

# SwissSidechain: a molecular and structural database of non-natural sidechains

David Gfeller<sup>1</sup>, Olivier Michielin<sup>1,2,3,\*</sup> and Vincent Zoete<sup>1,\*</sup>

<sup>1</sup>Swiss Institute of Bioinformatics (SIB), Quartier Sorge, Bâtiment Génopode, CH-1015 Lausanne, <sup>2</sup>Ludwig Institute for Cancer Research, CH-1015 Lausanne and <sup>3</sup>Pluridisciplinary Center for Clinical Oncology (CePO), Centre Hospitalier Universitaire Vaudois, CH-1015 Lausanne, Switzerland

Received August 15, 2012; Revised September 20, 2012; Accepted September 29, 2012

## ABSTRACT

**Amino acids form the building blocks of all proteins. Naturally occurring amino acids are restricted to a few tens of sidechains, even when considering post-translational modifications and rare amino acids such as selenocysteine and pyrrolysine. However, the potential chemical diversity of amino acid sidechains is nearly infinite. Exploiting this diversity by using non-natural sidechains to expand the building blocks of proteins and peptides has recently found widespread applications in biochemistry, protein engineering and drug design. Despite these applications, there is currently no unified online bioinformatics resource for non-natural sidechains. With the SwissSidechain database (<http://www.swissidechain.ch>), we offer a central and curated platform about non-natural sidechains for researchers in biochemistry, medicinal chemistry, protein engineering and molecular modeling. SwissSidechain provides biophysical, structural and molecular data for hundreds of commercially available non-natural amino acid sidechains, both in L- and D-configurations. The database can be easily browsed by sidechain names, families or physico-chemical properties. We also provide plugins to seamlessly insert non-natural sidechains into peptides and proteins using molecular visualization software, as well as topologies and parameters compatible with molecular mechanics software.**

## INTRODUCTION

Amino acid sidechains confer their specific properties to all proteins. They govern the specific folding of polypeptide chains into distinct 3D structures, build

complementary binding interfaces between members of macromolecular complexes or create enzyme catalytic sites by spatially arranging together different chemical groups (1). Despite the large diversity of existing proteins, it is clear that nature has not exhaustively sampled the set of macromolecules with specific functions that can be chemically synthesized. This is especially true because proteins are made out of a limited number (twenty in most cases) of genetically encoded amino acids. Although this limited number may confer advantages in terms of genetic encoding and synthesis of proteins *in vivo*, it clearly restricts protein's functional diversity.

Introducing new chemistry into existing proteins is therefore an attractive strategy for biochemical studies, protein engineering and human therapeutics. A promising approach is to expand the repertoire of amino acids by introducing non-natural amino acids into existing proteins or peptides (2). Thereby, new regions of the chemical space can be explored that have not been already sampled along evolution. Non-natural amino acids, and especially L-alpha amino acids where the sidechain has only one single bond with the backbone (i.e. excluding amino acids with branching at C $\alpha$  or bonds with other backbone atoms), are particularly interesting because they do not affect the chemical properties of the protein backbone. As such they are more likely to keep the global fold and secondary structure elements of a polypeptide chain unchanged. For instance, in a recent study, it was shown that non-natural sidechains dramatically increase the affinity of amyloid fiber inhibitors (3). Similarly, several cyclic and other kinds of modified peptides with non-natural amino acid sidechains have been developed for therapeutics use, such as cilengtide (4) or crafilzomib (5). Non-natural sidechains have also been used as independent ligands. For instance, L-3,4-dihydroxyphenylalanine is a non-natural amino acid used in Parkinson disease treatment (6), and 5-hydroxytryptophan (oxitriptan) has been used as an antidepressant (7). Another commonly used type of non-natural sidechains consists of D-amino acids.

\*To whom correspondence should be addressed. Tel: +41 21 6924422; Fax: +41 21 6924065; Email: [vincent.zoete@unil.ch](mailto:vincent.zoete@unil.ch)  
Correspondence may also be addressed to Olivier Michielin. Tel: +41 21 6924053; Fax: +41 21 6924065; Email: [olivier.michielin@unil.ch](mailto:olivier.michielin@unil.ch)

D-amino acids are enantiomers of L-amino acids where the sidechain and the hydrogen atom attached to the C $\alpha$  atom have been switched. D-amino acids are known to increase resistance to protease degradation and are thus frequently used to enhance the biological stability of peptide-based drugs or biological tools (8,9).

In addition to therapeutics use, non-natural sidechains have found many other applications in biochemistry and protein studies (10). These include photo-crosslinking amino acids to probe *in vivo* protein interactions (11,12), fluorescent amino acids used as markers of specific proteins (13) or phosphorylated amino acid mimetics to probe the effect of post-translational modifications (14).

The importance of expanding the building blocks of proteins beyond the 20 natural ones is further emphasized by noting that a remarkably high degree of optimization is often observed in the choice of amino acid sequences for many proteins. Optimized sequences stand as an important hurdle to protein or peptide engineering based only on genetically encoded amino acids.

Experimental research on non-natural sidechains is based mainly on solid-phase synthesis technology (15) and genetically encoding non-natural sidechains into micro-organisms or cell cultures (2). The former technique uses *in vitro* chemical peptide synthesis to assemble polypeptide chains with both natural and non-natural residues. This technology is routinely used to synthesize non-natural peptides and can nowadays be applied to small proteins of up to hundred amino acids (16). The latter is typically based on generating novel unique codon-tRNA pairs, together with the corresponding aminoacyl tRNA synthetases in a given organism (2). This approach enables production of non-natural sidechain containing polypeptide chains in micro-organisms that can be used as *in vivo* biological tools (13) or as modified peptide libraries for ligand screening (17). Recent progresses currently enable encoding simultaneously hundreds of different non-natural sidechains in the same organism (18).

Bio- and chemo-informatics, structural modeling and visualization tools are powerful to guide and interpret experimental studies with both natural and non-natural amino acids. Recently, we and others developed different computational tools to better characterize non-natural sidechains and predict the effect of introducing them into proteins or peptides (19–22). This includes sidechain 3D coordinates, biomolecular parameters describing properties such as bond length flexibility, as well characterization of conformational diversity in terms of rotamers (23). For the latter, different strategies have been elaborated to predict rotamer probabilities, using either solely free energy-based calculations (20,22), or combining physics-based calculations together with statistical analysis of conserved properties in natural sidechain rotamer libraries (19).

Here, we introduce the SwissSidechain database, a unified and integrated resource providing access to curated biochemical and structural information for 210 non-natural sidechains. Many of the data are unique to SwissSidechain (e.g. rotamers, biomolecular parameters), whereas the rest has been collected from existing resources to provide a rapid access to this information. The

SwissSidechain database further includes visualization and molecular modeling tools. Different options to browse the non-natural sidechains based on their chemical properties or their families are also provided. The database can be accessed at <http://www.swissidechain.ch>.

### Content of the SwissSidechain database

The SwissSidechain database contains molecular and structural data for 210 non-natural alpha amino acid sidechains, both in L- and D-configurations, in addition to the 20 natural ones. These amino acids have been selected based on two main criteria: first, the presence of non-natural sidechains in publicly available protein structures in the PDB (Protein Data Bank) (24) and second, the commercial availability. Each amino acid in SwissSidechain has been given a three- or four-letter code. For all sidechains present in the PDB, we kept the existing three-letter code. For other sidechains, a four-letter code was created. These include many D-amino acids whose L-configuration is present in the PDB. For them, a 'D' was simply added in front of the three-letter code (e.g. NLE stands for L-Norleucine and DNLE for D-Norleucine). For the rest, four-letter codes were chosen so as to recall as much as possible the full chemical name (e.g. AZDA for L-azido-alanine), and always starting with a 'D' for non-natural sidechains in D-configuration (e.g. DZDA for D-azido-alanine).

SwissSidechain data are available in different files describing the chemical structure (SMILES and 2D chemical drawing) and 3D coordinates (PDB, Mol2) of these sidechains, files describing their physico-chemical properties (molecular weight, volume, logP, pKa, partial charges, bond/angle/torsion constants), as well as files describing the probability of their possible conformations (rotamers). Among the different experimentally measurable chemical properties of non-natural sidechains in SwissSidechain, volumes have been calculated with CHARMM (Chemistry at HARvard Molecular Mechanics) (25) and molecular weights with OpenBabel (26). LogP values give information about the hydrophobicity. Experimental LogP for both the full amino acid and the sidechain only have been manually collected from literature when available. For the rest of the sidechains, LogP values have been predicted with XlogP3 (27). PKa values for each protonatable group of the full amino acids have been retrieved from literature when available and computed with MarvinSketch (ChemAxon, <http://www.chemaxon.com>) otherwise. Model parameters, such as bond, angle and torsion constants, which provide information about sidechain flexibility and dynamics, are provided as parameter files compatible with the CHARMM force field (28) and can be readily used in the CHARMM (25) and GROMACS (GROningen MACHine for Chemical Simulations) (29) molecular mechanics software. Similarly, partial atomic charges, which are useful to predict possible interactions (e.g. H-bond, electrostatic) of a sidechain with its surrounding atoms, are provided as topology files in the CHARMM (25) and GROMACS (29) format.

The computation of these model parameters is described in detail elsewhere (19).

Finally, rotamers have been generated as described in Gfeller *et al.* (19). To include information about sidechain conformations in recent crystal structures, we developed the current version of SwissSidechain based on an up-to-date rotamer library for natural sidechains, instead of using rotamer libraries published in 2002 (23). In practice, we performed a global analysis of natural sidechain conformations in all crystallographic structures available in the PDB (as on May 2012) with resolution lower or equal to 1.75Å. This updated rotamer library of natural sidechains, and especially the conserved probabilities of the first dihedral angles of long sidechains, was used in all our analysis to predict rotamers for non-natural sidechains, as described in Gfeller *et al.* (19). We further expand rotamer libraries to D-amino acids, both natural and non-natural. For non-chiral sidechains, rotamer probabilities can be readily generated using the properties of mirror images. In particular, a D-configuration characterized by backbone dihedral angles  $\phi$  and  $\psi$  and sidechain dihedral angles  $\chi_i$  is equivalent to the L-configuration with dihedral angles  $-\phi$ ,  $-\psi$  and  $-\chi_i$

( $i = 1, \dots, N$ , where  $N$  stands for the number of freely rotating dihedral angles along the sidechain). In other words,  $P_D(\chi_1, \dots, \chi_N | \phi, \psi) = P_L(-\chi_1, \dots, -\chi_N | -\phi, -\psi)$ , where  $P_D$  stands for the conformational probability of a D-amino acid and  $P_L$  stands for the conformational probability of the same amino acid in L-configuration (30). Note that this does not consider interactions with other residues, neither whether the previous and the next residues are in L- or D-configuration. However, such information is typically not explicitly considered in rotamer libraries. For chiral sidechains, the previous relation does not hold because of the additional asymmetric centers. Therefore, we used the same strategy based on molecular dynamics simulations and renormalization that was used for L-sidechains in Gfeller *et al.* (19) to predict the rotamer libraries for chiral D-sidechains.

To help exploring the structural environment of non-natural sidechains incorporated in polypeptide chains, we developed visualization plugins for PyMOL and UCSF Chimera (31). These plugins enable users to mutate residues in protein structures to any sidechain (natural and non-natural, in both L- and D-configurations) present in the SwissSidechain database. This is

### Navigation

- Home
- Browse
- Visualization Plugins
- MD simulations
- D-amino acids
- Download
- License
- Contacts

## Non-natural sidechain 3MY

**3D view**

<b>Name</b>	3-Chloro-tyrosine
<b>Molecular Weight [g/mol]</b>	216.64
<b>logP</b>	-1.14 (predicted)
<b>Sidechain logP</b>	2.56 (predicted)
<b>Volume [Å<sup>3</sup>]</b>	224.29
<b>Sidechain Volume [Å<sup>3</sup>]</b>	138.63
<b>pKa</b>	1.56 / 9.47 / 7.84 (predicted)
<b>SMILES</b>	<chem>OC(=O)[C@H](C)[C@H](Cc1ccc(O)c(Cl)c1)O)[NH3]</chem>
<b>D-amino acid code</b>	D3MY

**PDB References:**

**L-amino acid**    PDB | Ligand

**D-amino acid**    PDB | Ligand

**PubChem**    L: 110992

**CAS number**    L: 7423-93-0 | D: 35608-63-0

**Download:**

Files	L-amino acid	D-amino acid
PDB		
MOL2		
2D structures (PNG)		
backbone-dependent rotamers		
backbone-independent rotamers		
Gromacs topology file		
CHARMM topology file		

Browsing options

Pymol + Chimera Plugins

Molecular Dynamics parameter files

**GROMACS**    **CHARMM**

Biochemical parameters

PDB

PROTEIN DATA BANK

PubChem

Structure files

Rotamer files

Topology files

**Figure 1.** Sidechain general information page. 2D and 3D views of the full amino acids are provided. Names, SMILES and several biochemical parameters are listed on the right part of the page, as well as links to other structural or chemical databases. Download options are provided in the lower part.

particularly useful to evaluate the structural environment of a new residue and detect possible clashes that may arise on mutation. Moreover, the resulting structures can be readily used for future investigations with molecular mechanics software. Detailed instructions for installing and running these plugins are provided on the SwissSidechain web site, together with short movies describing standard use for mutating a residue to a non-natural sidechain in a protein structure.

### Web interface

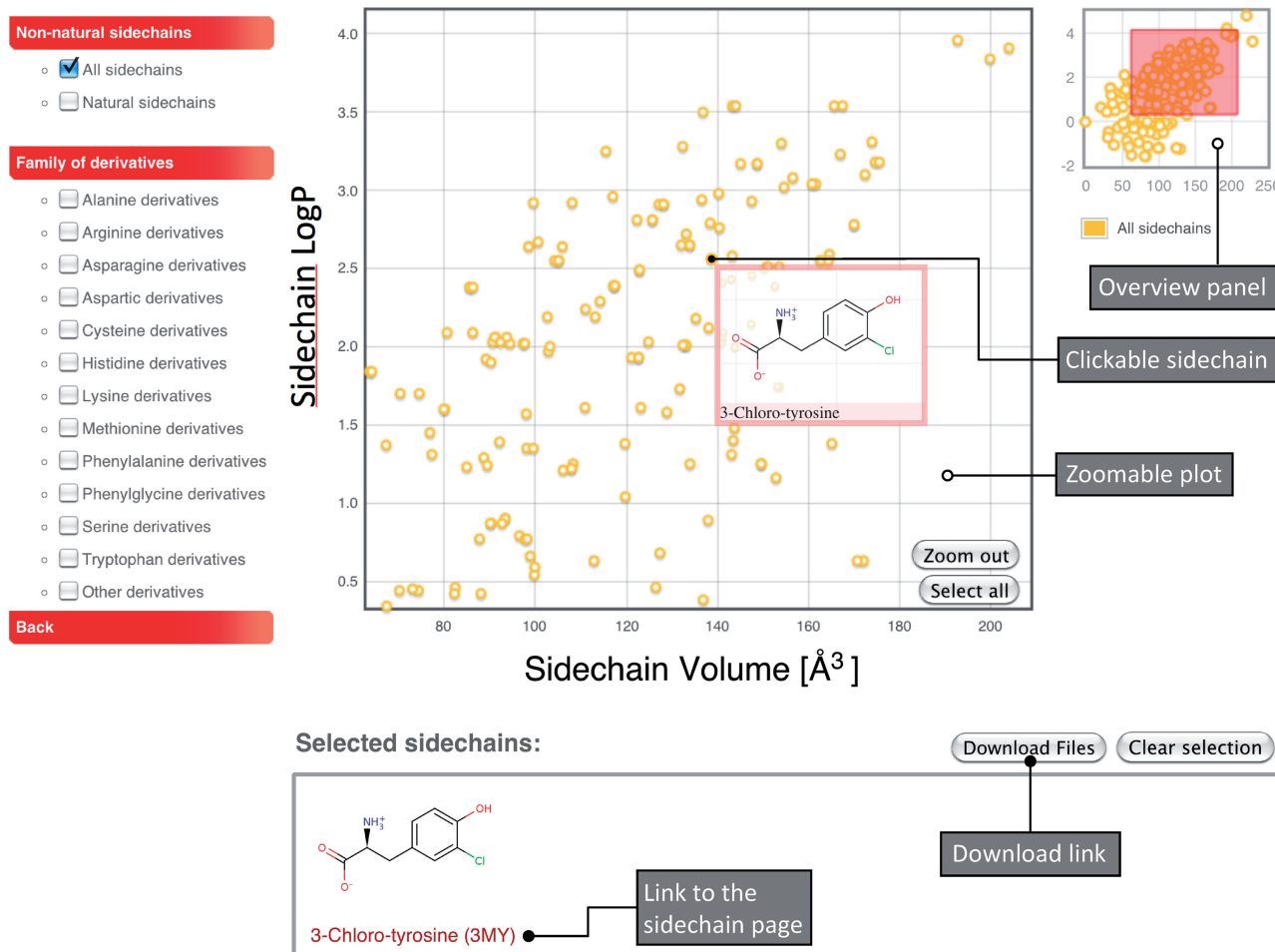
For each sidechain, all data are collected in dedicated web pages containing 2D and 3D structures, the different names, biochemical properties, links to existing databases and download options. An example is shown in Figure 1. The biochemical parameters listed on the right have been computed as described above. Links to existing databases include the PDB (24), the PDB Ligand Expo (32) and PubChem (33). Files to be downloaded consist of structure files containing 3D coordinates (PDB and Mol2 formats), 2D structures, rotamers (backbone dependent and

independent) and topology files for molecular mechanics software such as CHARMM (25) and GROMACS (29) Parameter files are available at the Molecular Dynamics (MD) simulations pages. These data are provided for both L- and D-configurations. The amino acid 3D structures for the L-configuration can be visualized online thanks to the open-source Java viewer for chemical structures Jmol (<http://www.jmol.org>). In addition we provide bulk download options for all data at the download page of SwissSidechain.

### Browsing options

SwissSidechain data can be browsed in large alphabetical tables containing the three- or four-letter codes and the sidechain full names, 2D structures, internal links to the sidechain pages, external links to the PDB, as well as download links for structural files. Both a full table and tables restricted to sub-families of natural sidechain derivatives (e.g. methionine derivatives) are provided.

Another alternative is to query data based on sidechain physico-chemical properties. Two particularly important



**Figure 2.** Sidechain chemical properties browsing plot. Each sidechain is represented by a yellow circle positioned according to the sidechain volume and logP. On mouseover, the amino acid 2D structure and full name appear. Sidechains can be selected by clicking on the circles or using the 'select all' button to select all sidechains in the plot. Selected sidechains appear in the box below the plot, where users can follow the links to the sidechain web page (see Figure 1) or download all structural and molecular files. Users can zoom on the graph by selecting subregions (see red rectangle in the overview panel). Specific families of sidechain derivatives can be selected in the menu on the left.



parameters are the volume and the hydrophobicity (logP) of the sidechains. In SwissSidechain, we provide an interactive 2D plot of these parameters. This enables users to zoom, visualize and select specific subsets of non-natural sidechains (see Figure 2). Sidechains can be selected by clicking on the corresponding circles or selecting a region of the graph and using the 'select all' button. Selected sidechains appear in the box below the graph and links to sidechain pages can be followed from there. This approach is especially useful for protein engineering and ligand design, to select a subset of sidechains with specific properties.

## CONCLUSION AND OUTLOOK

Exploiting naturally evolved, and often well optimized, proteins or peptides while not being restricted to the limited set of 20 natural amino acids is a promising approach for protein engineering with applications in biochemistry, protein structure and function analysis, and for drug design (3,10). Toward this goal, non-natural sidechains are being increasingly used in experimental studies. The SwissSidechain database provides a unified web resource for hundreds of non-natural sidechains. The biochemical parameters allow researchers to rapidly retrieve sidechains with specific properties, such as volume or hydrophobicity. The different visualization tools enable seamless introduction of non-natural sidechains in protein and peptide structures. The model parameters, such as bond and angle flexibility or partial atomic charges are useful to investigate non-natural sidechains with molecular mechanics software. Most of the non-natural sidechains are commercially available, so that researchers interested in using the predictions obtained with the SwissSidechain data can easily test them experimentally. The naming is consistent with the PDB so that published experimental structures can be readily analyzed with the tools provided in SwissSidechain.

In the current version of SwissSidechain, we focus on alpha amino acid sidechains that do not modify the backbone atoms (i.e. without branching at C $\alpha$ , modification of the nitrogen atom or longer backbones such as beta amino acids). These amino acids in their L-configuration are less likely to lead to large structural remodeling when incorporated into proteins or peptides. Therefore they are often used in protein engineering experiments, and predictions of their effect on protein structure and function are in general more accurate. The same amino acids in D-configuration have found many applications in biochemistry and drug discovery. Better characterizing D-amino acid properties is also useful to study D-retro-inverso peptides that can provide mimetics of L-peptides (34), or D-peptides obtained by screening L-peptides against D-targets and using the mirror-image property of D-enantiomers (35). Structural modeling of these types of non-natural sidechains is becoming feasible, although it is still more challenging. This is for instance the case with peptides binding to peptide recognition domains, such as Post synaptic density protein, Drosophila disc large tumor suppressor, and Zonula

occludens-1 protein or Src Homology 3 domains (36,37). As the binding interface is often relatively short, one may be able to probe *in silico* the effect of alpha-amino acids in D-configuration (38). For these different reasons, we have included data for the D-configuration of all amino acid sidechains present in the current version of SwissSidechain.

As experimental techniques keep progressing, more and more non-natural sidechains will become commercially available or be encountered in protein structures. We plan to include these new sidechains in future updates of the SwissSidechain database. Other developments may include focusing on different families of non-natural sidechains, such as fluorescent or photo-cleavable amino acids.

## ACKNOWLEDGEMENTS

The authors would like to thank Eric Pettersen for valuable help in building the UCSF Chimera plugin, and Aurélien Grosdidier for insightful comments on the database.

## FUNDING

European Molecular Biology Organization long-term fellowship [ALTF 241-2010 to D.G.]; Swiss Institute of Bioinformatics. Funding for open access charge: Swiss Institute of Bioinformatics.

*Conflict of interest statement.* None declared.

## REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2007) *Molecular Biology of the Cell*, 5th edn. Garland Science, New York.
- Wang, L., Xie, J. and Schultz, P.G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 225–249.
- Sievers, S.A., Karanicolas, J., Chang, H.W., Zhao, A., Jiang, L., Zirafi, O., Stevens, J.T., Munch, J., Baker, D. and Eisenberg, D. (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature*, **475**, 96–100.
- Burke, P.A., DeNardo, S.J., Miers, L.A., Lamborn, K.R., Matzku, S. and DeNardo, G.L. (2002) Cilengitide targeting of alpha(v)beta(3) integrin receptor synergizes with radioimmunotherapy to increase efficacy and apoptosis in breast cancer xenografts. *Cancer Res.*, **62**, 4263–4272.
- Parlati, F., Lee, S.J., Aujay, M., Suzuki, E., Levitsky, K., Lorens, J.B., Micklem, D.R., Ruurs, P., Sylvain, C., Lu, Y. *et al.* (2009) Carfilzomib can induce tumor cell death through selective inhibition of the chymotrypsin-like activity of the proteasome. *Blood*, **114**, 3439–3447.
- Smith, Y., Wichmann, T., Factor, S.A. and DeLong, M.R. (2012) Parkinson's disease therapeutics: new developments and challenges since the introduction of levodopa. *Neuropsychopharmacology*, **37**, 213–246.
- Turner, E.H. and Blackwell, A.D. (2005) 5-Hydroxytryptophan plus SSRIs for interferon-induced depression: synergistic mechanisms for normalizing synaptic serotonin. *Med. Hypotheses*, **65**, 138–144.
- Welch, B.D., VanDemark, A.P., Heroux, A., Hill, C.P. and Kay, M.S. (2007) Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl Acad. Sci. USA*, **104**, 16828–16833.
- Schorderet, D.F., Manzi, V., Canola, K., Bonny, C., Arsenijevic, Y., Munier, F.L. and Maurer, F. (2005) D-TAT transporter as an ocular peptide delivery system. *Clin. Experiment. Ophthalmol.*, **33**, 628–635.

10. Wang, Q., Parrish, A.R. and Wang, L. (2009) Expanding the genetic code for biological studies. *Chem. Biol.*, **16**, 323–336.
11. Ai, H.W., Shen, W., Sagi, A., Chen, P.R. and Schultz, P.G. (2011) Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem*, **12**, 1854–1857.
12. Kessler, B., Michielin, O., Blanchard, C.L., Apostolou, I., Delarbre, C., Gachelin, G., Gregoire, C., Malissen, B., Cerottini, J.C., Wurm, F. *et al.* (1999) T cell recognition of haptens. Anatomy of T cell receptor binding of a H-2kd-associated photoreactive peptide derivative. *J. Biol. Chem.*, **274**, 3622–3631.
13. Wang, J., Xie, J. and Schultz, P.G. (2006) A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc.*, **128**, 8738–8739.
14. Lemke, E.A., Summerer, D., Geierstanger, B.H., Brittain, S.M. and Schultz, P.G. (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.*, **3**, 769–772.
15. Kent, S.B. (1988) Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.*, **57**, 957–989.
16. Mandal, K. and Kent, S.B. (2011) Total chemical synthesis of biologically active vascular endothelial growth factor. *Angew. Chem. Int. Ed. Engl.*, **50**, 8029–8033.
17. Tian, F., Tsao, M.L. and Schultz, P.G. (2004) A phage display system with unnatural amino acids. *J. Am. Chem. Soc.*, **126**, 15962–15963.
18. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. and Chin, J.W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, **464**, 441–444.
19. Gfeller, D., Michielin, O. and Zoete, V. (2012) Expanding molecular modeling and design tools to non-natural sidechains. *J. Comput. Chem.*, **55**, 1525–1535.
20. Renfrew, P.D., Choi, E.J., Bonneau, R. and Kuhlman, B. (2012) Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS One*, **7**, e32637.
21. Revilla-Lopez, G., Rodriguez-Roperio, F., Curco, D., Torras, J., Isabel Calaza, M., Zanuy, D., Jimenez, A.I., Cativiela, C., Nussinov, R. and Aleman, C. (2011) Integrating the intrinsic conformational preferences of noncoded alpha-amino acids modified at the peptide bond into the noncoded amino acids database. *Proteins*, **79**, 1841–1852.
22. Revilla-Lopez, G., Torras, J., Curco, D., Casanovas, J., Calaza, M.I., Zanuy, D., Jimenez, A.I., Cativiela, C., Nussinov, R., Grodzinski, P. *et al.* (2010) NCAD, a database integrating the intrinsic conformational preferences of non-coded amino acids. *J. Phys. Chem. B*, **114**, 7413–7422.
23. Dunbrack, R.L. Jr (2002) Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.*, **12**, 431–440.
24. Rose, P.W., Beran, B., Bi, C., Bluhm, W.F., Dimitropoulos, D., Goodsell, D.S., Prlic, A., Quesada, M., Quinn, G.B., Westbrook, J.D. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
25. Brooks, B.R., Brooks, C.L. 3rd, Mackerell, A.D. Jr, Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
26. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R. (2011) Open Babel: An open chemical toolbox. *J. Cheminform.*, **3**, 33.
27. Wang, R., Fu, Y. and Lai, L. (1997) A new atom-additive method for calculating partition coefficients. *J. Chem. Inf. Model.*, **37**, 615–621.
28. MacKerell, J.A.D., Bashford, D., Bellott, M., Dunbrack, R.L. Jr, Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem. G*, **102**, 3586–3616.
29. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.*, **26**, 1701–1718.
30. Mitchell, J.B. and Smith, J. (2003) D-amino acid residues in peptides and proteins. *Proteins*, **50**, 563–571.
31. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
32. Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) Ligand Depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics*, **20**, 2153–2155.
33. Wang, Y., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Zhou, Z., Han, L., Karapetyan, K., Dracheva, S., Shoemaker, B.A. *et al.* (2012) PubChem's BioAssay Database. *Nucleic Acids Res.*, **40**, D400–D412.
34. Cardo-Vila, M., Giordano, R.J., Sidman, R.L., Bronk, L.F., Fan, Z., Mendelsohn, J., Arap, W. and Pasqualini, R. (2010) From combinatorial peptide selection to drug prototype (II): targeting the epidermal growth factor receptor pathway. *Proc. Natl Acad. Sci. USA*, **107**, 5118–5123.
35. Funke, S.A. and Willbold, D. (2009) Mirror image phage display—a method to generate D-peptide ligands for use in diagnostic or therapeutical applications. *Mol. Biosyst.*, **5**, 783–786.
36. Tonikian, R., Xin, X., Toret, C.P., Gfeller, D., Landgraf, C., Panni, S., Paoluzi, S., Castagnoli, L., Currell, B., Seshagiri, S. *et al.* (2009) Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. *PLoS Biol.*, **7**, e1000218.
37. Gfeller, D., Butty, F., Wierzbicka, M., Verschueren, E., Vanhee, P., Huang, H., Ernst, A., Dar, N., Stagljar, I., Serrano, L. *et al.* (2011) The multiple-specificity landscape of modular peptide recognition domains. *Mol. Syst. Biol.*, **7**, 484.
38. Yongye, A.B., Li, Y., Giulianotti, M.A., Yu, Y., Houghten, R.A. and Martinez-Mayorga, K. (2009) Modeling of peptides containing D-amino acids: implications on cyclization. *J. Comput. Aided. Mol. Des.*, **23**, 677–689.