **2008**, *73*, 3418.
- (36) Bisetty, K.; Gomez-Catalan, J.; Alemán, C.; Giralt, E.; Kruger, H. G.; Perez, J. J. *J. Pept. Sci.* **2004**, *10*, 274.
- (37) Bisetty, K.; Govender, P.; Kruger, H. G. *Biopolymers* **2006**, *81*, 339.
- (38) Cornell, W. D.; Gould, I. R.; Kollman, P. A. *J. Mol. Struct. (Theochem)* **1997**, *392*, 101.
- (39) de Meijere, A. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 809.
- (40) Toniolo, C.; Crisma, M.; Formaggio, F.; Peggion, C.; Broxterman, Q.; Kaptein, B. *J. Inclusion Phenom. Macrocyclic Chem.* **2005**, *51*, 121.
- (41) (a) Bodnar, R. J. *Peptides* **2009**, *30*, 2432. (b) McLaughlin, P. J. *Proenkephalin-Derived Opioid Peptides*. In *Handbook of Biologically Active Peptides*; Kastin, A. J., Ed.; Academic Press: London, 2006; pp 1313–1318. (c) Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Morgan, B. A.; Morris, H. R. *Nature* **1975**, *258*, 577.
- (42) (a) Cai, X.; Dass, C. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1. (b) Graham, W. H.; Carter II, E. S.; Hicks, R. P. *Biopolymers* **1992**, *32*, 1755. (c) Griffin, J. F.; Lings, D. A.; Smith, G. D.; Blundell, T. L.; Tickle, I. J.; Bedarkar, S. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 3272. (d) Higashijima, T.; Kobayashi, J.; Nagai, U.; Miyazawa, T. *Eur. J. Biochem.* **1979**, *97*, 43. (e) Spirtes, M. A.; Schwartz, R. W.; Mattice, W. L.; Coy, D. H. *Biochem. Biophys. Res. Commun.* **1978**, *81*, 602. (f) Khaled, M. A.; Long, M. M.; Thompson, W. D.; Bradley, R. J.; Brown, G. B.; Urry, D. W. *Biochem. Biophys. Res. Commun.* **1977**, *76*, 224.
- (43) (a) Koca, J.; Carlsen, H. J. *J. Mol. Struct.* **1995**, *337*, 17. (b) Perez, J. J.; Loew, G. H.; Villar, H. O. *Int. J. Quantum Chem.* **1992**, *44*, 263.

- (44) (a) Hansmann, U. H. E.; Okamoto, Y.; Onuchic, J. N. *Proteins* **1999**, *34*, 472. (b) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.
- (45) Sanbonmatsu, K. Y.; García, A. E. *Proteins* **2002**, *46*, 225.
- (46) Marcotte, I.; Separovic, F.; Auger, M.; Gagné, S. M. *Biophys. J.* **2004**, *86*, 1587.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (48) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

- (49) (a) Crisma, M.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Royo, S.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *Chem.—Eur. J.* **2006**, *12*, 251. (b) Royo, S.; De Borggraeve, W. M.; Peggion, C.; Formaggio, F.; Crisma, M.; Jiménez, A. I.; Cativiela, C.; Toniolo, C. *J. Am. Chem. Soc.* **2005**, *127*, 2036.
- (50) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

JP102092M