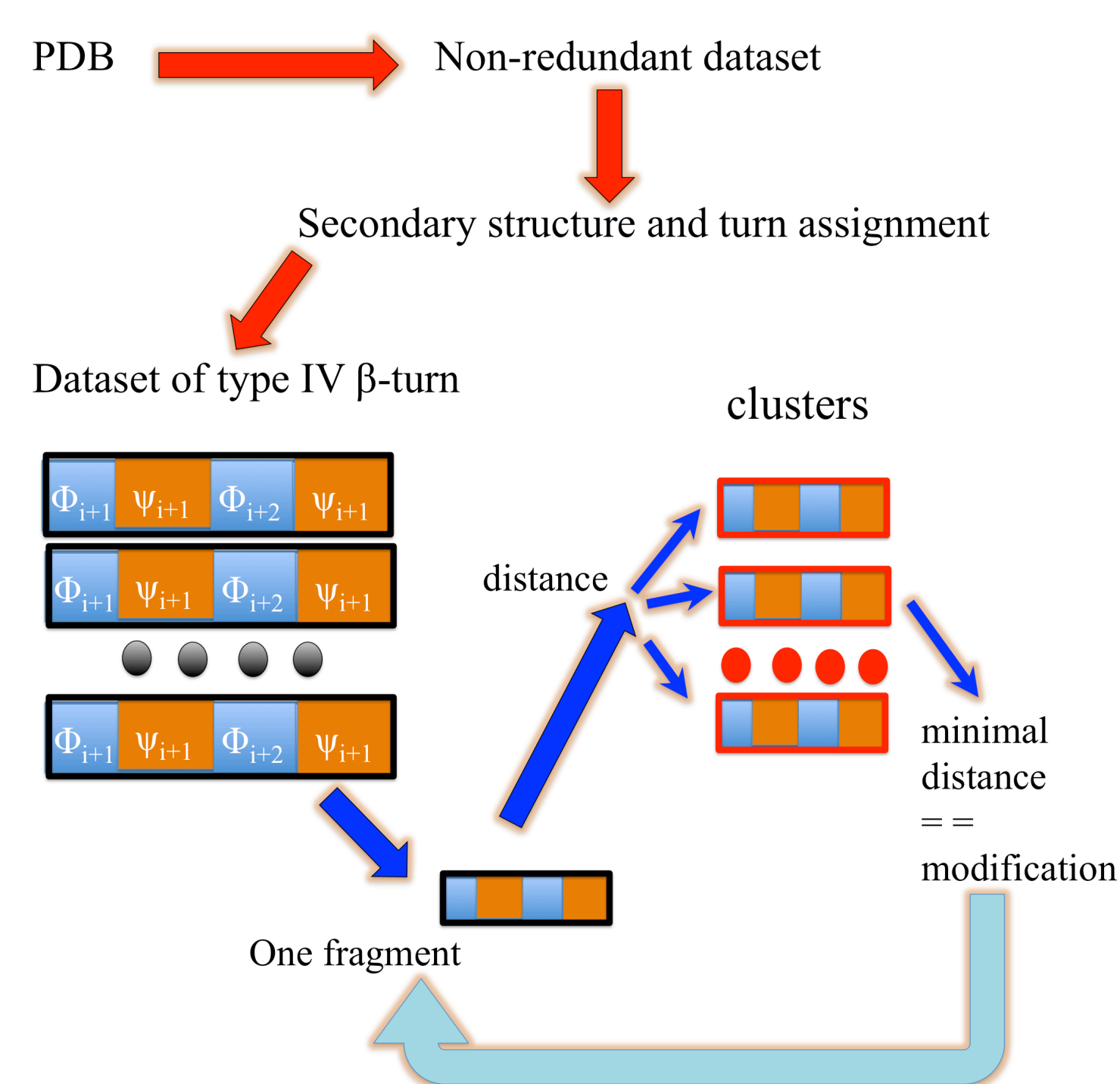




Abstract

The functional properties of a protein depend mainly on its three-dimensional (3D) structure. They are classically assigned, visualized and analysed through the prism of classical secondary structures composed of repetitive parts (α -helices for 1/3rd of the residues and β -strands for 1/5th, resp.) connected by coil. Two other repetitive structures also exist, namely the PolyProline II and the β -turns. The β -turns have been characterized by a hydrogen bond between N-H and C=O of residues i and $i+3$ by Venkatachalam [1]. He also characterized the first β -turn types. Later novel turns were defined, some being discarded, leading to a final collection of type I, I', II, II', IV, VIa1, VIa2, VIb, and VIII β -turns. Types VIa1, VIa2 and VIb are characterized by the presence of a cis-Proline at residue $i+2$. Turns that do not fit any of the above criteria are classified as type IV [2]. β -turn IV, *i.e.* the miscellaneous category, represents near 1/3rd of β -turn residues in protein structure, and is the second most frequent β -turn. 25 years have passed since the last proposition of an extension of the classical definition of β -turns. *Over all we have known*, it seems a good moment to dig them and to see if some new recurrent conformations are not hidden into this miscellaneous type. An automatic clustering approach based on the rules of β -turn type assignment was designed to search for recurrent new turns inside this miscellaneous type. The four most occurring clusters defined the new β -turn types. Surprisingly, these types, named IV₁, IV₂, IV₃ and IV₄, represent half of the type IV β -turns, and are more frequent than many established ones. Type IV₁, is in the neighbourhood of type II but with very different amino acid composition, while IV₂ is close to type VIII with related amino acid content. Types IV₃ and IV₄ are in the same dihedral angle region than frequent β -turn type I, but with distinct dihedral angle values [3].

In silico protocol

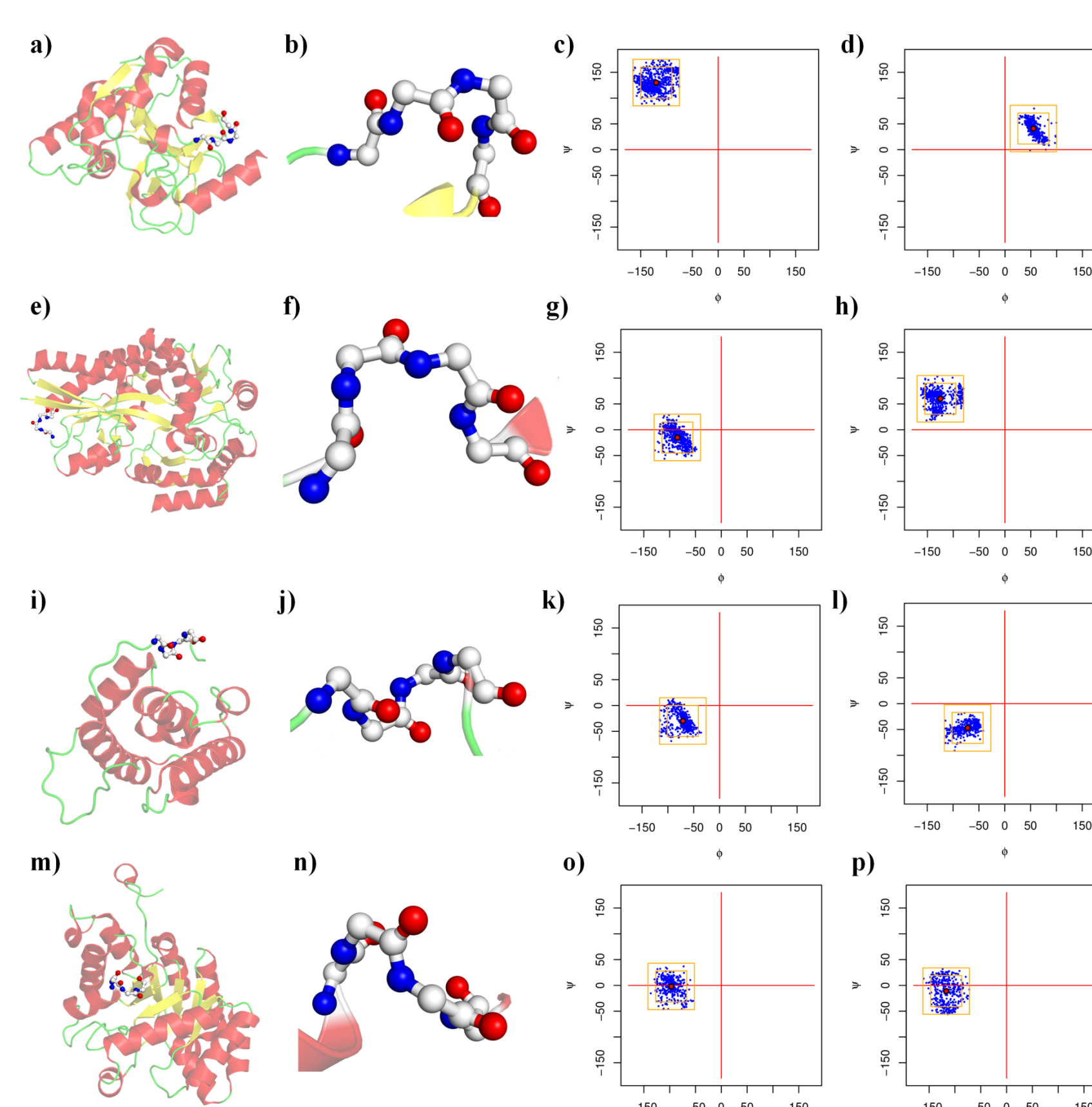


A specific clustering approach was designed to cluster type IV β -turns by using the classical rule, allowing $\pm 30^\circ$ for all angles, with the exception of one at $\pm 45^\circ$ for the defined values. The clustering derived from Self-Organizing Maps (SOM, without diffusion between the clusters). The training was carried out in 2 successive parts; the first one limited the potential bias of initialization, and the second refined the clustering by using the specific rules for β -turn types.

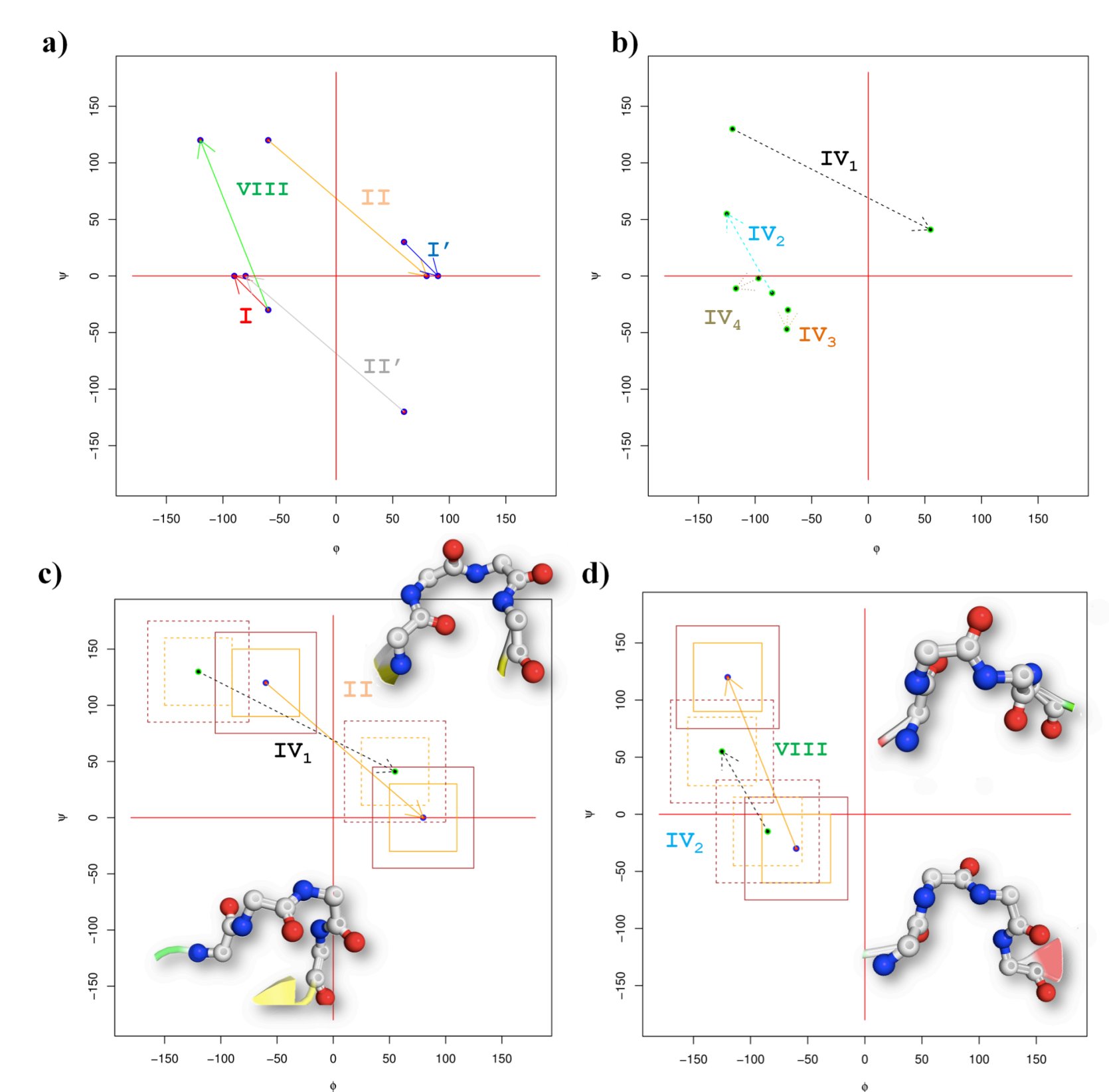
Amino acid analyses

	i	$i+1$	$i+2$	$i+3$
β -turn I	(+)	cPghStND	PSEK	whSTNDe
	(-)	IVLmAfYwQERK	IVLmFywqGhtn	IVLmAPG
β -turn II	(+)	qP	Pek	GN
	(-)	sD	ivLywcGst	IVLmAFywqPSTdERK
β -turn VIII	(+)	acPGs	PDek	IVFyhNd
	(-)	IvImqerk	ilfycGh	lAPG
β -turn I'	(+)	Fst	GNd	Gn
	(-)	Q	ivpt	ivlapsterk
β -turn II'	(+)	St	G	stN
	(-)	Vlafqpter	p	ivlp
β -turn VI _{a1}	(+)	Vp	afYp	P
	(-)	N	ilt	ip
β -turn VI _{a2}	(+)	N	ne	P
	(-)	G	ivlagstderk	h
β -turn VI _b	(+)	P	y	P
	(-)	G	ivlagstderk	pr
β -turn IV ^{ori}	(+)	CPGStnD	PGsndk	gHtND
	(-)	IVlaqerk	IVLmAc	IVLaqP
β -turn IV ₁	(+)	cG	hnek	GhND
	(-)	Ve	g	ivlaptk
β -turn IV ₂	(+)	PgS	pstDk	hND
	(-)	IvImrk	ivafg	ivIPG
β -turn IV ₃	(+)	CsnD	aP	vmt
	(-)	Ivak	eq	pg
β -turn IV ₄	(+)	Cpgsnd	ftmD	whtNd
	(-)	Vle	ivag	lpg
β -turn IV _{misc}	(+)	cPgnd	PGn	GhtNd
	(-)	Ivlar	ivLmAfC	ivLay

The four new turns



New β -turn representation. (a-d) Type IV₁, (e-h) type IV₂, (i-l) type IV₃, and (m-p) type IV₄. A turn close to the ideal values of its type (a, e, i, m) within a protein and (b, f, j, n) a close-up of the turn. (c, g, k, o) Ramachandran plot (ϕ , ψ) of residue $i+1$ and (d, h, l, p) of residue $i+2$; red dots are the ideal values. The number of observations of both residues is strictly identical.

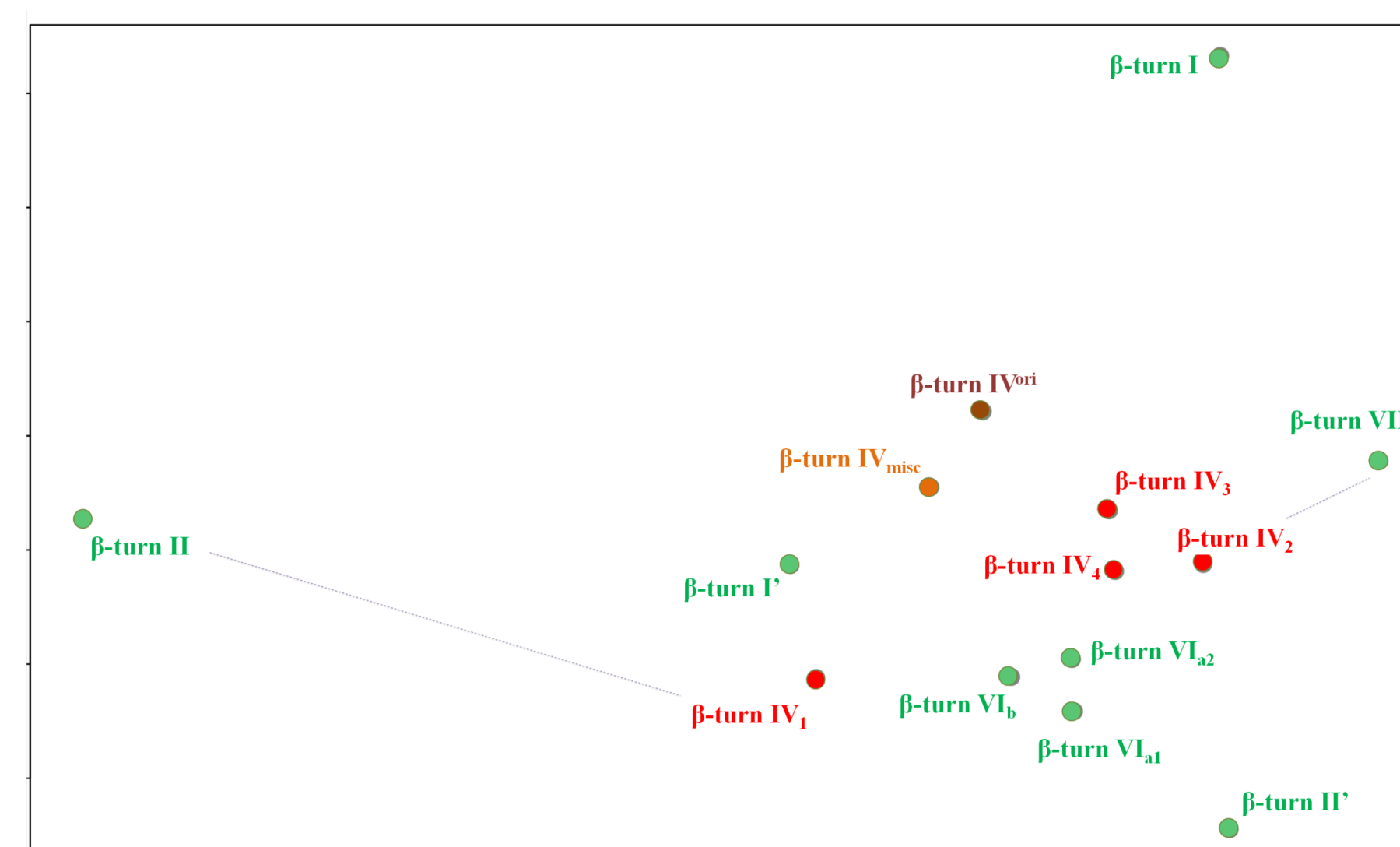


Ramachandran plot of the different β -turn types. An arrow connects the dihedral angle values of residue $i+1$ to residue $i+2$. (a) Classical β -turns, (b) new β -turns, (c) a close-up of type II and IV₁ β -turns, and (d) on type VIII and IV₂ β -turns, the first square corresponds to the $\pm 30^\circ$ rule, and the second one to the $\pm 45^\circ$ rule.

Conclusion

From an unsupervised classification, based exclusively on dihedral angles, four new types were defined. The two most occurring, type IV₁ and IV₂ β -turns, are linked to existing type II and VIII β -turns, but with very distinct features. On the one hand, type IV₂ and VIII β -turns shared striking amino acid compositional features with minor differences; although type IV₂ β -turn can be associated with stabilizing hydrogen bonds contrary to type VIII β -turn. While on the other hand, type IV₁ and II β -turns are very close in terms of dihedral angles but are very distinct in terms of amino acids content. It is clearly depicted that type II β -turn is highly specific while type IV₁ β -turn has more classical propensities, being closer to type I' β -turn than type II β -turn.

β -turn type	(%)	
1	I	38.21
2	IV ^{ori}	31.72
2	IV ^{misc}	16.44
3	II	11.81
4	VIII	9.84
5	IV ₁	5.10
6	I'	4.10
7	IV ₂	3.95
8	IV ₃	3.53
9	IV ₄	2.70
10	II'	2.51
11	VI _b	0.88
12	VI _{a1}	0.73
13	VI _{a2}	0.20



Sammon map of amino acid behaviours of the different β -turns. Classical turns are in green while new turns are in red.

Amino acid's Z-score of β -turn types. Colors underline the difference of new turns and the original type IV^{ori} (in green new over- or under representation, in blue inversion of over- or under representation,

References

- [1] Venkatachalam CM (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 6:1425-1436.
- [2] Hutchinson EG, Thornton JM (1996) PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci 5:212-220.
- [3] de Brevern AG (2016) Extension of the classical classification of β -turns. Sci Rep. 6:33191.

Acknowledgments

This work was supported by grants from the French Ministry of Research, University of Paris Diderot – Paris 7, French National Institute for Blood Transfusion (INTS), French Institute for Health and Medical Research (INSERM). AdB also acknowledges the Indo-French Centre for the Promotion of Advanced Research / CEFIPRA for collaborative grants (numbers 3903-E and 5302-2). This study was supported by grants from the Laboratory of Excellence GR-Ex, reference ANR-11-LABX-0051. The labex GR-Ex is funded by the programme "Investissements d'avenir" of the French National Research Agency, reference ANR-11-IDEX-0005-02. Calculations were performed on an SGI cluster granted by Conseil Régional Ile de France and INTS (SESAME Grant).