

Analysis of the interface variability in NMR structure ensembles of protein–protein complexes



Luisa Calvanese^{a,b,c}, Gabriella D'Auria^{a,b,c}, Anna Vangone^d, Lucia Falcigno^{a,b,c}, Romina Oliva^{e,*}

^a CIRPeB, University of Naples "Federico II", via Mezzocannone 16, 80134 Naples, Italy

^b Department of Pharmacy, University of Naples "Federico II", via Mezzocannone 16, 80134 Naples, Italy

^c Institute of Biostructures and Bioimaging – CNR, via Mezzocannone, 16, 80134 Naples, Italy

^d Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Faculty of Science–Chemistry, Utrecht University, Utrecht, Netherlands

^e Department of Sciences and Technologies, University Parthenope of Naples, Centro Direzionale Isola C4, I-80143 Naples, Italy

ARTICLE INFO

Article history:

Received 1 February 2016

Received in revised form 7 March 2016

Accepted 8 March 2016

Available online 9 March 2016

Keywords:

Protein complex

NMR

Interface

Inter-residue contacts

CONSRANK

COCOMAPS

ABSTRACT

NMR structures consist in ensembles of conformers, all satisfying the experimental restraints, which exhibit a certain degree of structural variability. We analyzed here the interface in NMR ensembles of protein–protein heterodimeric complexes and found it to span a wide range of different conservations. The different exhibited conservations do not simply correlate with the size of the systems/interfaces, and are most probably the result of an interplay between different factors, including the quality of experimental data and the intrinsic complex flexibility. In any case, this information is not to be missed when NMR structures of protein–protein complexes are analyzed; especially considering that, as we also show here, the first NMR conformer is usually not the one which best reflects the overall interface.

To quantify the interface conservation and to analyze it, we used an approach originally conceived for the analysis and ranking of ensembles of docking models, which has now been extended to directly deal with NMR ensembles. We propose this approach, based on the conservation of the inter-residue contacts at the interface, both for the analysis of the interface in whole ensembles of NMR complexes and for the possible selection of a single conformer as the best representative of the overall interface. In order to make the analyses automatic and fast, we made the protocol available as a web tool at: <https://www.mol-nac.unisa.it/BioTools/consrank/consrank-nmr.html>.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Protein–protein interactions are at the basis of most biological processes in the cell (Costanzo et al., 2010; Pawson and Nash, 2000). Many human diseases are in fact the result of abnormal protein–protein interactions (Ryan and Matthews, 2005; Vidal et al., 2011). The detailed characterization on a structural basis of protein–protein complexes is thus a crucial step in understanding and possibly mediating protein–protein interactions (Gonzalez-Ruiz and Gohlke, 2006; Mullard, 2012; Nisius et al., 2012). Nowadays, thousands of experimental structures of protein–protein complexes are available from the Protein Data Bank (Berman et al., 2000), a significant fraction of them being solved by solution NMR. Although NMR usually becomes the method of choice when

attempts to crystallize the complex have failed, in principle it offers the advantage of depicting a dynamic rather than a static view of the structure (Foster et al., 2007). Mobile regions of NMR structures are indeed also used to infer protein intrinsic disorder (Di Domenico et al., 2012; Martin et al., 2010; Potenza et al., 2015).

Results of NMR structure determination are released to the PDB as ensembles of N conformers (or models), with N typically around 20, all of them satisfying the experimental restraints. Authors are requested to identify a representative model and place it as first model in the ensemble. To this aim, models are generally ranked either on the basis of the target function (TF), measuring the agreement between a structure and the used set of experimental constraints, or if energetically minimized, based on the lowest energy (LE). Submitted conformers will feature a certain degree of structural variability. In particular, when coming to biomolecular complexes, more or less significant differences will be observed regarding their biological interface and this dynamical view of NMR complexes could in principle add valuable information on the recognition process.

* Corresponding author.

E-mail addresses: luisa.calvanese@unina.it (L. Calvanese), gabriella.dauria@unina.it (G. D'Auria), a.vangone@uu.nl (A. Vangone), lucia.falcigno@unina.it (L. Falcigno), romina.oliva@uniparthenope.it (R. Oliva).

A number of useful tools have been developed to the aim of: clustering models from an NMR ensemble in conformationally related subfamilies (Kelley and Sutcliffe, 1997), analysing and visualizing their structural variability/mobility (Martin et al., 2010; Scott and Straus, 2015; Sikic and Carugo, 2009; Theobald and Wuttke, 2006), identifying conserved regions and/or quantifying their similarity (Kelley and Sutcliffe, 1997; Konagurthu et al., 2010; Snyder and Montelione, 2005), or selecting the best representative (for instance to be used in protein structure prediction) (Kelley and Sutcliffe, 1997; Tosatto and Battistutta, 2007). Tools that allow enriching the conformational ensemble to get insight into the protein dynamics (Laughton et al., 2009) have also been proposed. However, all these tools have been conceived and developed for the analysis of single proteins, not for protein complexes. In fact, no tool specifically devoted to the analysis of what is of major interest in a protein complex, i.e. its biological interface, has been proposed to date. To fill this gap, we propose here a method to automatically analyze the interface in NMR ensembles of protein complexes and to possibly select a structure that is the best representative of the ensemble, in terms of inter-residue contacts at the interface. The binding interface and the network of contacts between two proteins are indeed crucial features to take into account in the study of a protein–protein complex (and its related function). The importance of interfacial contacts has been established, for instance, in the assessment of the quality of docking models (Lensink and Wodak, 2013) and, more recently, in the prediction of complexes binding affinity (Vangone and Bonvin, 2015). In this context, we have previously shown that the conservation of inter-residue contacts can be successfully applied to the analysis of the interface in structure ensembles of protein–protein complexes, both from docking (Oliva et al., 2015, 2013; Vangone et al., 2014, 2013, 2012) and molecular dynamics (MD) simulations (Abdel-Azeim et al., 2014a,b; Lancellotti et al., 2015; Oliva et al., 2015). In particular, we introduced numerical values for the conservation, in a given structure ensemble, both of single contacts (as the fraction of models featuring that contact) and of the overall interface (as the fraction of inter-residue contacts common to a given percentage of models). The contacts conservation can also be visually represented in a “consensus contact map”, i.e. an inter-molecular contact map where the conservation rates of the different contacts are represented on a gray scale (the more conserved the contact, the darker the dot). We have also proposed a consensus approach for the ranking of docking model ensembles, based on their ability to match the most conserved contacts at the interface (Oliva et al., 2013). Based on this approach, by definition, the model that is the best representative of the overall interface will be top ranked.

Herein, we extend the above approach to the analysis of NMR structures of protein complexes and in particular to all the 70 structures of heterodimeric complexes with a PDB entry, for which 20 conformers have been released. Based on the measures we introduced (Vangone et al., 2012), we show here that this set of protein–protein complexes span a large range of interface variability, that is not simply related to the different size of the interacting proteins or of the complex interface. Therefore, the interface variability, peculiar to each complex, should be taken into account when analyzing a solution NMR structure of a protein complex, or crucial information on the contacts at the interface risks to be missed. We also show that the 1st NMR conformer is usually not the best representative of the overall interface. Therefore, our ranking approach can be considered as a complementary tool to select one representative of the NMR structure ensemble, when attention is focused on its biological interface.

On these bases, we propose our approach to analyze NMR conformational ensembles of protein–protein complexes, in order to: (i) quantify and visualize their interface variability and (ii) select an interface representative. All described analyses can be automat-

ically performed through a dedicated web tool (now added as “NMR functionalities” to the CONSRANK server (Chermak et al., 2015)).

2. Materials and methods

2.1. NMR complexes

70 PDB structures were selected from the PDB database (<http://www.ebi.ac.uk/pdbe/>) (Velankar et al., 2010, 2015)), updated to May 5th 2015, based on the following criteria: solved by solution NMR, containing two entities, polymer type “protein”, both chains at least 30 amino acids long, 20 conformers submitted. Oligomers larger than dimers and structures obtained by docking simulations have been excluded. Although no filter has been used to avoid redundancy, the derived set of structures is non-redundant in terms of sequence. Indeed, when the same two proteins interact in two different complexes, at least one of them corresponds to different fragments and/or sources. The only two exceptions are represented by the structures corresponding to the PDB IDs: 2ktf and 2lof, with chain B sharing, within longer fragments, a 27-mer sequence 96% identical, and structures 2lqh and 2lqi, as they feature exactly the same protein sequences, but they interestingly exhibit completely different binding modes.

2.2. Interface analysis

Analyses of the conformer ensembles were performed on the above set of 70 structures using the now added NMR functionalities in the CONSRANK web server (Chermak et al., 2015), available at <https://www.molnac.unisa.it/BioTools/consrank/consrank-nmr.html>, where a sample output is also available, by the “load precalculated data” option. For details on the CONSRANK output see (Oliva et al., 2013). Briefly, given an ensemble of N conformers, for each inter-residue contact the conservation rate, CR_{kl} , is defined as in Eq. (1),

$$CR_{kl} = nc_{kl}/N, \quad (1)$$

where nc_{kl} is the total number of conformers where residues k and l are in contact. The conservation rate thus ranges between $CR_{kl} = 0$, if the contact between residues k and l is never observed, to $CR_{kl} = 1$, if the contact is observed in all the conformers. Once the conservation rates have been calculated, for each conformer i a score is first calculated as in Eq. (2), where M_i is the total number of contacts in conformer i . Then, a normalized score, \bar{S}_i , is calculated as in Eq. (3):

$$S_i = \sum_1^{M_i} CR_{kl}, \quad (2)$$

$$\bar{S}_i = S_i/M_i \quad (3)$$

Conformers are ranked according to their \bar{S}_i value, basically corresponding to the average conservation of the inter-residue contacts in each conformer. C_{70} is calculated as in Eq. (4), where nc_{70} is the total number of inter-residue contacts conserved in at least 70% of the analyzed frames and nc_i is the total number of inter-residue contacts in frame i . C_{50} and C_{90} are calculated similarly.

$$C_{70} = \frac{nc_{70}}{\sum_i^N \frac{nc_i}{N}} \quad (4)$$

Within this work, two residues are defined in contact if any pair of atoms belonging to the two residues is closer than a cut-off distance of 5 Å. Conservation rates are additionally plotted in the form of consensus contact maps for visualization purposes (Vangone

Table 1

Main features of analyzed NMR structures, with relative interface conservation (C50, C70 and C90) and number of the NMR model top ranked by CONSRANK.

PDB ID	Description	Release year	Selection criteria ^a	Length (L) ^b	Length (R) ^c	Residues at the interface	Interface area (Å ²)	C50 ^d	C70	C90	NMR model top ranked by CONSRANK
1cee	Cdc42/ACK	1999	TF	59	181	72	1413	0.837	0.906	0.974	20
1cf4	Cdc42/WASP	1999	LE	44	184	95	1868	0.935	0.642	0.398	20
1e0a	Cdc42/Pak1	2000	TF	46	184	65	1322	1.001	0.866	0.701	19
1eci	Ectatomin A/B	1995	na	34	37	38	722	1.036	0.893	0.793	16
1ees	Cdc42/Pak3	2000	LE	46	178	97	1945	0.849	0.530	0.350	3
1kbh	CBP/p300	2002	TF	47	59	81	1655	0.981	0.840	0.707	9
1l8c	Hif-1 alpha/CBP	2002	TF	51	95	96	2228	0.921	0.816	0.734	9
1otr	Cue2/ubiquitin	2003	TF	49	76	30	555	0.950	0.765	0.581	14
1q5w	Npl4 Zn finger/ubiquitin	2004	TF	31	76	24	390	1.005	0.908	0.778	15
1r8u	CBP TAZ1/CITED2	2004	TF	50	100	83	1894	0.949	0.853	0.731	20
1sse	YAP1	2004	LE	35	86	25	463	0.892	0.679	0.552	20
1tlh	T4 AsiA/Sigma70 region 4	2004	TF	68	89	51	1044	0.939	0.805	0.687	1
1uel	HR23B/S5a	2004	LE	48	95	37	723	0.925	0.834	0.722	9
1wr1	Dsk2p UBA/ubiquitin	2005	LE	58	76	22	431	0.919	0.813	0.777	16
1z00	ERCC1/XPF	2005	LE	84	84	74	1533	0.946	0.875	0.834	16
2dt7	SF3a60/SF3a120	2006	TF	38	85	31	636	0.825	0.825	0.795	20
2e30	p22/NHE1 fragment	2006	LE	43	195	78	1511	0.930	0.723	0.563	13
2fho	p14/SF3b155	2005	LE	47	87	39	911	0.934	0.863	0.755	14
2fv4	Spc24/Spc25	2006	LE	58	89	55	1004	0.978	0.728	0.432	1
2g3q	Ede1-UBA/ubiquitin	2006	LE	43	76	26	449	1.016	0.806	0.631	16
2ggp	Atx1/CCC2a	2006	TF	71	73	27	488	0.924	0.765	0.633	11
2h7d	Talin-1/integrin beta3-Pip5k1c (chimera)	2007	LE	34	101	46	1001	1.060	0.883	0.805	17
2jt4	Sla1-SH3/ubiquitin	2007	TF	71	76	32	595	1.041	0.795	0.384	7
2jzb	NusA/aCTD	2008	LE	70	81	23	344	0.984	0.656	0.492	17
2k1r	MNK1/ATOX1	2008	TF	68	73	29	427	1.036	0.725	0.440	3
2k2u	TFIIH-Tfb1/VP16	2008	LE	35	115	31	566	0.898	0.838	0.659	13
2k42	WASP/EspF(U)	2008	TF	36	72	48	950	0.876	0.783	0.676	12
2k4a	SYT1/FGF1	2008	LE	128	133	34	618	1.038	0.898	0.645	5
2k7a	SH2-ITK/SH3-ITK	2009	TF	63	108	45	738	0.972	0.739	0.544	9
2ka4	CBP-TAZ1 (340–439)/STAT2-TAD (786–838)	2009	LE	128	133	103	2324	0.913	0.814	0.734	1
2ka6	CBP-TAZ2 (1764–1855)/STAT1-TAD (710–750)	2009	LE	45	92	69	1551	0.916	0.862	0.775	14
2kbx	PINCH/ILK	2008	LE	70	171	42	641	0.699	0.385	0.070	14
2kc8	RelE/RelBc	2009	LE	33	95	68	1482	0.962	0.864	0.678	15
2kdu	Calmodulin/Unc13a	2009	LE	36	148	69	1455	0.851	0.664	0.551	14
2kfk	Bem1/Cdc24-PB1	2009	LE	78	86	36	660	0.960	0.826	0.670	3
2khs	SNase121/SNase (111–143)	2009	LE	35	121	66	1432	0.817	0.649	0.456	6
2kj4	VEK-30/plasminogen	2009	LE	32	87	31	624	0.935	0.805	0.623	18
2kje	CBP/E1A	2009	LE	42	92	60	1265	0.941	0.780	0.631	11
2knb	Park2/endophilin-A1	2009	LE	62	76	31	650	0.976	0.803	0.631	16
2knc	Alphallb/Beta3	2009	LE	54	79	42	763	0.836	0.809	0.783	19
2kt5	Ref2/ICP27	2011	LE	40	124	39	694	0.742	0.611	0.545	5
2ktf	Rad30b-UBM2 (Hs, 45-aa)/ubiquitin	2010	LE	32	76	38	712	0.985	0.887	0.867	13
2kwu	Rad30b-UBM2 (Mm: 673–717)/ubiquitin	2010	LE	36	76	33	563	0.995	0.931	0.642	16
2kwv	Rad30b-UBM1 (Mm: 487–532)/ubiquitin	2010	LE	36	76	33	582	0.927	0.711	0.556	2
2kxh	FIR/FBP	2010	LE	31	199	38	689	0.894	0.723	0.647	20
2l0f	Rad30b-UBM2 (Hs, 32-aa)/ubiquitin	2010	LE	45	76	48	842	1.016	0.903	0.758	9
2l14	CBP/p53	2011	LE	49	59	53	1166	0.735	0.591	0.399	12
2l27	ECD1-CRF-R1/peptide agonist	2011	LE	38	84	27	552	0.864	0.702	0.459	16
2l29	IGF2/IGF2R	2012	LE	67	142	40	792	0.891	0.629	0.393	7
2l2l	p66alpha/MBD2	2011	LE	36	43	34	662	0.901	0.685	0.577	4
2l53	Calmodulin/NaV1.5 (IQ motif)	2011	LE	31	148	43	972	0.957	0.827	0.714	18
2l9s	PHF12/Sin3a	2011	TF	45	94	50	1061	0.858	0.771	0.543	7
2ld7	Sap30/Sin3a	2011	LE	75	94	81	1635	0.972	0.872	0.709	7
2lkm	PHF12/MORF4L1	2012	LE	42	172	67	1512	0.987	0.870	0.743	3
2lqh	CBP/FOXO3a	2012	LE	52	87	78	1510	0.795	0.588	0.450	11
2lqi	CBP/FOXO3a	2012	LE	52	87	80	1568	0.903	0.677	0.493	20
2luh	Vta1/Vps60	2012	LE	59	167	78	1638	1.016	0.864	0.806	14
2lww	RelA-TAD/CBP-TAZ1	2013	LE	70	100	105	2438	0.897	0.834	0.658	16
2m5a	Spa/ZpA963 (synthetic affibody)	2013	TF	58	58	40	819	0.929	0.856	0.820	10
2mjf	Rsa1/Hit1	2014	LE	40	95	71	1623	0.968	0.866	0.786	11
2mre	Rad18-UBZ/ubiquitin	2014	TF	33	79	26	539	0.901	0.817	0.704	4
2pq4	RPA14/32	2007	LE	35	90	40	646	0.935	0.853	0.633	19
2rms	Sin3a/Sap25	2008	TF	61	71	41	741	0.940	0.850	0.761	14

(continued on next page)

Table 1 (continued)

PDB ID	Description	Release year	Selection criteria ^a	Length (L) ^b	Length (R) ^c	Residues at the interface	Interface area (Å ²)	C50 ^d	C70	C90	NMR model top ranked by CONSRANK
2nrn	TFIIIE-alpha/TFIIH-p62	2008	LE	62	108	52	1242	0.898	0.820	0.618	3
2roz	Fe65L1/Apbb2	2008	TF	32	136	45	843	0.911	0.867	0.794	16
2rr3	VAP-A/OSBP1	2010	LE	44	128	36	742	0.921	0.844	0.691	19
2rsy	CSK/CBP	2013	TF	38	99	48	884	1.076	0.840	0.790	1
2ru4	YIIM2/MAII (<i>de novo</i> protein)	2014	LE	84	115	67	1261	0.680	0.613	0.493	20
4beh	RRP1/RPP2	2013	LE	114	116	54	1076	0.902	0.752	0.602	13
4uqt	IST3/CWC26	2014	LE	35	89	38	798	1.008	0.851	0.764	7

^a Selection criterion as reported in the corresponding wwPDB entry. *Abbreviations*: TF = lowest target function value; LE = lowest energy value; na = not available.

^b Length of the shorter interacting protein (ligand).

^c Length of the longer interacting protein (receptor).

^d By definition, C50 is obtained dividing the number of contacts common to at least 50% of the models by the average number of contacts in all the models (see Section 2). Therefore, it may happen that the numerator is larger than the denominator, thus making the ratio slightly exceeding 1.

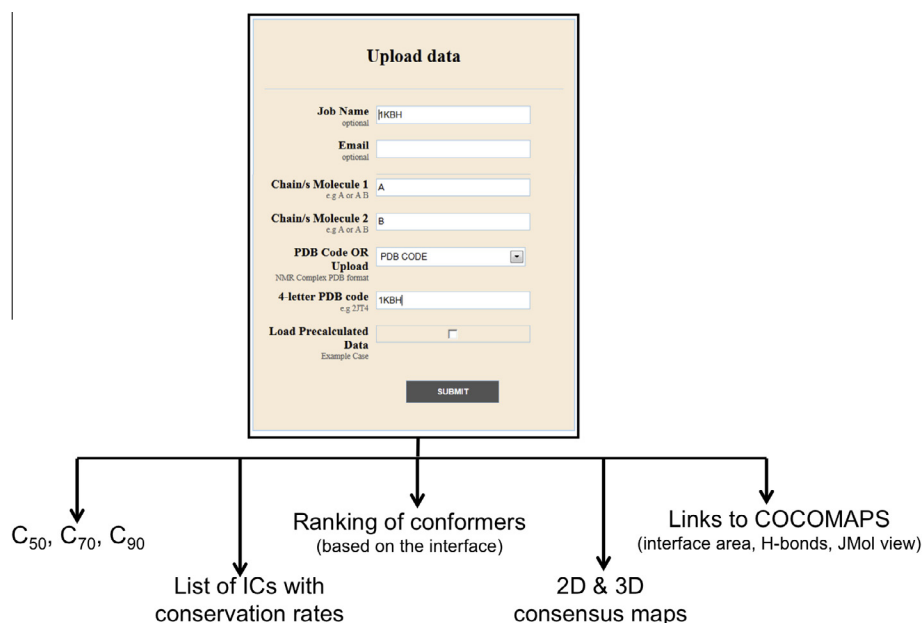


Fig. 1. Example input and schematic output for an NMR protein complex. ICs stay for “inter-residue contacts”.

et al., 2012). Interface areas and number of residues at the interface for single conformers were calculated with NACCESS (Hubbard and Thornton, 1993) through the COCOMAPS server (Vangone et al., 2011). Information on the complexes, such as chain length, complex name, method of structure selection and release date were extracted from the corresponding PDB entries.

The structure corresponding to the PDB entry 2kbn has not been used for the calculation of Pearson's correlation coefficients, as it clearly is an outlier, having all the C_{50} , C_{70} and C_{90} values falling more than 1.5 times the interquartile range below the first quartile of data (thresholds of 0.776, 0.514 and 0.236, respectively) (Moore and McCabe, 1998).

3. Results and discussion

We searched the PDBE (Velankar et al., 2010, 2015) for protein-protein complexes solved by solution NMR. We limited our search to heterodimeric complexes. As we are interested in characterizing protein-protein complexes rather than peptide-protein complexes, we have excluded too short interactors, by setting a lower threshold of 30 residues on the shorter (ligand) chain. Furthermore, to make any measure of the interface conservation independent of

the ensemble size, we only selected structures for which 20 conformers (indeed the most common choice) were made available from the PDB by the authors of the structures. Complexes whose structures were obtained based on docking simulations were not considered. This left us with a set of 70 structures, whose details are reported in [Table 1](#), together with some results of herein performed analyses.

To estimate the interface conservation of each NMR ensemble, we used an approach based on the conservation of contacts at the interface, which is embodied in the CONSRANK algorithm and web server (Chermak et al., 2015; Oliva et al., 2013). The CONSRANK web server has been originally conceived and developed for the analysis of ensemble of docking models, to be uploaded by the user. Now, a functionality to directly download an NMR structure from the PDB and unpack it in the single conformer structures has been provided. It is available at the URL: <https://www.mol-nac.unisa.it/BioTools/consrank/consrank-nmr.html>. The input interface is extremely friendly. Users are only requested to fill up three fields i.e. the PDB 4-character alphanumeric identifier (PDB ID) and the chain identifiers for the two interactor chains (see Fig. 1 for an example). Optionally, users can also upload their own PDB file, specify the job name and/or indicate an e-mail address to which results will be sent.

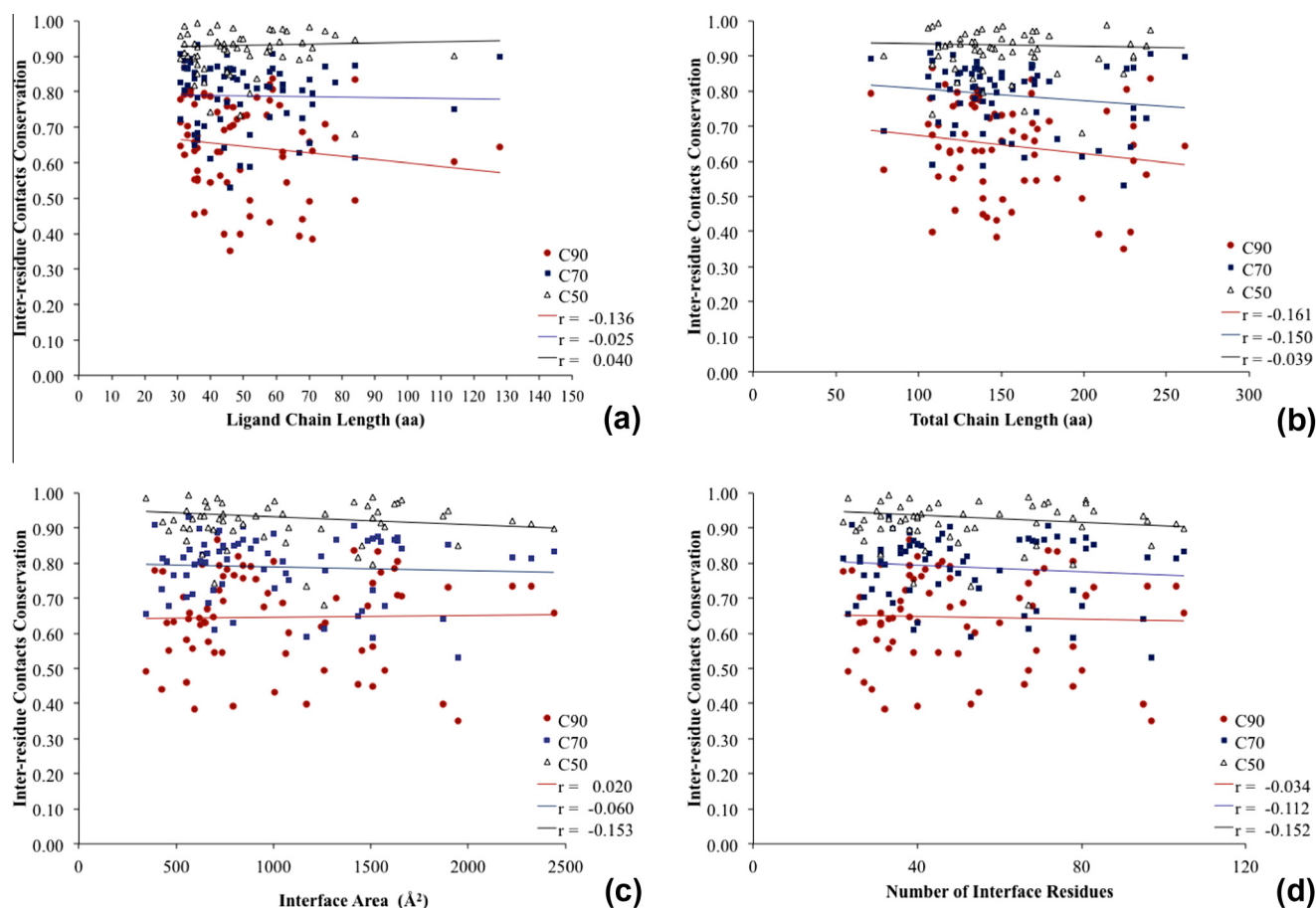


Fig. 2. Conservation of the contacts at the interface *versus*: (a) the chain length of the shorter interactor (ligand), (b) the total chain length (ligand + receptor), (c) the interface area and (d) the total number of residues at the interface. C50, C70 and C90 represent the fraction of contacts common to at least 50%, 70% and 90% of the conformers, respectively.

The output, available in few seconds, is a standard CONSRANK output, reporting: (i) a table with the inter-residue contacts ranked according to their conservation rate (CRkl, corresponding to the fraction of models in which the given contact is found), (ii) values reflecting the fraction of inter-residue contacts common to 50%, 70% and 90% of the models in the ensemble (C₅₀, C₇₀ and C₉₀, respectively), reflecting the overall conservation of the interface contacts in the ensemble, (iii) a table with the analyzed conformers, ranked according to how well they match the most conserved contacts in the models ensemble. Please note that the CONSRANK ranking does not relate in any way to the experimental quality of the models. Simply, conformers that are the best representatives of the overall interface will be top ranked according to this scheme.

Further, 2D and 3D consensus maps (where the third dimension is given by the conservation rate of each inter-residue contact) are provided, which can be zoomed and navigated to visualize the identity of the residues pairs corresponding to a given contact, as well as its conservation. It is important to notice that results of the above analyses are independent of the mutual orientation of the conformers in the input PDB file. Finally, a direct link is given for each NMR model to the corresponding COCOMAPS analyses, providing detailed information on its interface, including the interface area, a list of the inter-molecular H-bonds and an 3D visualization of the complex in Jmol (<http://www.jmol.org/>) (Vangone et al., 2011). A sample output to be immediately visualized is available from the web interface.

In the following, we report results of the CONSRANK analyses on the above set of NMR protein complexes. First of all, we consid-

ered the values of the C₅₀/C₇₀/C₉₀ parameters, as they immediately reflect the overall conservation/variability of the complex interface (the higher the number, the more conserved the interface). As a first observation, we found that such values are quite diverse in the set of structures we analyzed. After excluding the 2kbx structure, which is a clear outlier (see Section 2) having an extremely poor interface conservation, we observe C₅₀, C₇₀ and C₉₀ values spanning the ranges 0.68–1.1, 0.53–0.93 and 0.35–0.87 and assuming average values of 0.93 ± 0.08 , 0.79 ± 0.09 and 0.65 ± 0.13 , respectively. This means, for example, that 65% of inter-residue contacts are conserved on average in 90% of models (C₉₀), a value that raises to 87% and decreases to 35% for the structures featuring the least and most variable interface, respectively.

In Fig. 2, the C₅₀, C₇₀ and C₉₀ values are plotted *versus* the complex size (measured as the ligand and ligand plus receptor chain length), and the size of the complex interface (measured as the interface area and the total number of residues at the interface). Values of the corresponding Pearson's correlation coefficients are also reported. From Fig. 2 it is clear that the conservation of contacts at the interface in the analyzed structures set does not simply correlate either with the size of the complex or with the size of the interface, as one would maybe expect. Values of the Pearson's correlation coefficients stay indeed within the range of ± 0.16 , corresponding to a minimum *P*-value of 0.19.

Most probably, the interface conservation of each NMR structure is the result of a complex interplay between the quality of the experimental data, the methodology adopted for building the three-dimensional models and the intrinsic complex flexibility.

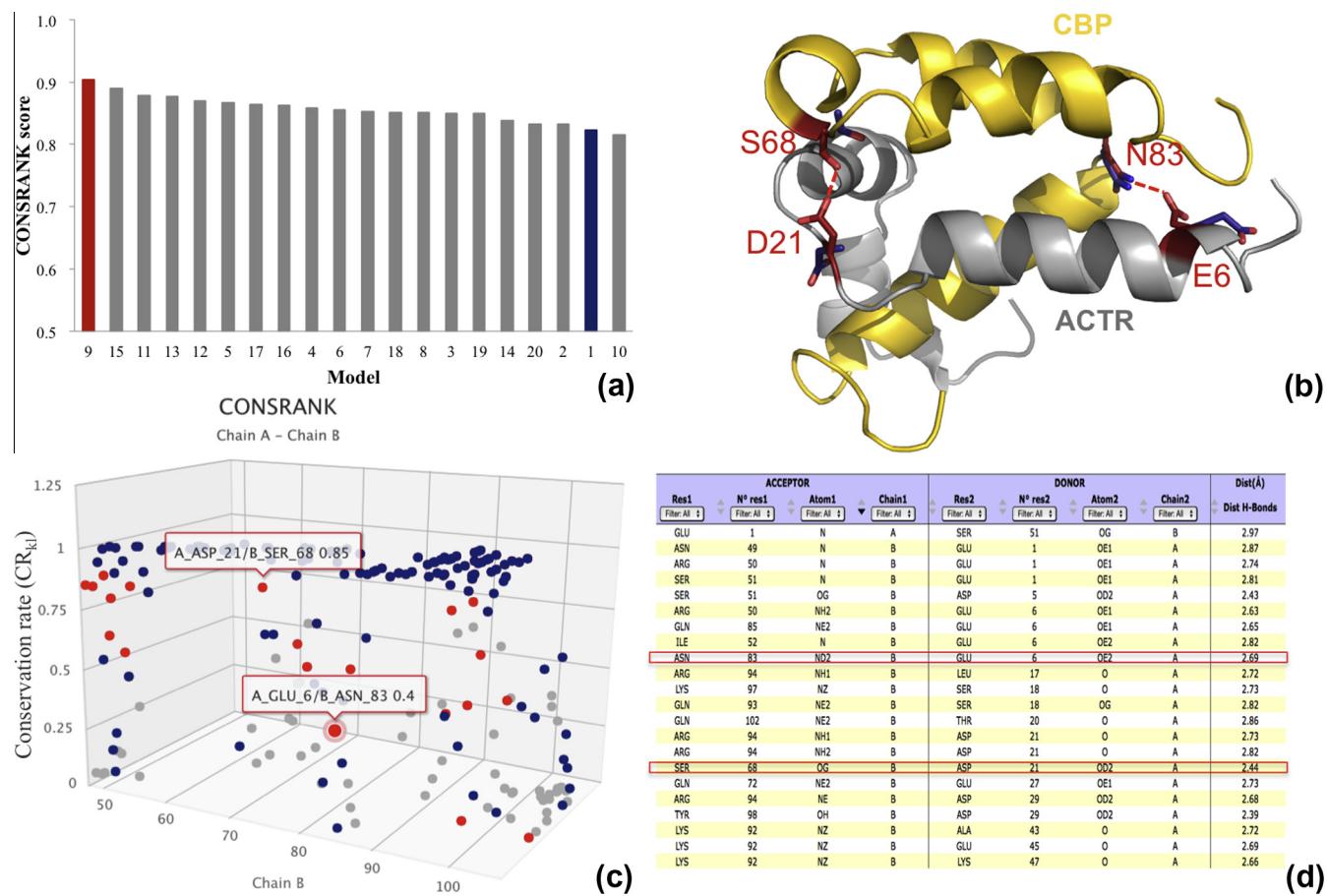


Fig. 3. Results of the analysis of the 1KBH complex. (a) Plot of the CONSRANK ranking of the 20 NMR conformers. (b) 3D representation of the complex: conformer ranked 1st by CONSRANK (corresponding to the 9th submitted NMR conformer) is shown as a cartoon; side chains of residues CBP-Ser68 and Asn83 and of ACTR-Asp21 and Glu6 for the 1st CONSRANK and the 1st NMR conformers are shown as red and blue sticks, respectively. (c) Interactive CONSRANK 3D consensus map: contacts corresponding to the consensus are colored gray, contacts corresponding to the 1st CONSRANK and the 1st NMR conformers are colored red and blue, respectively. Contacts corresponding to CBP-Ser68/ACTR-Asp21 and CBP-Asn83/ACTR-Glu6 are labeled, with values of the corresponding conservation rates, 0.85 and 0.40, indicated. Chains A and B represent the ACTR and sequences, respectively; the third (vertical) axis represents the conservation rate of each contact in the NMR conformers ensemble. (d) List of inter-molecular H-bonds for the 1st CONSRANK conformer, as obtained by COCOMAPS. H-bonds corresponding to the contacts at point (c) are boxed in red.

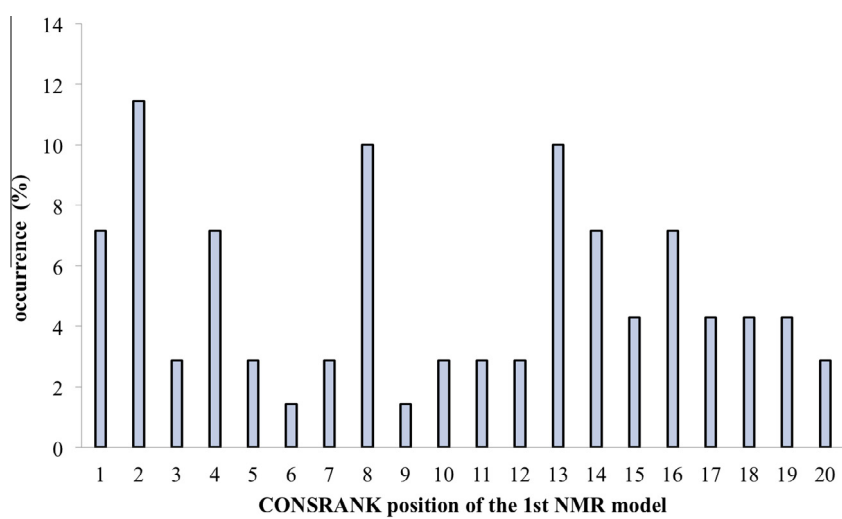


Fig. 4. CONSRANK ranking of the 1st NMR model. The plot reports the fraction of 1st NMR models ranked by CONSRANK in the 1st to the 20th position.

Unraveling this issue is outside the scopes of this work. Valuable articles are available from the literature, which report analyses and tools addressing the not trivial issue of assessing and comparing the quality of NMR structures, including (Doreleijers et al.,

2012; Laskowski et al., 1996; Montelione et al., 2013; Nicastro et al., 2006; Rosato et al., 2013; Vuister et al., 2014). We just notice here that we could not discern a significant correlation (Pearson's coefficients within 0.23 and *P*-values above 0.057, Fig. S1) between the interface variability and the NMR structure release year, which has been proposed as an indirect measure of the structure quality, in view of the substantial improvements in NMR spectroscopic structure determination over the time (Andrec et al., 2007).

However, as a matter of fact, a researcher interested in using the NMR structure of a protein complex should consider that there may be a certain variability in the interface of the different released conformers. That means that analyzing only the 1st NMR model, which is indeed the default choice for many tools, including molecular visualizers (Basse et al., 2013; DeLano, 2002; Tina et al., 2007) may lead to miss crucial information on the complex interface. We show an example of this in Fig. 3, for the NMR structure of the complex between the complementary interaction domains of mouse CBP (cAMP responsive CRE-binding protein) and human ACTR (activator for thyroid hormone and retinoid receptors), two incompletely structured domains that fold synergistically to form a tight complex (PDB ID: 1kbh (Demarest et al., 2002)). Interaction between these two proteins mediates the hormone response, which in turn regulates the expression of genes that are essential for development, reproduction and homeostasis. As shown in Fig. 3a, the 1st NMR model here, does not correspond to the model top ranked by CONSRANK. *C*₅₀, *C*₇₀ and *C*₉₀ values for this complex (0.981, 0.840 and 0.707, respectively) are above the average values we calculated for the whole set of structures. That notwithstanding, two intermolecular contacts discussed as crucial for the interaction (Demarest et al., 2002), namely H-bonds between the side chains of CBP-Asn83 and ACTR-Glu6 and of CBP-Ser68 and ACTR-Asp21, are not present in the 1st NMR conformer, where in fact the side chains of ACTR-Glu6 and Asp21 do not even point towards the CBP protein (see Fig. 3b). The above contacts are in fact observed in only 8 (40%) and 17 (85%) conformers, respectively, as can be seen from the interactive 3D map shown in Fig. 3c. Not surprisingly, both these H-bonds are instead observed in the conformer ranked 1st by CONSRANK (Fig. 3b), which can thus satisfactorily be used to describe the complex interface. The complete list of inter-molecular H-bonds, as obtained by accessing the COCOMAPS link available from the CONSRANK output, is reported in Fig. 3d for this model, including the above two H-bonds.

When looking at how are the models within the analyzed NMR ensembles ranked by CONSRANK, we indeed found that only for a minority of them (less than 8%) the 1st NMR conformer is also the one selected by CONSRANK as the best representative of the ensemble interface (Fig. 4). In Fig. 4, we indeed report the CONSRANK ranking position of the 1st NMR model for the 70 analyzed structures, while in the last column of Table 1 we listed the NMR model, for each structure, which is top ranked by CONSRANK. From these data, it is clear that the 1st NMR model is roughly equally ranked by CONSRANK between positions 1–20 and that the best representative of the overall interface can correspond to any of the 20 submitted NMR conformers.

4. Conclusions

We have shown here that the interface in NMR structures of protein heterodimeric complexes can be from poorly to very well conserved, within the set of conformers selected for submission to the PDB, and this different behavior is not simply the result of the different systems and interfaces size. Considering that the main hallmark of the NMR structures is the possibility to capture the dynamics of the system, we believe that this information should not be missed in the analysis phase. Furthermore, the 1st NMR

conformer is often not the best representative of the overall interface.

For these reasons, we propose here an approach, based on the conservation of inter-residue contacts, which allows analyzing the conformers ensemble altogether, by providing both: (i) numerical values of specific contacts and overall interface conservation, and (ii) the selection of one conformer whose interface best reflects that of the overall structure ensemble. This approach to the analysis of NMR structures of protein complexes can be of utility both to NMR spectroscopists, while reporting a novel structure, and to structural biologists who want to get insight into PDB structures for several applications, such as the design of site-specific mutations, the design of inhibitors of the interaction, or the selection of a representative structure for a template-based docking procedure.

Remarkably, all described analyses can be performed automatically in few seconds to minutes, by accessing the “NMR functionalities” now added to the CONSRANK web server at <https://www.molnuc.unisa.it/BioTools/consrank/consrank-nmr.html>.

The server is also suited to deal with nucleic acids, therefore, the proposed approach can straightforwardly be extended to the analysis of protein-nucleic acid NMR complexes.

Funding

RO has been supported by Regione Campania (LR5-AF2008).

Acknowledgment

We thank Dr. Andrea Petta for assistance in implementing the CONSRANK NMR functionalities.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2016.03.008>.

References

- Abdel-Azeim, S., Chermak, E., Vangone, A., Oliva, R., Cavallo, L., 2014a. MDcons: intermolecular contact maps as a tool to analyze the interface of protein complexes from molecular dynamics trajectories. *BMC Bioinformatics* 15 (Suppl. 5), S1.
- Abdel-Azeim, S., Oliva, R., Chermak, E., De Cristofaro, R., Cavallo, L., 2014b. Molecular dynamics characterization of five pathogenic Factor X mutants associated with decreased catalytic activity. *Biochemistry* 53, 6992–7001.
- Andrec, M., Snyder, D.A., Zhou, Z., Young, J., Montelione, G.T., et al., 2007. A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing. *Proteins* 69, 449–465.
- Basse, M.J., Betzi, S., Bourgeois, R., Bouzidi, S., Chetrit, B., et al., 2013. 2P2ldb: a structural database dedicated to orthosteric modulation of protein–protein interactions. *Nucleic Acids Res.* 41, D824–D827.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., et al., 2000. The protein data bank. *Nucleic Acids Res.* 28, 235–242.
- Chermak, E., Petta, A., Serra, L., Vangone, A., Scarano, V., et al., 2015. CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts. *Bioinformatics* 31, 1481–1483.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., et al., 2010. The genetic landscape of a cell. *Science* 327, 425–431.
- DeLano, W.L., 2002. The PyMOL Molecular Graphics System, <<http://www.pymol.org>>.
- Demarest, S.J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., et al., 2002. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415, 549–553.
- Di Domenico, T., Walsh, I., Martin, A.J., Tosatto, S.C., 2012. MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* 28, 2080–2081.
- Doreleijers, J.F., Vranken, W.F., Schulte, C., Markley, J.L., Ulrich, E.L., et al., 2012. NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acids Res.* 40, D519–D524.
- Foster, M.P., McElroy, C.A., Amaro, C.D., 2007. Solution NMR of large molecules and assemblies. *Biochemistry* 46, 331–340.

- Gonzalez-Ruiz, D., Gohlke, H., 2006. Targeting protein–protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* 13, 2607–2625.
- Hubbard, S.J., Thornton, J.M., 1993. 'NACCESS' Computer Program.
- Kelley, L.A., Sutcliffe, M.J., 1997. OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And Domains. *Protein Sci.* 6, 2628–2630.
- Konagurthu, A.S., Reboul, C.F., Schmidberger, J.W., Irving, J.A., Lesk, A.M., et al., 2010. MUSTANG-MR structural sieving server: applications in protein structural analysis and crystallography. *PLoS ONE* 5, e10048.
- Lancellotti, S., Peyvandi, F., Pagliari, M.T., Cairo, A., Abdel-Azeim, S., et al., 2015. The D173G mutation in ADAMTS-13 causes a severe form of congenital thrombotic thrombocytopenic purpura. A clinical, biochemical and *in silico* study. *Thromb. Haemost.* 115, 51–62.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.
- Laughton, C.A., Orozco, M., Vranken, W., 2009. COCO: a simple tool to enrich the representation of conformational variability in NMR structures. *Proteins* 75, 206–216.
- Lensink, M.F., Wodak, S.J., 2013. Docking, scoring, and affinity prediction in CAPRI. *Proteins* 81, 2082–2095.
- Martin, A.J., Walsh, I., Tosatto, S.C., 2010. MOBI: a web server to define and visualize structural mobility in NMR protein ensembles. *Bioinformatics* 26, 2916–2917.
- Montelione, G.T., Nilges, M., Bax, A., Guntert, P., Herrmann, T., et al., 2013. Recommendations of the wwPDB NMR validation task force. *Structure* 21, 1563–1570.
- Moore, D.S., McCabe, G.P., 1998. Introduction to the Practice of Statistics, third ed. W.H. Freeman & Co Ltd, New York.
- Mullard, A., 2012. Protein–protein interaction inhibitors get into the groove. *Nat. Rev. Drug Discov.* 11, 173–175.
- Nicastro, G., Habeck, M., Masino, L., Svergun, D.I., Pastore, A., 2006. Structure validation of the Josephin domain of ataxin-3: conclusive evidence for an open conformation. *J. Biomol. NMR* 36, 267–277.
- Nisius, B., Sha, F., Gohlke, H., 2012. Structure-based computational analysis of protein binding sites for function and druggability prediction. *J. Biotechnol.* 159, 123–134.
- Oliva, R., Chermak, E., Cavallo, L., 2015. Analysis and ranking of protein–protein docking models using inter-residue contacts and inter-molecular contact maps. *Molecules* 20, 12045–12060.
- Oliva, R., Vangone, A., Cavallo, L., 2013. Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins* 81, 1571–1584.
- Pawson, T., Nash, P., 2000. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* 14, 1027–1047.
- Potenza, E., Di Domenico, T., Walsh, I., Tosatto, S.C., 2015. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* 43, D315–D320.
- Rosato, A., Tejero, R., Montelione, G.T., 2013. Quality assessment of protein NMR structures. *Curr. Opin. Struct. Biol.* 23, 715–724.
- Ryan, D.P., Matthews, J.M., 2005. Protein–protein interactions in human disease. *Curr. Opin. Struct. Biol.* 15, 441–446.
- Scott, W.R., Straus, S.K., 2015. Determining and visualizing flexibility in protein structures. *Proteins* 83, 820–826.
- Sikic, K., Carugo, O., 2009. CARON – average RMSD of NMR structure ensembles. *Bioinformatics* 4, 132–133.
- Snyder, D.A., Montelione, G.T., 2005. Clustering algorithms for identifying core atom sets and for assessing the precision of protein structure ensembles. *Proteins* 59, 673–686.
- Theobald, D.L., Wuttke, D.S., 2006. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* 22, 2171–2172.
- Tina, K.G., Bhadra, R., Srinivasan, N., 2007. PIC: protein interactions calculator. *Nucleic Acids Res.* 35, W473–W476.
- Tosatto, S.C., Battistutta, R., 2007. TAP score: torsion angle propensity normalization applied to local protein structure evaluation. *BMC Bioinformatics* 8, 155.
- Vangone, A., Abdel-Azeim, S., Caputo, I., Sblattero, D., Di Niro, R., et al., 2014. Structural basis for the recognition in an idiotype–anti-idiotype antibody complex related to celiac disease. *PLoS ONE* 9, e102839.
- Vangone, A., Bonvin, A.M., 2015. Contacts-based prediction of binding affinity in protein–protein complexes. *eLife* 4, e07454.
- Vangone, A., Cavallo, L., Oliva, R., 2013. Using a consensus approach based on the conservation of inter-residue contacts to rank CAPRI models. *Proteins*.
- Vangone, A., Oliva, R., Cavallo, L., 2012. CONS-COCOMAPS: a novel tool to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. *BMC Bioinformatics* 13 (Suppl. 4), S19.
- Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., Oliva, R., 2011. COCOMAPS: a web application to analyse and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* 27, 2915–2916.
- Velankar, S., Best, C., Beuth, B., Boutselakis, C.H., Cobley, N., et al., 2010. PDBe: protein data bank in Europe. *Nucleic Acids Res.* 38, D308–D317.
- Velankar, S., van Ginkel, G., Alhroub, Y., Battle, G.M., Berrisford, J.M., et al., 2015. PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.*
- Vidal, M., Cusick, M.E., Barabasi, A.L., 2011. Interactome networks and human disease. *Cell* 144, 986–998.
- Vuister, G.W., Fogh, R.H., Hendrickx, P.M., Doreleijers, J.F., Gutmanas, A., 2014. An overview of tools for the validation of protein NMR structures. *J. Biomol. NMR* 58, 259–285.