

# A graph theoretical approach for assessing bio-macromolecular complex structural stability

Carlos Adriel Del Carpio · Mihai Iulian Florea · Ai Suzuki · Hideyuki Tsuboi · Nozomu Hatakeyama · Akira Endou · Hiromitsu Takaba · Eiichiro Ichiishi · Akira Miyamoto

Received: 7 November 2008 / Accepted: 25 February 2009 / Published online: 29 April 2009  
© Springer-Verlag 2009

**Abstract** Fast and proper assessment of bio macromolecular complex structural rigidity as a measure of structural stability can be useful in systematic studies to predict molecular function, and can also enable the design of rapid scoring functions to rank automatically generated bio-molecular complexes. Based on the graph theoretical approach of Jacobs et al. [Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF (2001) Protein flexibility predictions using graph theory. *Proteins: Struct Funct Genet* 44:150–165] for

expressing molecular flexibility, we propose a new scheme to analyze the structural stability of bio-molecular complexes. This analysis is performed in terms of the identification in interacting subunits of clusters of flappy amino acids (those constituting regions of potential internal motion) that undergo an increase in rigidity at complex formation. Gains in structural rigidity of the interacting subunits upon bio-molecular complex formation can be evaluated by expansion of the network of intra-molecular inter-atomic interactions to include inter-molecular inter-atomic interaction terms. We propose two indices for quantifying this change: one local, which can express localized (at the amino acid level) structural rigidity, the other global to express overall structural stability for the complex. The new system is validated with a series of protein complex structures reported in the protein data bank. Finally, the indices are used as scoring coefficients to rank automatically generated protein complex decoys.

C. A. Del Carpio (✉) · H. Tsuboi · N. Hatakeyama · A. Endou · H. Takaba · A. Miyamoto  
Department of Applied Chemistry,  
Graduate School of Engineering, Tohoku University,  
6-6-11-1302 Aoba, Aramaki,  
Aoba-ku, Sendai 980-8579, Japan  
e-mail: carlos@aki.che.tohoku.ac.jp

M. Iulian Florea  
Department of Electrical, Information and Physics Engineering,  
School of Engineering, Tohoku University,  
6-6-04, Aramaki Aza Aoba,  
Aoba-ku, Sendai 980-8579, Japan

A. Suzuki · A. Miyamoto  
New Industry Creation Hatchery Center (NICHe),  
Tohoku University,  
6-6-10 Aoba, Aramaki,  
Aoba-ku, Sendai 980-8579, Japan

E. Ichiishi  
School of Materials Science,  
Japan Advanced Institute of Science and Technology,  
Nomi-city, Ishikawa 923-1211, Japan

E. Ichiishi  
Graduate School of Medicine, Tohoku University,  
Aoba-ku, Sendai 980-0872, Japan

**Keywords** Protein–protein interaction · Complex structural rigidity · Biomolecular docking · Scoring function

## Introduction

Fueled by advances in information technologies and hardware sophistication, research projects entailing large-scale genome-wide analysis have experienced exponential growth in the post-genome era. At the center of these initiatives has been the outstanding work directed at elucidating the structure and function of bio-molecules in general and proteins in particular, the problem of protein–protein interactions (PPIs) and complex formation being ubiquitous to the entire spectrum of intracellular biochemical processes in a living organism.

Difficulties inherent to the problem, however, have prevented major breakthroughs in the field, and unrelenting effort is being devoted to overcoming these, as epitomized by recent international contests (e.g., CAPRI [2]) designed to assess the problem-solving ability of automatic methodologies developed for this purpose. A wide range of methodologies that attempt to solve the problem can be found in the literature [3–7], the most frequently referenced being the grid-scoring algorithm developed by Katchalski-Katzir et al. [8] used to dock pairs of protein molecules regarded as rigid bodies. Since the main difficulty stems precisely from the fact that proteins are intrinsically flexible molecules, several improvements to this seminal technique have been implemented in order to rank higher close-to-native decoys and decrease the number of false positives generated by the original algorithm.

One such attempt is constituted by our recently developed system “MIAX” (bio Macromolecular Interaction Assessment Computer System), which endows receptor and ligand molecules with a flexibility character, which, embedded within the grid-scoring algorithm in what we call a soft docking algorithm, performs much better than rigid body docking alone, especially in cases where independently crystallized structures known to interact are docked [9] (the unbound docking problem).

Notwithstanding this limited flexibility accounting term and the improved performance of the algorithm—particularly in recognizing the interaction sites on both molecules—the central problem stemming from collective motions within each protein molecule has to be focused and addressed from a completely different angle.

Common sense when manipulating this type of intramolecular motion within a protein naturally points at molecular dynamics (MD) simulations as the most suitable technique available with which to determine this type of internal motion in macromolecules. However, the collective internal motions, that are critical to protein function are characteristically low frequency motions corresponding to internal substructures such as loops and domains, the simulation of which using MD methods require intrinsically large simulation times and computational infrastructure. The limitation of this type of approach is underscored by attempts to process biologically transcendental molecular systems such as protein–protein, protein–RNA, or protein–DNA complexes, the sheer size of these molecules undermining the completeness of, and our ability to interpolate, the results.

We have attempted to overcome this problem by developing an algorithm based on graph theoretical instances oriented to determine the flexibility of a protein molecule in terms of the number of degrees of freedom it possesses. Based on the latter, we can then identify and compute the number of internal collective motions that

proteins may express that exert non ambiguous effects in their interaction behavior with other molecules. Moreover, the algorithm allows the analytical determination of the rigidity achieved by complex formation as resulting in the loss of degrees of freedom of the subunits comprising it. The algorithm is generated by treating protein molecules as networks of inter- and intra-molecular atomic interactions dictated by the three-dimensional structure of the macromolecule (complex and/or constituting independent subunits) and its inter-atomic connectivity. Quantifying the degree of loss of flexibility of the interacting subunits and the gain in rigidity of the complex may assist not only in the determination and assignment of specific functions to well defined regions and domains in the interacting protein (RNA, or DNA) molecules but also in the development of scoring schemes to assess the correctness of protein complex conformations predicted by automatic procedures. Here, we introduce two indices to quantify the loss of flexibility: one at the local or amino acid level, and a second at a global level, which can be used as an index to express the structural stability achieved by the interacting subunits in the bio-molecular complex.

Furthermore, combination of this methodology with MD simulations can result in faster engines to predict the real time behavior of protein complexes, and thus it may be of great utility to large-scale genome-wide function elucidation studies.

## Methods

While a protein fold may be stabilized in its native state by mainly covalent bonding (such as disulfide bridges) and intra-molecular electrostatic interactions, protein complexes on the other hand are formed essentially by weak interactions, namely van der Waals, ionic, and hydrophobic forces, that play a determinant role in the way biomolecules interact as well as in the stabilization of the protein complex. Computing interacting energies gives therefore a general assessment of the stability of a molecular system (i.e., a bio-molecule, a complex of two or more monomers, or an aggregation of molecules in general). Analysis of structural changes in terms of these energies, however, may not directly give a clear account of the contribution to this loss of stability of rearrangements and internal collective motions of sub-structures unless one performs a meticulous analysis correlating them with structural change, let alone the accuracy level of the force field used in such studies.

Given the large number of bio-processes in which biomolecules in general, and proteins in particular, take part in within an organism, the need to express molecular stability in a rather systematic and direct way is critical for genome-

wide functional analyses. A method to express the stability of a bio-molecule in terms of the putative conformational changes undergone by that bio-molecule can be realized by calculating the changes in the number of degrees of freedom (DOF) associated with any particular chemical (or physical) transformation.

Although enumerating the number of DOFs in a molecule involves computationally expensive procedures, a methodology based on network rigidity and constraint counting proposed by Jacobs et al [10–12], has proved a worst case performance of  $O(n^2)$ , for a network involving  $n$  sites. We have developed an algorithm for biological molecules to account for flexibility loss and rigidity gain upon bio-molecular complex formation that retains some of the basic tenets of Jacob's algorithm but that has broader applicability to cases where weak forces such as those involved not only in protein complexes but in the formation of any type of bio-molecular complex, as mentioned above, play transcendental roles. The graph theoretical instances on which the algorithm is based are described in the next section.

#### Protein structural rigidity estimation based on constraint counting

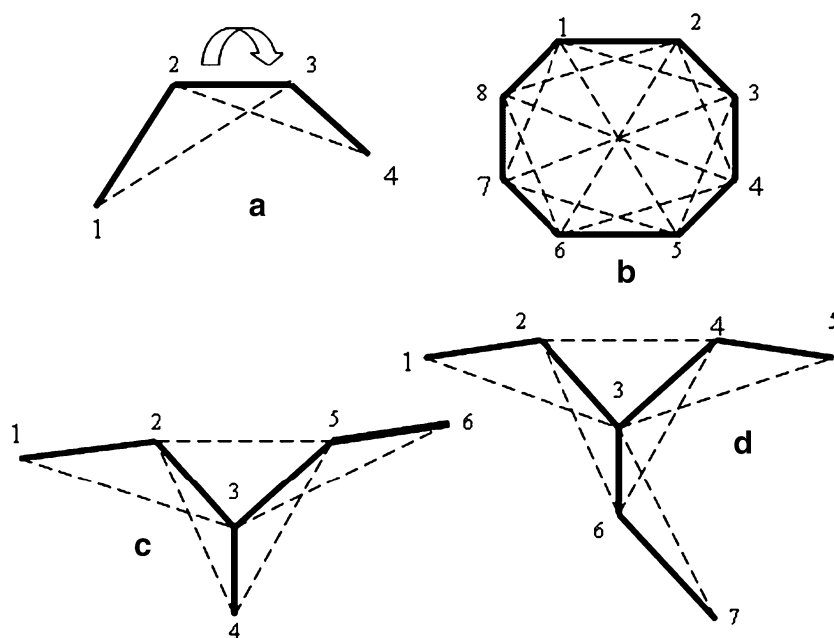
The notion behind Jacob's algorithm expressed in graph theoretical terms can be illustrated by the graphs (networks of nodes) shown in Fig. 1, where a hinge joint (Fig. 1a) is formed by edge joining two triangles, Fig. 1b shows a rigid cluster, and Fig. 1c and d show three and four rigid clusters, respectively.

A hinge joint or torsional angle between nodes 1–4 in a chain of atoms connected in the order 1–2–3–4 (i.e., 1 and 4 not directly connected but rather via central atoms 2 and 3; Fig. 1a) implies intrinsically a rotation motion around the edge (which may or may not be realizable, depending on the interaction energy between atoms 1 and 4).

According to this logic, the graph in Fig. 1a can be described as composed of two rigid clusters or graphs (triangles with vertices of nodes 1,2,3 and 2,3,4) joined by an independent hinge constituted by the edge joining vertices 2 and 3.

The same notion applied to the description of the graph in Fig. 1b shows no independent hinges since rotation of any of the nodes with respect to any of the joints is precluded by connections of that particular node to other different nodes, the graph thus constituting a single rigid cluster.

The graph in Fig. 1c is constituted of three rigid clusters, i.e., triangle (1,2,3), tetrahedron (2,3,4,5) and triangle (3,5,6), and independent hinges on edges (2,3) and (3,5); while the graph in Fig. 1d is constituted by four rigid clusters, the triangles (1,2,3), (3,4,5), and (3,6,7) plus a tetrahedron with vertices (2,3,4,6), independent hinges being constituted by edges (2,3), (3,4) and (3,6); a further element in the last two graphs is the pivot node (node 3) into which the rigid clusters converge. The rigidity of a network built in this way will be proportional to the number of edges (constraints) from which it is constituted since they constrain possible movements of the vertices; consequently, the number of DOF of such networks will decrease with the increase in the number of edges linking the nodes.



**Fig. 1** Distance constraint algorithm for molecular rigidity estimation. **a** Hinge joint, **b** rigid cluster, **c** three rigid clusters, **d** four rigid clusters

The number of DOFs in a network can be estimated by constraint counting (number of independent edges or connections among the nodes) [10–12].

If all the constraints are independent, the number of DOF,  $F$ , can be expressed as:

$$F \approx 3N - C \quad (1)$$

where  $N$  is the number of nodes and  $C$  the number of constraints. Since all the constraints are not independent, the real number of DOFs in a network of  $N$  nodes in a  $d$ -dimensional space is lower than in Eq. 1, the number of constraints to make all the nodes of a set of  $n \geq d$  nodes mutually rigid being no more than:

$$F = dn - \frac{d(d+1)}{2} \quad (2)$$

Each rigid region in a network can thus be identified whenever there are exactly enough constraints distributed among a set of nodes to make the nodes composing the network mutually rigid, or, put more simply, by identifying redundant constraints in the network. For a planar network, this logic is stated by the Laman theorem [12].

Identification of the redundant constraints for a network constituted by  $N$  nodes is frequently performed by the use of a pebble game algorithm constructed for 2 or 3 dimensions [10, 13, 14].

Application of the algorithm to identify constrained (rigid) regions in an isolated bio-molecule or a protein complex can be performed by constructing a similar network from the connectivity (covalent bonding) relationship among the constituting atoms and the interaction forces that hold them at their equilibrium positions.

Since our interest is namely the identification of changes in the flexibility of protein monomers upon interaction and/or complex formation, besides constraints derived from covalent bonding, angle stretching, and N–H···O type hydrogen bonding (typical or conventional hydrogen bonds), the need arises for consideration of constraints emerging from other weak inter-molecular interactions such as hydrogen bonds of the C–H···O type, ionic bonds, and hydrophobic interactions prevailing in PPIs. Here, we show that the contribution of C–H···O type hydrogen bonds, introduced as a new term together with the hydrophobic interaction among the atoms constituting the interaction interfaces, is not negligible and adds to the consistency and rigidity of the network of interactions in complex structures.

Bio-molecular complex structure as a network of constraints

Building the network of constraints for a protein (or any biomolecule such as RNA or DNA), as mentioned earlier, requires identification of all the forces holding the atoms at

their equilibrium positions. The first type of constraints are constituted by central-force constraints, which in this case are covalent bonds (1–2 type interactions or nearest neighbor interactions, as shown by solid lines between atoms 1 and 2, 2 and 3, or 3 and 4 in Fig. 1a, for example), stretch angle interactions (1–3 type interactions; interaction between atoms 1 and 3, or 2 and 4 in Fig. 1a) and N–H···O hydrogen bonding type interactions (interaction between atoms 1 and 4 in Fig. 1a). In the case of a hydrogen bond, a constraint will be added to the donor and the carbon atom connected to the acceptor, as well as to the H and O atoms involved in the hydrogen bond (vide infra). In the algorithm presented here, threshold energy values for the stretching bond and conventional hydrogen bond interactions are set so that only interactions above these threshold values are considered in the set of constraints or edges in the network. For the angle stretching energies calculated using the MM3 force field, the threshold value is set within a range of 0.02 kcal mol<sup>-1</sup>. For conventional hydrogen bonding, the threshold is set to -1.0 kcal mol<sup>-1</sup>, which is equivalent to a distance between the hydrogen atom and the donor of 3.7 Å.

Since the main objective of the present work is the evaluation of protein complex structures, weak forces that keep the interacting molecules bound have to be taken into account as force constraints. As mentioned above, two such constraints have been considered here and details of their computations are given in the following sections.

#### C–H···O hydrogen bonds

Hydrogen bonds in general are key elements in the interplay between stability and specificity in protein folding as well as in PPI. Moreover, upon complex formation, the strength and directionality of hydrogen bonds make them one of several important factors discriminating the orientation of the interacting subunits. In fact, recent studies by Jiang and Lai [15], and Panigrahi and Desiraju [16] on the role of CH···O type hydrogen bonds in protein folding and PPIs have shown that the average energy contribution of a conventional H-bond is 30% while that of a CH–O is as much as 17%, by no means an insignificant number. In the specific case of a C<sub>α</sub>, this atom is an activated carbon donor because it is bound to the electron-withdrawing amide N–H and C=O groups, and hydrogen bonds between main-chain C<sub>α</sub>–H groups and backbone or side-chain oxygen atoms are frequently observed in soluble interacting proteins. However, side chain carbon atoms are also involved in C–H···O hydrogen bonds with side chain O atoms—Gly, Phe, and Tyr being noteworthy for the number of this type of hydrogen bonds they form. Furthermore, although this type of interaction had been regarded as a negligible, low energy interaction, ab-initio calculations

have demonstrated that these energies may reach a level of 2.5 to 3.0 kcal mol<sup>-1</sup> in vacuum [17, 18]. Consequently, the relevance of these interactions cannot be neglected, and in the present work they are computed using a conventional hydrogen bond potential function where threshold values for the hydrogen atoms and the acceptor are set to  $d_{\text{H}} < 3.5$  Å. and  $\zeta > 120$  degrees (or  $\zeta > 90$  degrees when  $d_{\text{H}} < 3.0$  Å) as shown in Fig. 2 [15].

Atoms involved in a C–H···O type hydrogen bond are assigned constraints in a similar way as for atoms involved in conventional hydrogen bonds, as described earlier.

The relevance of C–H···O type hydrogen bonds in the present algorithm is shown by comparing the effects of considering and disregarding them on the algorithm results. This computational experiment is described in the **Results** section (vide infra).

Since X-ray solved protein structures (as well as RNA and DNA structures) reported in the protein data bank (PDB) lack hydrogen atoms coordinates, these are computed using the REDUCE program [19].

#### *Force constraints derived from the hydrophobic interaction term*

A further constraint besides C–H···O hydrogen bonds holding atoms at their equilibrium positions in protein complexes with respect to each other is the hydrophobic interaction; hydrophobic groups on the surfaces of the protein molecules tend to aggregate in an aqueous environment. Accounting for this factor in the context of the network model presented here is accomplished by computing van der Waals interactions among atoms for which the hydrophobic potential is high. Hydrophobic potentials for the atoms of the interacting proteins are calculated using the scheme of Bresseur [20]. Here again, a threshold value equivalent to 2.0 kcal mol<sup>-1</sup> is established for this type of interaction, with van der Waals interaction energies calculated using the MM3 force field. Therefore, all van der Waals interactions among atoms of high hydrophobic potentials with energies below the threshold are considered non-central force constraints, in contrast to covalent bonds and angle stretching constraints.

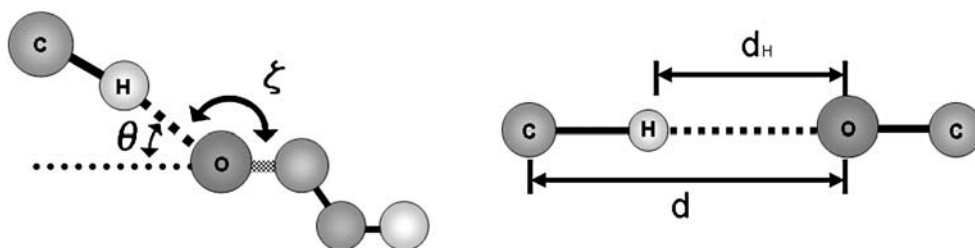
#### Stabilizing water molecules

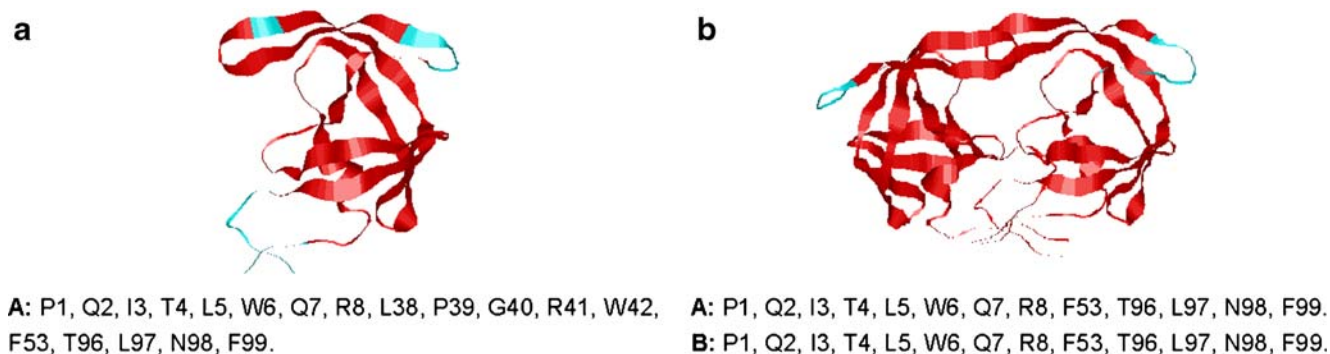
A further factor leading to stability (or instability of the complex) is the breaking (or in several cases the formation) of hydrogen bonds that specific amino acids in the interacting isolated subunits have with particular water molecules in their environment when interaction occurs. To take into account the contribution of the breaking or formation of these bonds to increased/decreased rigidity of the complex, the methodology proposed here takes into consideration the position of water molecules that are reported together with the crystal structures of proteins in the PDB. Although a more exhaustive analysis would place hypothetical water molecules in the proximity of polar groups, in this report we limit ourselves to the positions of water molecules reported in PDB. Since the present methodology is static in nature (displacement of atoms or molecules are ignored), whenever the position of a water molecule is superimposed by the atomic positions of any of the monomers, water molecules drop out from the calculation, taking into account only the number of hydrogen bonds that the water molecule formed with any of the interface atoms of any of the interacting subunits. Inclusion of specific water molecules as part of the target system to analyze the effect of the hydrogen bonds does not change the underlying constraint counting methodology to identify regions of high and low rigidity in the protein monomers and complexes.

#### Quantitative expression of protein structural flexibility change upon complex formation

The procedure described so far can be applied to the evaluation of the role that flexibility of the protein subunits plays in the formation of a new complex molecule. Counting the DOF for each protein structure using the constraint counting methodology permits the differentiation of high flexible regions or domains in the protein from domains of high rigidity. Flexible domains are characterized by a high number of DOF, i.e., a high number of independent hinges [14], while highly rigid domains will be characterized by a low number, or no, floppy modes. An

**Fig. 2** Constants involved in computation of C–H···O hydrogen bond energies



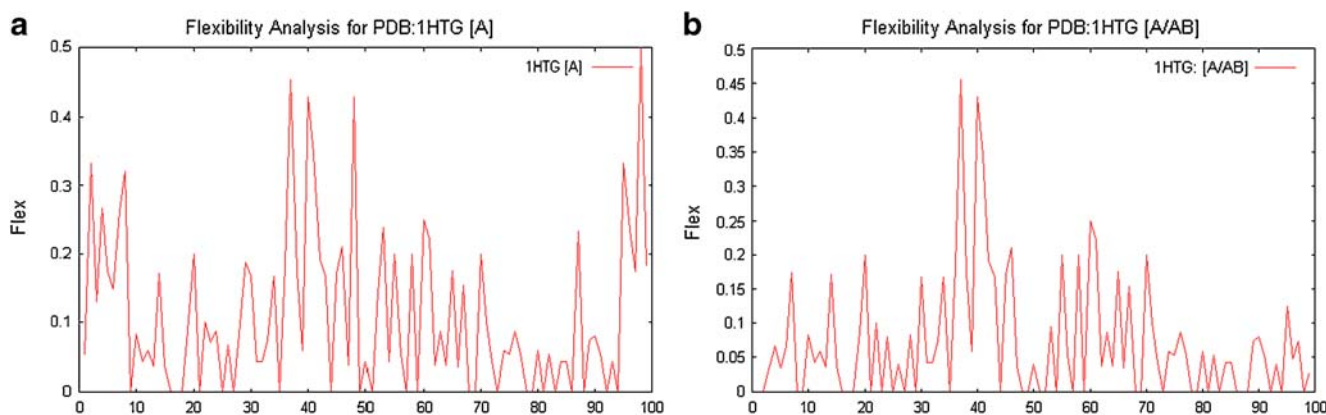


**Fig. 3** Flexible regions for the human immunodeficiency virus type 1 (HIV-1) protease monomer in the isolated (**a**; 1HHP) and complexed (**b**; 1HTG) state. Text: *A* Flappy amino acids in isolate monomer, *B*

Amino acids changing from flexible to rigid. Regions mapped in *light color* indicate the position of flexible regions in the backbone of the protein

**Table 1** List of C–H···O type pairs at the interaction interface in 1HTG [human immunodeficiency virus type 1 (HIV-1) protease monomer in the complexed state]

Donor					Acceptor				
Atom number	PDB name	Amino acid		Chain	Atom	PDB Name	Amino acid		Chain
		Name	No.				Name	No.	
3105	CG2	THR	96	B	24	OE1	GLN	2	A
2963	CZ	ARG	87	B	89	O	TRP	6	A
2961	CD	ARG	87	B	113	O	GLN	7	A
2000	CA	GLY	27	B	416	OD2	ASP	25	A
1720	CG	PRO	9	B	424	O	THR	26	A
1721	CD	PRO	9	B	424	O	THR	26	A
1942	CD2	LEU	23	B	438	O	GLY	27	A
1699	CZ	ARG	8	B	459	OD2	ASP	29	A
2349	CG2	ILE	50	B	768	O	GLY	48	A
2363	CA	GLY	51	B	801	O	GLY	51	A
3150	CB	PHE	99	B	1463	O	ILE	93	A
3147	CA	PHE	99	B	1482	O	GLY	94	A
3150	CB	PHE	99	B	1482	O	GLY	94	A
1636	CG	LEU	5	B	1489	O	CYS	95	A
1637	CD1	LEU	5	B	1489	O	CYS	95	A
1571	CD	PRO	1	B	1547	O	PHE	99	A
1361	CZ	ARG	87	A	1653	O	TRP	6	B
1359	CD	ARG	87	A	1677	O	GLN	7	B
436	CA	GLY	27	A	1980	OD2	ASP	25	B
156	CG	PRO	9	A	1988	O	THR	26	B
157	CD	PRO	9	A	1988	O	THR	26	B
378	CD2	LEU	23	A	2002	O	GLY	27	B
135	CZ	ARG	8	A	2023	OD2	ASP	29	B
838	CG2	ILE	54	A	2346	O	ILE	50	B
839	CD1	ILE	54	A	2346	O	ILE	50	B
1548	CB	PHE	99	A	3065	O	ILE	93	B
1545	CA	PHE	99	A	3084	O	GLY	94	B
1548	CB	PHE	99	A	3084	O	GLY	94	B
72	CG	LEU	5	A	3091	O	CYS	95	B
73	CD1	LEU	5	A	3091	O	CYS	95	B
7	CD	PRO	1	A	3149	O	PHE	99	B



**Fig. 4** Flex index for the monomer in the isolated (a) and complexed (b) states in 1HTG

index to quantify the rigidity and/or flexibility of each amino acid constituting the protein can be calculated by:

$$flex = \frac{\text{Number of independent hinges in the amino acid}}{\text{Number of constraints for the amino acid}} \quad (3)$$

Applying the pebble game algorithm [9–11] and the computation of the *flex* index using Eq. 3, allows the clustering of amino acids in highly rigid and highly flexible domains. Characterizing and quantifying the transformation of these clusters of amino acids in the transition undergone by isolated protein monomers in the process of complex formation can lead to the establishment of general factors affecting the orientation and conformation changes observed before and after interaction. Moreover, since highly flexible domains in the protein can characterize regions of characteristic collective motion, the influence that this type of dynamic exerts on the way proteins interact with one another can also be studied by the algorithm presented here.

Since the present article is about the rigidity gain of monomers upon interaction and complex formation, a further index,  $\Delta Rix$ , that expresses the real gain of rigidity undergone by the monomers composing the resulting complex structure has been proposed and calculated. This index is defined as the difference between the flex index for the structure in the isolated state and in the complexed state or, in other words, the change in flexibility with opposite sign. This can be expressed as in Eq. 4:

$$\Delta Rix_{A/AB}^{(i)} = -\left(flex_{A/AB}^{(i)} - flex_A^{(i)}\right) \quad (4)$$

Where  $\Delta Rix_A^{(i)}$  is the rigidity change index for the *i*th amino acid in the sequence of monomer A, and  $flex_{A/AB}^{(i)}, flex_A^{(i)}$  are the flexibility indices computed using Eq. 3 for monomer A in the complexed and isolated states, respectively.

While flex and  $\Delta Rix$  indices express local flexibility and rigidity change at the amino acid level, one can define a

global index to express the total change in the overall rigidity for the system. We propose a topological index,  $S_i$ , of stability for the complex with respect to the isolated monomers that can account for the total change in rigidity upon complex formation and can be computed as:

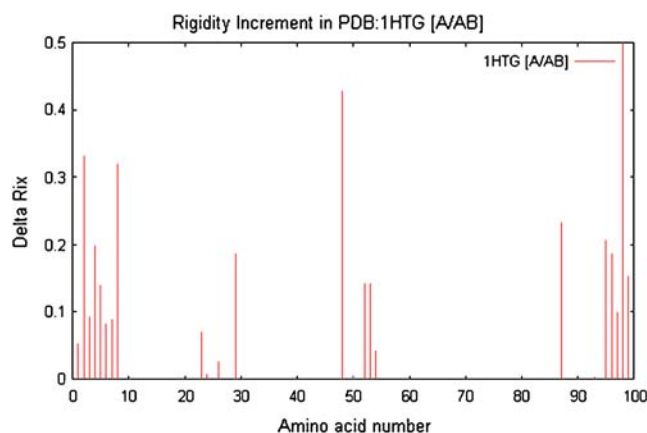
$$S_i = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^{n_j} \Delta Rix_{A_j}^{(i)} \quad (5)$$

Where  $S_i$  is the total stability index for the complex,  $N$  the number of monomers composing the complex,  $n_j$  the number of amino acids composing monomer  $A_j$  and  $\Delta Rix_{A_j}^{(i)}$  the rigidity change index for the *i*th amino acid in monomer  $A_j$  as defined in Eq. 4.

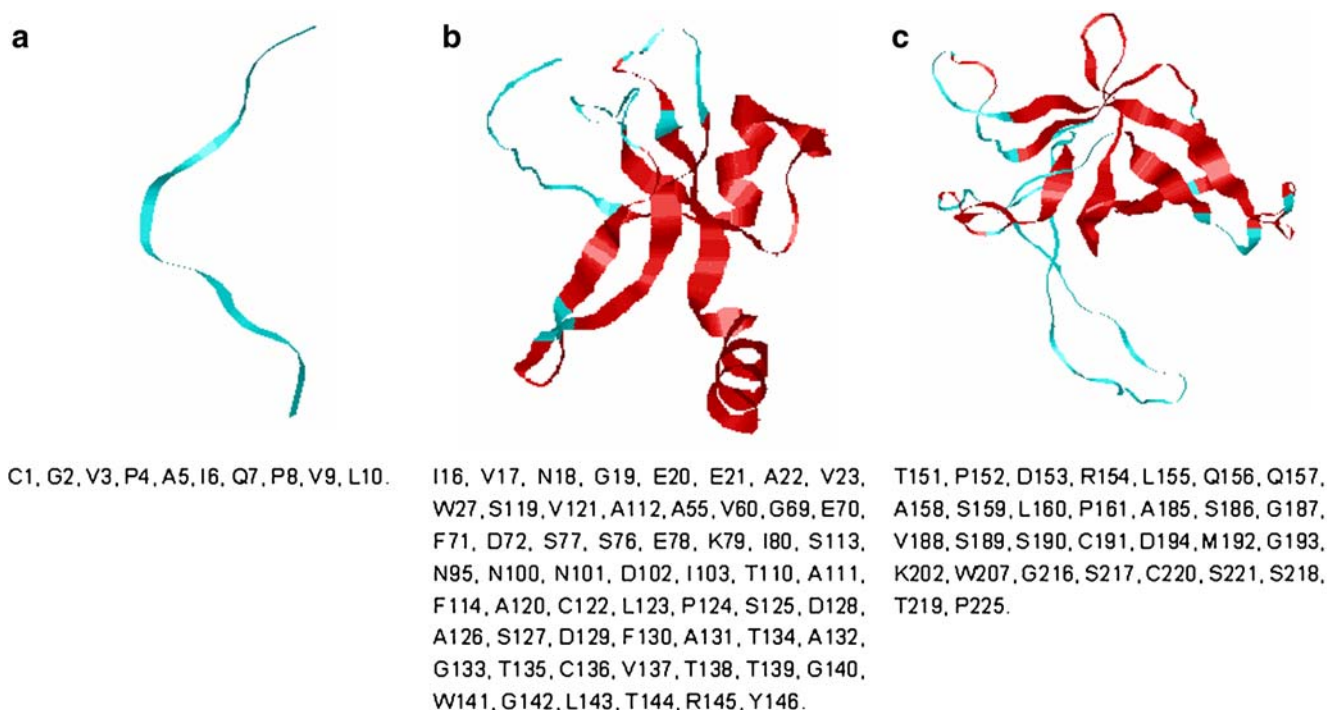
We apply the methodology to the evaluation of several protein complex prototypes and discuss the effect of flexibility and/or rigidity of the interacting molecules on the final complex structures.

## Results

The methodology has been coded in the C programming language, and the complete analysis system has been



**Fig. 5**  $\Delta Rix$  index for the monomer in 1HTG



**Fig. 6** Flexible regions for the monomers constituting the gamma-chymotrypsin L-para-chloro-1-aceoamido boronic acid inhibitor complex (IVGC) in chain A (**a**), chain B (**b**), and chain C (**c**) in the isolated state. The composition of each flappy region is shown beneath the ribbon models

dubbed “RIGIX” for RIGIdity analysis Computer System, with the acronym for computer system, “CS”, being replaced by its phonetic equivalent “X”.

RIGIX was validated using a set of randomly selected protein complex structures reported in PDB. The  $\Delta$ Rix

index was computed using Eq. 4, the necessary flex indices being computed for each monomer, in both the complex ( $\text{flex}_{A/\text{complex}}$ ) and isolated ( $\text{flex}_A$ ) states. Clusters of flappy amino acids composing flexible regions in molecular systems are mapped on ribbon models of the monomers

**Table 2** List of C–H $\cdots$ O type pairs at the interaction interfaces in gamma-chymotrypsin L-para-chloro-1-aceoamido boronic acid inhibitor complex (IVGC)

Donor		Acceptor							
Atom number	PDB name	Amino acid		Chain	Atom	PDB Name	Amino acid		Chain
		Name	No.				Name	No.	
283	CA	SER	26	B	64	O	ILE	6	A
286	CB	SER	26	B	64	O	ILE	6	A
250	CG1	VAL	23	B	83	O	GLN	7	A
286	CB	SER	26	B	83	O	GLN	7	A
251	CG2	VAL	23	B	114	O	VAL	9	A
128	CA	LEU	10	A	213	OE2	GLU	20	B
67	CG2	ILE	6	A	264	O	PRO	24	B
68	CD1	ILE	6	A	264	O	PRO	24	B
65	CB	ILE	6	A	278	O	GLY	25	B
68	CD1	ILE	6	A	278	O	GLY	25	B
41	CB	PRO	4	A	285	O	SER	26	B
42	CG	PRO	4	A	285	O	SER	26	B
101	CB	PRO	8	A	285	O	SER	26	B
103	CD	PRO	8	A	285	O	SER	26	B
55	CB	ALA	5	A	1657	O	GLN	116	B
5	CB	CYS	1	A	1725	O	VAL	121	B



in the complexed and isolated states to allow visual inspection.

Here, we present three case studies in which a dimer, trimer and tetramer are treated in detail in order to discuss and evaluate the qualitative appropriateness of the methodology to assess complex structural stability, and the effectiveness of the proposed indices,  $\Delta R_{ix}$  and  $S_i$ , to quantitatively express it. We then present a series of other complexes to which the methodology has been applied, and discuss the generality of the proposed indices.

Detailed evaluation of the methodology was carried out using the dimer structure of human immunodeficiency virus type 1 (HIV-1) protease (PDB: 1HTG), the trimer of a serine protease (PDB: 1VGC), and a tetramer structure corresponding to the Bilin binding protein (PDB: 1BBP).

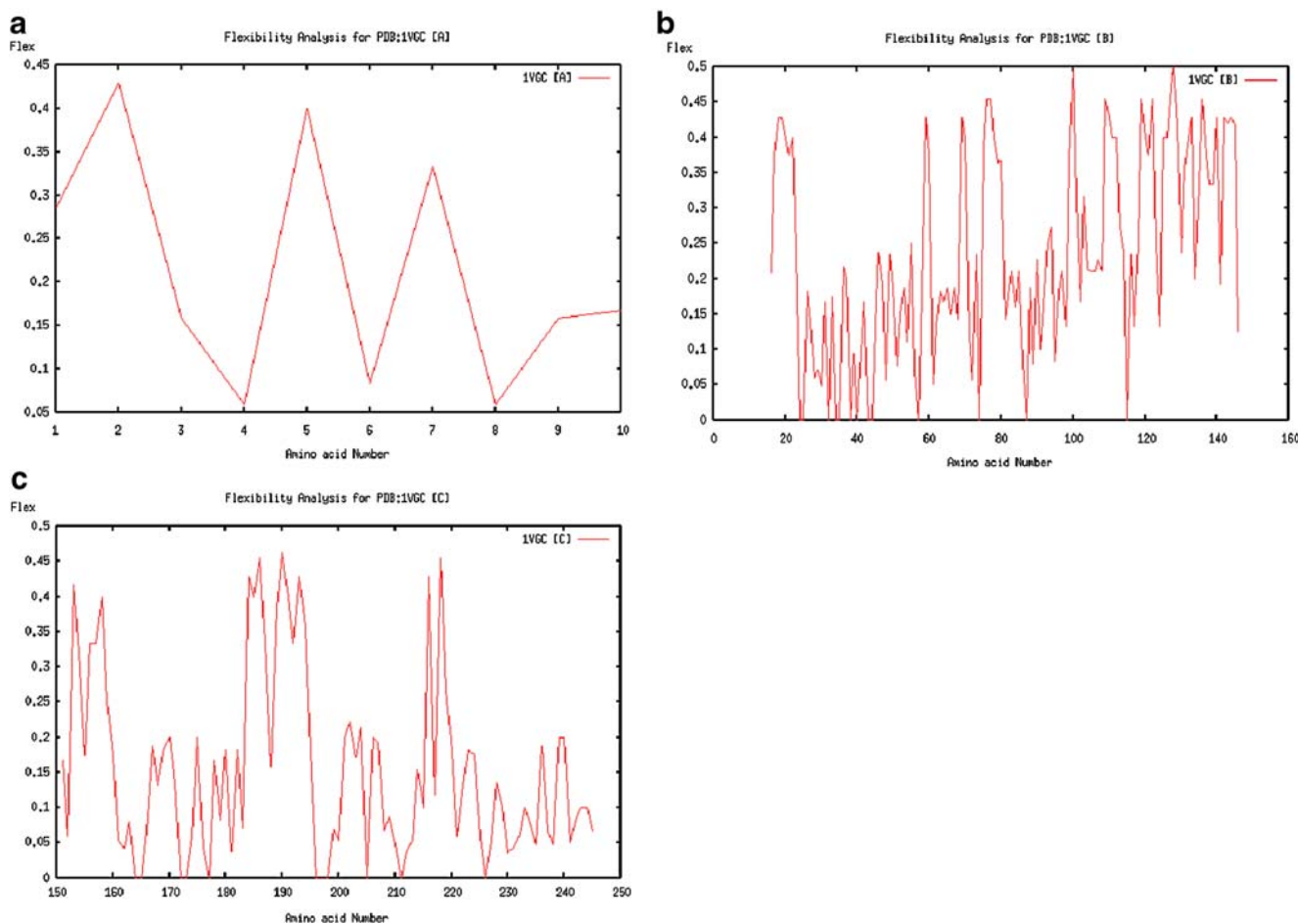
#### Case study 1: structural rigidity evaluation in HIV-1 protease

This structure is a prototype system for this type of analyses and we present it to validate our results, focusing on the

rigidity increment undergone by the subunits composing the complex when they interact to form the corresponding homo-dimer.

The 3D structures of the monomer and the complex can be found in PDB with codes 1HHP and 1HTG, respectively. The flappiest regions output by RIGIX, after addition of hydrogen atoms to the reported PDB structures, are mapped on ribbon models for the monomer in the isolated and complexed state as shown in Fig. 3a and b, respectively. All putative C–H $\cdots$ O hydrogen bonds in the complex are listed in Table 1. Similarly Fig. 4a and b illustrate plots of the flex index (Eq. 3) for the monomer in the isolated and complexed state, respectively.

The constitution of the flexible regions in terms of the names and numbers of the floppy amino acids of the monomer in the isolated state is listed below the corresponding ribbon model. The regions mapped in light color show the position of the flexible regions on the backbone of the protein molecules. Change in flexibility of the monomer in the isolated and complexed states is evident. Amino acids undergoing transformation from



**Fig. 7** Flex index plot for the isolated monomers of 1VGC. **a** Chain A, **b** chain B, **c** chain C

flexible to rigid in the complex are listed under the model of the complex structure. In the present case, while the structure of the isolated monomer is characterized by four regions of flexible amino acids, only one such cluster on each subunit remains flexible in the complexed structure.

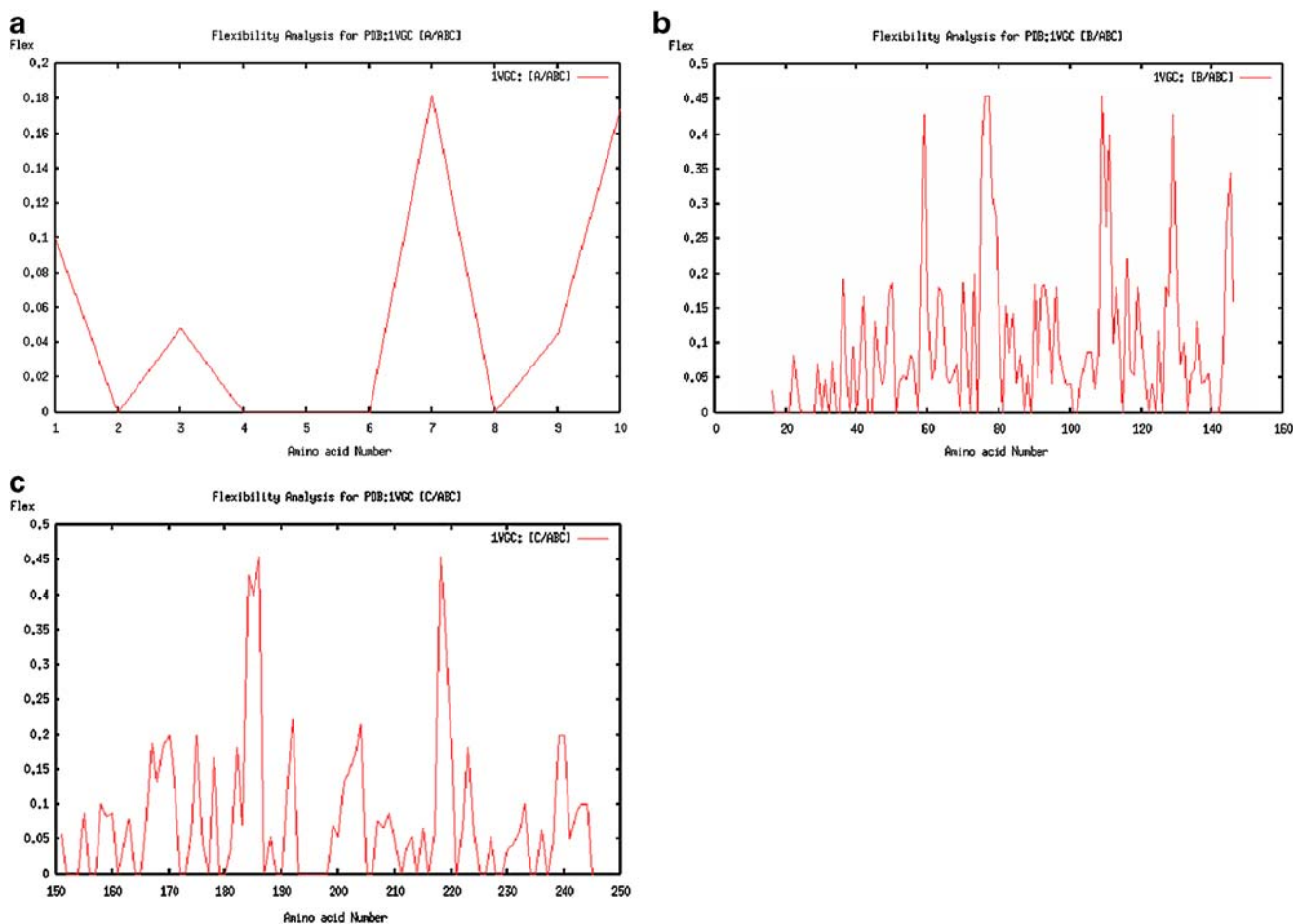
Increase of rigidity, evaluated by computing the  $\Delta R_{ix}$  index, is illustrated in Fig. 5.

Figure 5 shows the magnitude of rigidity gain for the amino acids of the interacting monomer.  $\Delta R_{ix}$  is zero when no change in flexibility, or rigidity, occurs. Therefore, the positions of the peaks in the  $\Delta R_{ix}$  plot represent amino acid positions that have gained rigidity, and the magnitude of this gain is represented by the height of the bar. Amino acids gaining rigidity belong mostly to the interaction interface of the monomers and their neighboring amino acids. Moreover, critical amino acids involved in the interaction between subunits become immediately apparent upon examining the magnitude of the index. Amino acids showing a stabilization greater than  $\Delta R_{ix}=0.2$  are GLN2, ARG8, GLY48 and ASN99. The gain in rigidity of these amino acids and their contribution to the overall structural

stability of the complex comply with the principles of structural biochemistry in that all these amino acids, and GLY in particular, are the most flexible. On the other hand, the *flex* index plot illustrates the regions in a particular subunit that remain flexible in the complex in spite of the interaction with the partner subunit. In the case of the HIV-1, these regions correspond to the amino acids listed under the ribbon model of the complex structure in Fig. 4b. The overall index of structural stability for the complex  $S_i$  is equal to 3.63.

Case study 2: structural rigidity analysis of serine protease complex PDB: 1VGC

Gamma-chymotrypsin L-para-chloro-1-aceoamido boronic acid inhibitor complex (PDB:1VGC) is a trimer structure, consisting of subunit A—a 13 amino acid long oligopeptide—and subunits B and C constituted of 131 and 91 amino acids, respectively. Applying RIGIX to each of the three subunits and the whole complex reveals the effects of the interaction on changes in structural flexibility of the monomers, and thus their gain of rigidity upon complex formation.



**Fig. 8** Flex index plot for the monomers of 1VGC in the complexed state. **a** Chain A, **b** chain B, **c** chain C

Ribbon models mapping the flexibility regions according to the results output by RIGIX are shown in Fig. 6a, b, and c for the oligopeptide (subunit A), subunit B (131 amino acids) and subunit C (97 amino acids), respectively.

All O–H inter-monomer pairs at a distance below 3.5 Å bearing the characteristics of C–H···O hydrogen bonds in the complex are listed in Table 2. Moreover Fig. 7 illustrates the plots of the flex index for the monomer in the isolated state, while Fig. 8 shows the plots of the flex index for the subunits in the complexed state. Finally, Fig. 9 shows plots of the  $\Delta$ Rix index.

Figure 10 shows a ribbon model of the trimer, where the light color indicates regions that remain flexible in the complex that are far from the interface between the subunits. The results generated by RIGIX for the trimer clearly show a pervasive increase in rigidity (or decrease in flexibility) of the three interacting subunits upon complex formation.

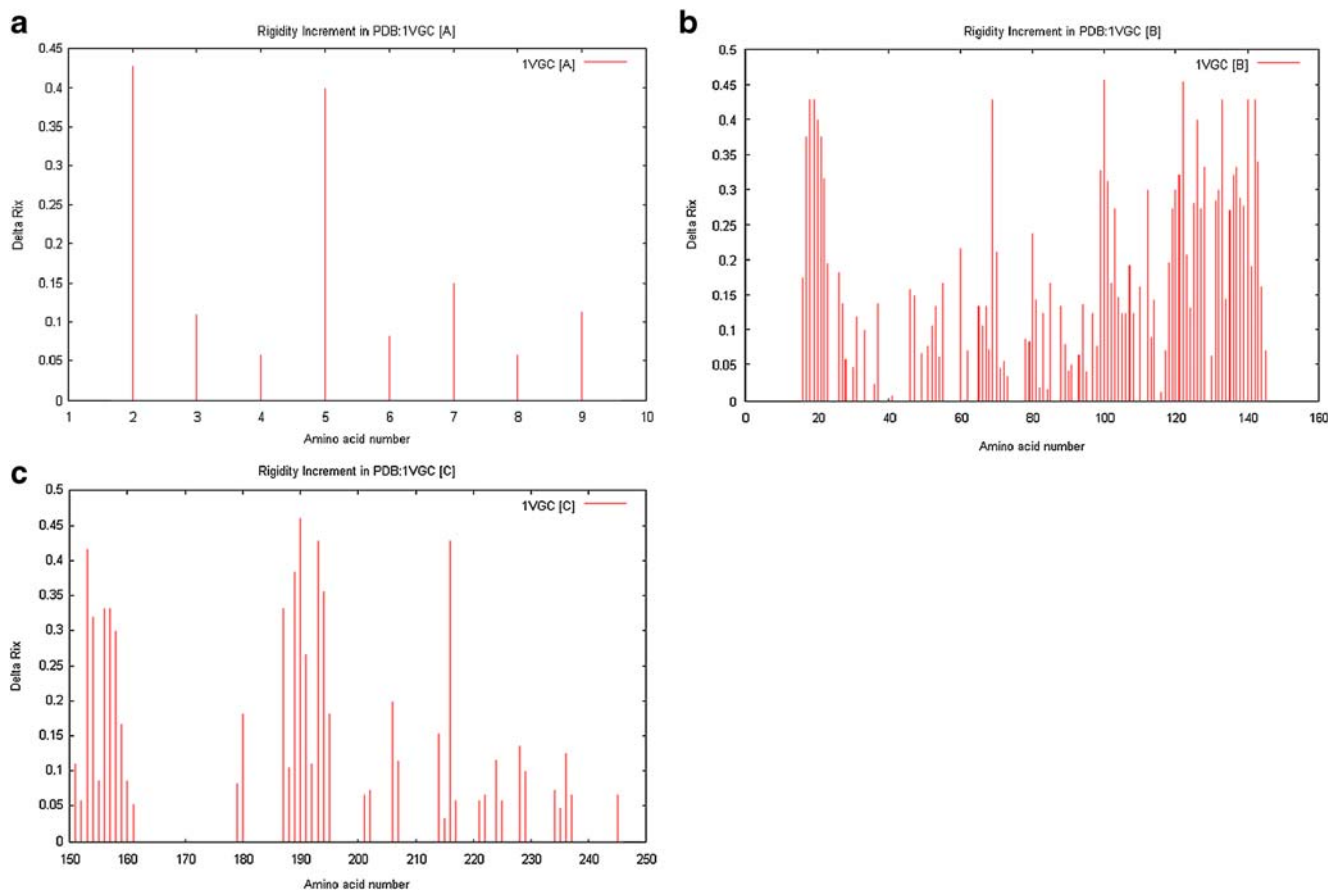
The structure of the oligopeptide changes from completely flexible in the isolated state to almost completely rigid in the complexed state, GLY2 and ALA5 being the amino acids gaining the most structural rigidity in the complex (Fig. 9a). The gain in rigidity of the other two subunits in the complex is similarly profuse, as can be seen

by comparing the respective plots in Figs. 7, 8 and 9 (7a and 8a showing the features for the first subunit in the isolated and complexed states, respectively, and 9a the increment in rigidity). Increasing rigidity is obviously associated with the amino acids located at the interfaces of interaction. The overall index of structural stability for the complex  $S_i$  is equal to 8.67.

### Case study 3: structural rigidity analysis of Bilin binding protein (PDB: 1BBP)

A further example that illustrates the ability of the system to evaluate and quantify changes in rigidity upon bio-molecular complex formation is the structural rigidity analysis of a homo-tetrameric protein complex. The complex studied here is Bilin binding protein (PDB: 1BBP), which has a repeating monomer of 172 amino acids. Application of RIGIX to each of the four subunits and to the whole complex also unveils in this case the change in structural rigidity undergone by the monomers as a consequence of the effects of the interaction and complex formation.

As in the previous examples, ribbon models showing RIGIX-predicted regions of high flexibility are shown in



**Fig. 9**  $\Delta$ Rix index plot for the monomers in 1VGC. **a** Chain A, **b** chain B, **c** chain C

Figs. 11 and 12 for the repeating unit in the isolated state and in the tetramer, respectively.

The set of O–H inter-monomer pairs at distances less than 3.5 Å bearing the characteristics of C–O···H hydrogen bonds in the complex are listed in Table 3 for the tetramer. Moreover, Fig. 13a illustrates the flex index plot for the monomer in the isolated state while Fig. 14 shows plots of the flex index for the repeating subunit in the complexed state. Similarly, Fig. 15 shows plots of the  $\Delta R_{ix}$  index for the monomer.

The corresponding plots of  $\Delta R_{ix}$ , i.e., plots showing the gain in structural rigidity of amino acids of the monomers upon complex formation, are illustrated in Fig. 15. As in the previous examples, the gain in rigidity of the amino acids in the complex is evident. In the present example, however, it is important to note the subtle differences in the patterns of  $\Delta R_{ix}$  for each monomer in spite of the 100% similarity of the sequences. These differences result from the interaction interfaces on the surface of each monomer, which are distinct from one another to differing degrees. In fact, comparison of the flex plots for each monomer after interaction with those before interaction shows that, for



N2, V3, A8, C9, P10, E11, V12, K13, P14, D161, P34, N35, V37, E38, S36, G41, K39, Y40, G52, K53, S54, H64, G65, N178, G78, D79, S80, K84, G95, G96, C119, V140, K139, L141, T142, S171, F170, C175, E172, K176, A173, A174, V177.

**Fig. 11** Flexible regions for the repeating unit in bilin binding protein (IBBP) in the isolated state

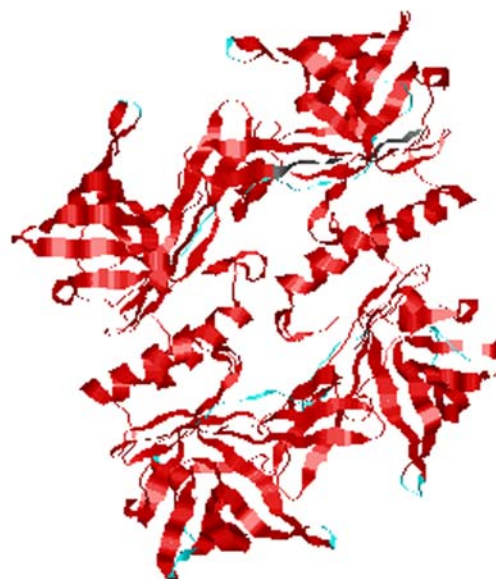


**A:** C1, G2, V3, P4, A5, I6, Q7, P8, V9, L10.

**B:** I16, V17, N18, G19, E20, E21, A22, V23, W27, S119, V121, A112, A55, G69, E70, F71, D72, S113, N95, N100, N101, D102, I103, F114, A120, C122, L123, P124, S125, D128, A126, S127, A131, T134, A132, G133, T135, C136, V137, T138, T139, G140, W141, G142, L143, T144.

**C:** P152, D153, R154, L155, Q156, Q157, A158, S159, L160, P161, G187, V188, S189, S190, C191, D194, M192, G193, K202, W207, G216, S217, C220, S221, P225.

**Fig. 10** Ribbon model for the complex of 1VGC with the remaining flexible regions in *light color*. *Text* List of amino acids changing from flappy to rigid



**A:** N2, V3, D161, P34, N35, V37, E38, S36, G41, K39, Y40, G52, K53, S54, H64, G65, C119, V140, K139, L141, T142, C175, E172, K176, A173, A174.

**B:** N2, V3, D161, P34, N35, E38, S36, V37, G41, K39, Y40, G52, K53, S54, T93, Y94, T98, C119, E123, D124, K139, V140, L141, T142, G143, F170.

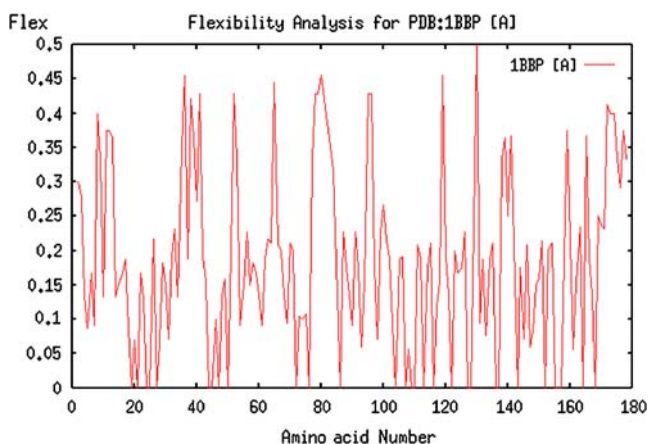
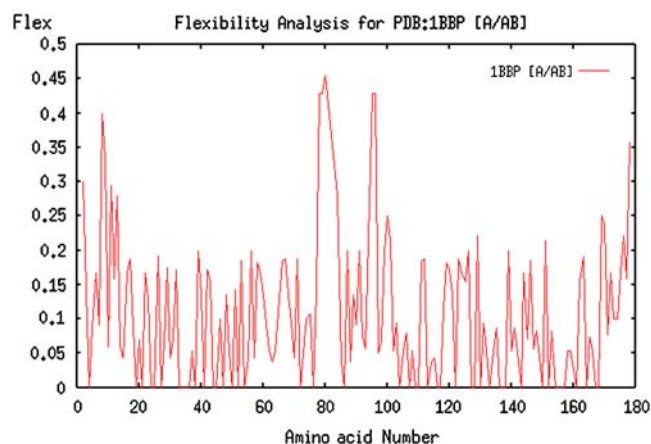
**C:** N2, V3, D161, N17, D161, P34, N35, E38, S36, K39, V37, G41, Y40, H64, G65, N178, G78, D79, C119, K139, V140, L141, T142, S171, F170, A173, A174, V177.

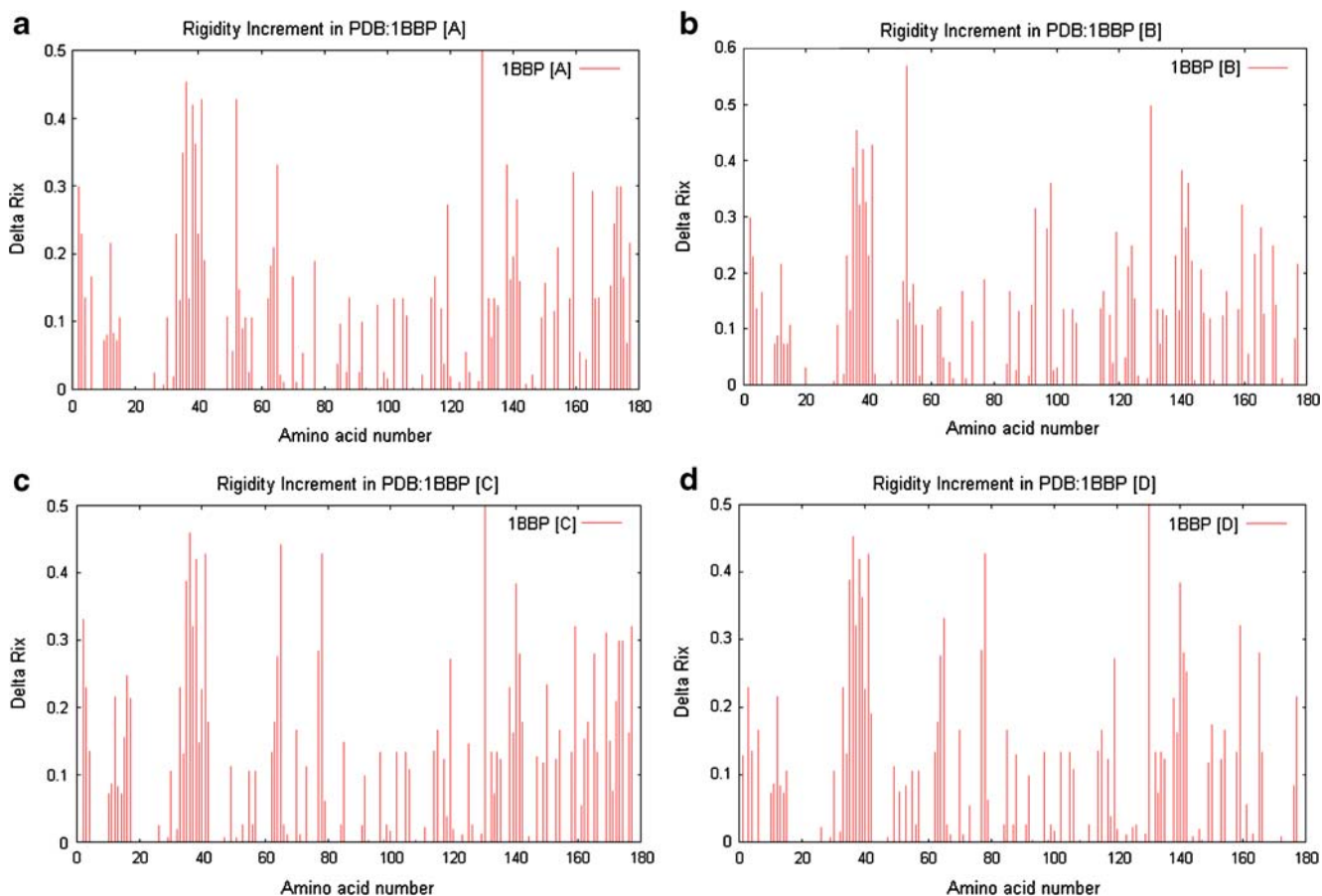
**D:** N1, V3, D161, P34, N35, E38, S36, V37, G41, K39, Y40, H64, G65, G78, D79, C119, K139, V140, L141, T142.

**Fig. 12** Flappy regions in tetrameric IBBP. *A–D:* Lists of amino acids changing from flexible to rigid at complex formation

**Table 3** List of C–H···O type pairs at the interaction interface in 1BBP

Donor					Acceptor				
Atom number	PDB name	Amino acid		Chain	Atom	PDB Name	Amino acid		Chain
		Name	No.				Name	No.	
5451	CB	ASN	2	C	7	OD1	ASN	2	A
5488	CE2	TYR	4	C	7	ODA	ASN	2	A
5490	OH	TYR	4	C	7	OD1	ASN	2	A
6087	CE	LYS	42	C	77	OD1	ASP	6	A
6433	CE1	HIS	64	C	77	OD1	ASP	6	A
5488	CE2	TYR	4	C	531	O	ASN	35	A
7350	CD	LYS	125	C	587	O	LYS	39	A
6061	CE2	TYR	40	C	1865	OD1	ASP	122	A
6056	CB	TYR	40	C	1901	O	LYS	125	A
6412	CG2	ILE	63	C	1901	O	LYS	125	A
6432	CD2	HIS	64	C	1901	O	LYS	125	A
4852	CB	SER	138	B	2308	OE1	GLU	150	A
4852	CB	SER	138	B	2309	OE2	GLU	150	A
4969	CE	LYS	146	B	2518	O	LEU	165	A
2375	CG1	ILE	154	A	5320	O	PHE	170	B
2377	CD1	ILE	154	A	5320	O	PHE	170	B
5	CB	ASN	2	A	5453	OD1	ASN	2	C
43	CZ	TYR	4	A	5453	OD1	ASN	2	C
42	CE2	TYR	4	A	5453	OD1	ASN	2	C
641	CE	LYS	42	A	5523	OD1	ASP	6	C
987	CE1	HIS	64	A	5523	OD1	ASP	6	C
42	CE2	TYR	4	A	5977	O	ASN	35	C
591	CE	LYS	39	A	7311	OD1	ASP	122	C
614	CE1	TYR	40	A	7311	OD1	ASP	122	C
966	CG2	ILE	63	A	7347	O	LYS	125	C
986	CD2	HIS	64	A	7347	O	LYS	125	C

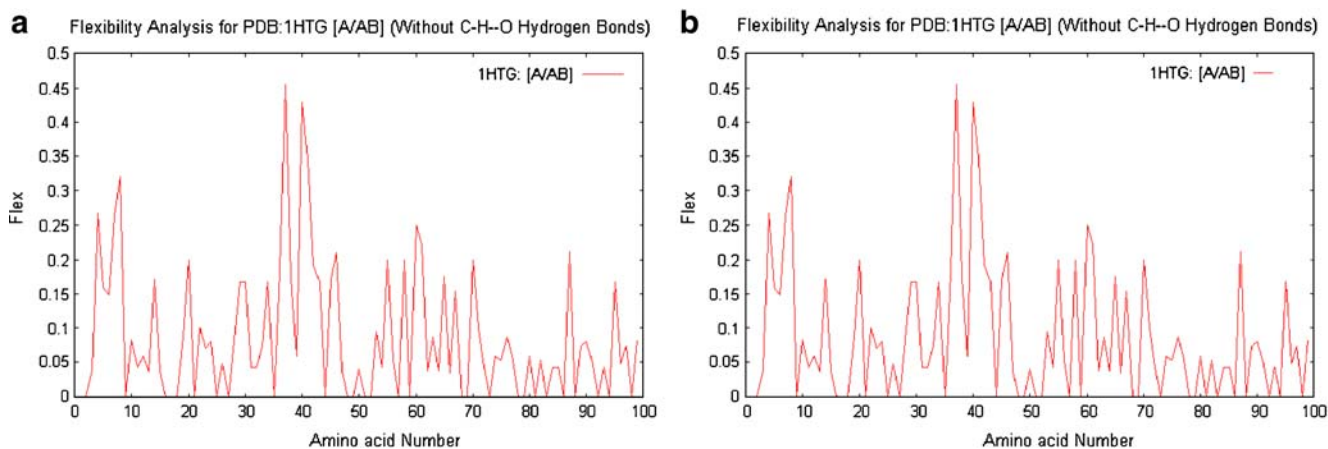
**Fig. 13** Flex index plot for the repeating monomer in 1BBP in the isolated state**Fig. 14** Flex index plot for the repeating monomer in 1BBP in the complexed state



**Fig. 15**  $\Delta$ Rix index plot for the monomers in 1BBP. **a** Chain A, **b** chain B, **c** chain C, **d** chain D

monomers A and B, the regions remaining flexible upon complex formation are those constituted by amino acids 8–12, 78–82, and 90–100. In the case of monomer C, the region comprising amino acids 78–82 has gained rigidity,

while in monomer D, instead of region 78–82, the region retaining high flexibility is that consisting of amino acids 48–50. Finally, the overall rigidity index for the tetramer is 14.89

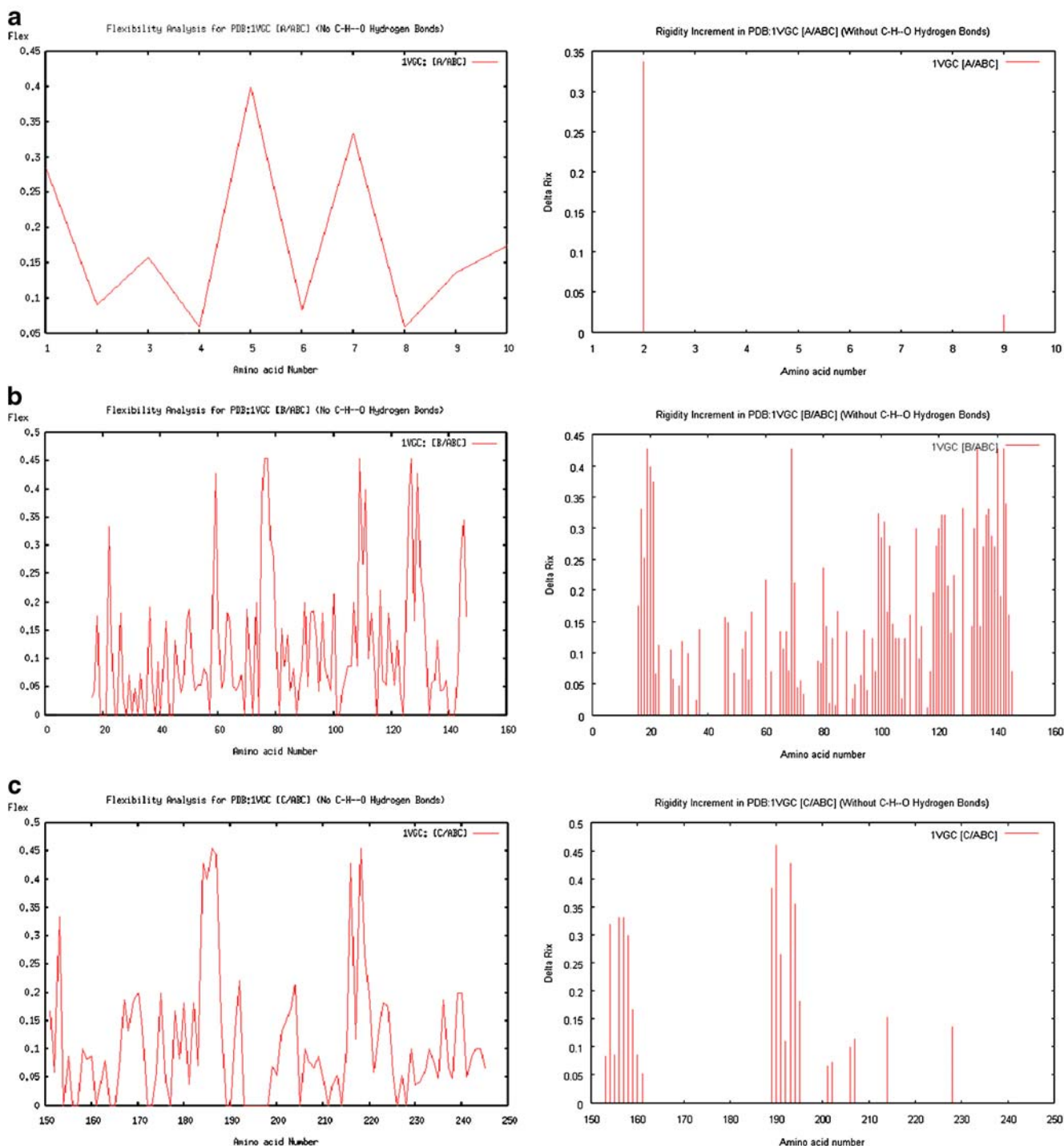


**Fig. 16** **a** Flex and **b**  $\Delta$ Rix index plots for the monomer in 1HTG when C-H...O hydrogen-bond-like pairs are not considered in the RIGIX analysis

The relevance of C-H $\cdots$ O type hydrogen bonds

A computer experiment was performed using RIGIX in order to demonstrate the importance of the C–O $\cdots$ H hydrogen bond in assessing the structural stability of the bio-molecular complexes, as postulated earlier in this

report. The computational experiment consisted of repeating the calculations for the molecules described in the previous section but disregarding any C–H $\cdots$ O type bond. The number of constraints removed by dismissing this type of hydrogen bond in the rigidity analysis are 31, 16, and 26 for 1HGT, 1VGC and 1BBP, respectively. Compared to the

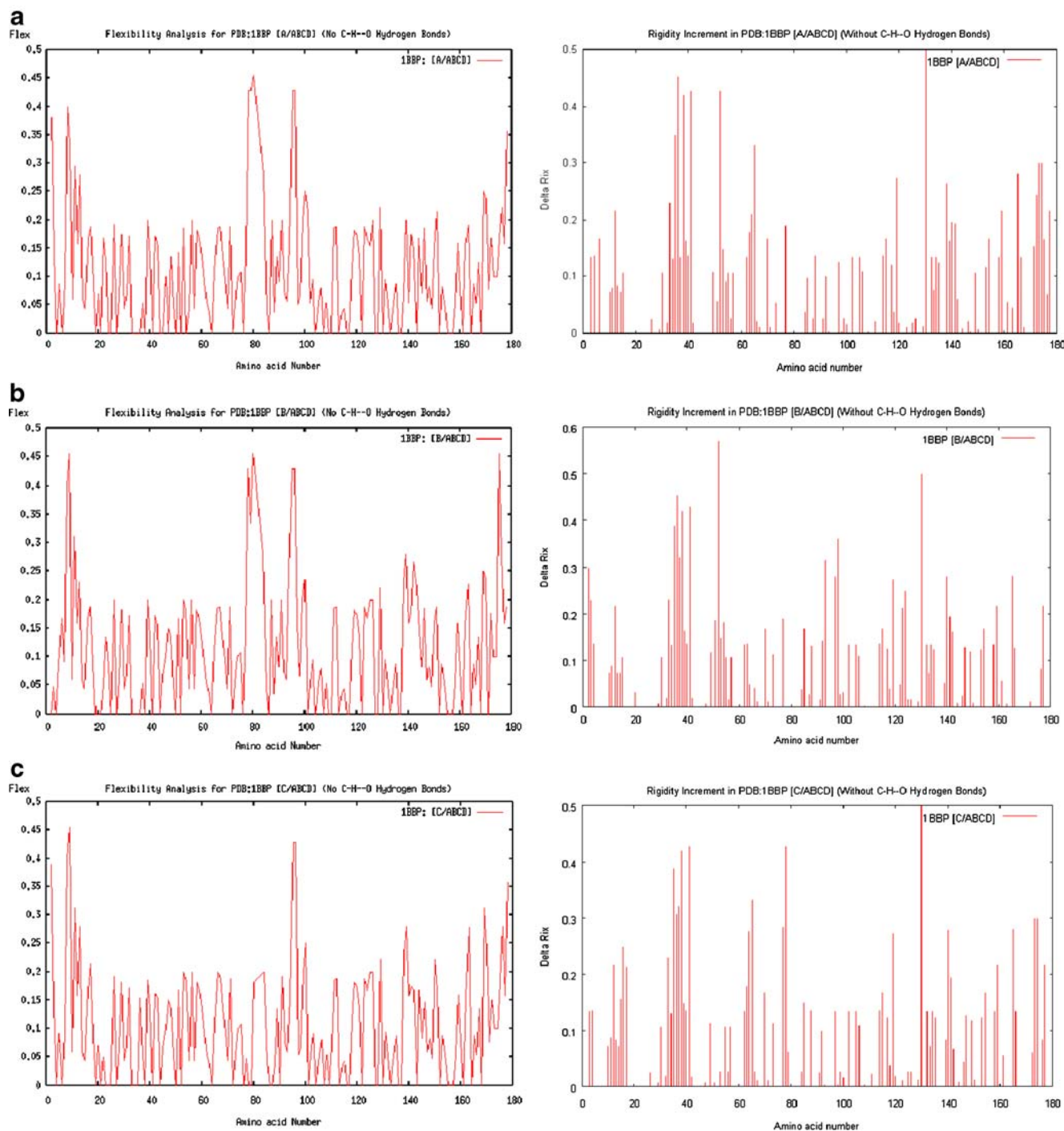


**Fig. 17** Plots of Flex (upper panels) and  $\Delta$ Rix (lower panels) indices for monomers of 1VGC in the complexed state when C–H $\cdots$ O hydrogen-bond-like pairs are not considered in RIGIX analysis. **a** Chain A, **b** chain B, **c** for chain C

total number of constraints for each complex, these numbers are less than 0.2% of the total constraints; however, they account for close to 40% of the constraints at the interface of the complex. The results of applying RIGIX to the evaluation of the structural rigidity of the complexes, dismissing the above-mentioned constraints, are shown in Figs. 16, 17 and 18 for the 1HTG, 1VGC, and

1BBP complexes, respectively. The overall or global rigidity indices for the conformers when C–H···O type hydrogen bonds are not accounted for are summarized in Table 4.

The relevance of considering C–H···O type bonds in the proposed methodology becomes evident when comparing the respective flex and  $\Delta R_{ix}$  plots for the monomers in the



**Fig. 18** Plots of Flex (*upper panels*) and  $\Delta R_{ix}$  (*lower panels*) indices for monomers of 1BBP in the complex state when C–H···O hydrogen-bond-like pairs are not considered in the RIGIX analysis. **a** Chain A, **b** chain B, **c** chain C, **d** chain D



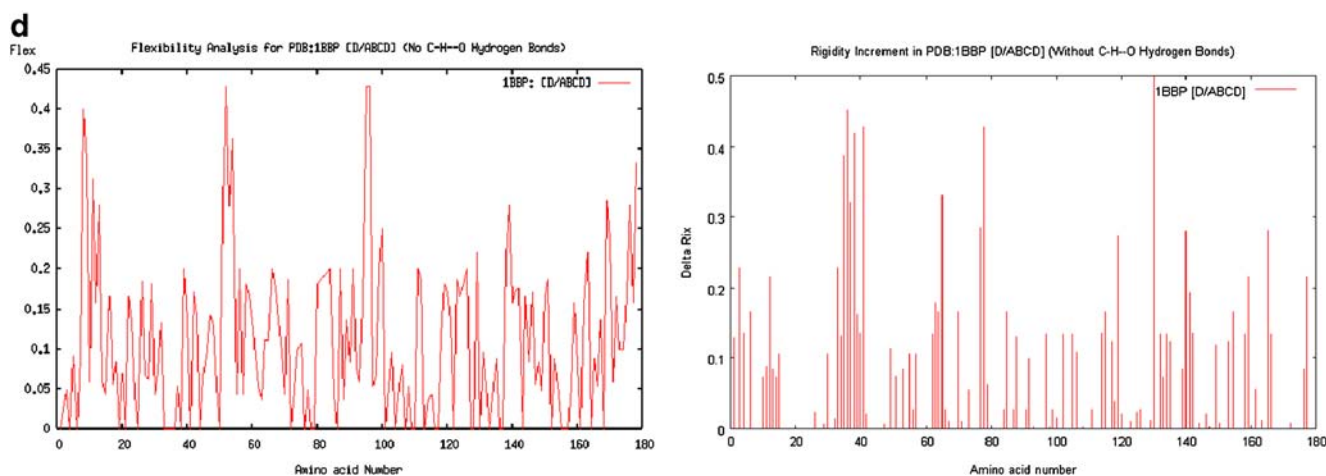


Fig. 18 (continued)

complexed state when these types of bonds are included or excluded in a RIGIX calculation. For the first complex, i.e., 1HTG, 20 amino acids increased in rigidity when the C–O $\cdots$ H constraints were considered (Fig. 4b), in contrast to only 10 amino acids when these constraints were not taken into account (Fig. 16b). Similarly, the flexible regions in the complexed molecule are almost unambiguously identified when the C–O $\cdots$ H type hydrogen bonds are taken into account, while wiggling of the flex plot for this monomer in the complexed state adds some ambiguity to the identification of the flexible regions. This is also reflected in the global structural stability index,  $S_i$ , for the complex, which is 3.63 when C–H $\cdots$ O type bonds are considered, and only 2.21 when no such type of bond is included in the calculation (Table 4).

Further evidence of the importance of considering C–H $\cdots$ O hydrogen bonds as constraints in the RIGIX analysis is obtained by performing the same analysis for the trimer and tetramer structures treated in the previous section.

Flexibility and rigidity plots for 1VGC when the C–O $\cdots$ H hydrogen-bond-like pairs are included are shown in Fig. 8 for flex and Fig. 9 for  $\Delta R_{ix}$ , while the plots for flex and  $\Delta R_{ix}$  when those constraints are excluded are shown in Fig. 17. The flex plots for the oligopeptide in 1VGC (Chain A) in Fig. 8 and Fig. 17 show accurately the importance of the C–H $\cdots$ O hydrogen, since where the bonds are included there is only one peak showing flexible amino acids (Fig. 8), while there are two when those bonds are not considered (Fig. 17). Moreover, the magnitudes of the flex

indices for the monomers in the complexes are lower when the hydrogen bonds are included than when they are not. Comparing the  $\Delta R_{ix}$  indices for the oligomer, we can see that only 1 amino acid increases in rigidity when no C–H $\cdots$ O constraint is considered (Fig. 17), while the rigidity of 8 of the 13 amino acids in the oligomer increases when they are considered (Fig. 9). Consideration of this type of hydrogen bond aids in unambiguous identification of the amino acids, and thereby, flexible regions that undergo the largest changes. The same conclusions can be drawn when the flexibility and rigidity plots are compared for monomers B and C of 1VGC. In addition, the global structural stability index  $S_i$  for the trimer is 8.67 when C–H $\cdots$ O bonds are considered, while it is only 6.91 when they are not (Table 5).

Furthermore, comparison of the RIGIX results when C–O $\cdots$ H type hydrogen bonds are included and excluded for the tetramer lead to the same results as for the previous examples. The figures to compare are Fig. 14 with Fig. 18a for flexibility (flex) and Fig. 15 with Fig. 18b for rigidity. Finally, in this case, the structural stability index  $S_i$  for the tetramer is 14.89 when the C–H $\cdots$ O bonds are considered, while it is only 12.62 when they are not (Table 4).

Consequently, inclusion of C–H $\cdots$ O type hydrogen bonds is critical in the identification of the amino acids whose structural rigidity has increased as an effect of interactions upon complex formation in a rather unambiguous way since they constitute more constraints for the atoms involved in hydrogen bond interaction, and conse-

**Table 4** Overall rigidity indices for complexes 1HGT, 1VGC, and 1BBP

Compound	Overall rigidity index ( $S_i$ ) with C–H $\cdots$ O bonds	Overall rigidity index ( $S_i$ ) w/o C–H $\cdots$ O bonds	No. of C–H $\cdots$ O bonds
1HTG	3.632	2.214	31
1VGC	8.979	6.907	16
1BBP	14.799	12.619	26

**Table 5** Rigidity evaluation for different protein complexes (flexible regions in light color, rigid regions in dark). \*Composition of the flexible regions. \*\*Amino acids changing from flappy to rigid upon complex formation

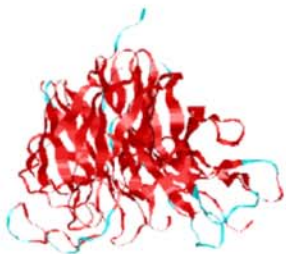
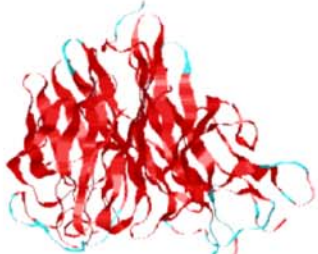

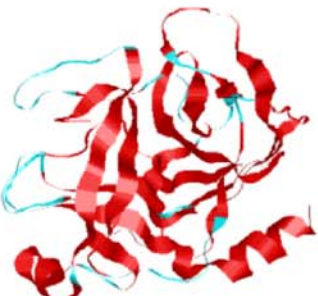

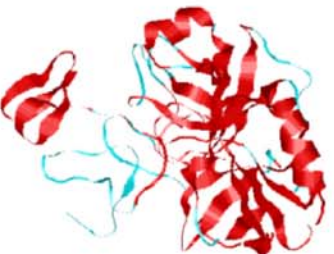



<p><b>1A4Q</b></p>	 <p><b>A:</b> E76, P77, E78, W79, Q87, G88, L453, L454, E104, N108, S109, A110, P123, K124, E125, P138, G139, G140, Y141, Y142, N143, G144, G433, D148, G345, T333, G334, P335, G340, D341, K342, G343, R344, G346, A357, S358, K359, D431, G432, K434, K435, T436, W437, H438, S439, S450, G451, Q452, W455, D462.</p>	 <p><b>B:</b> E76, P77, E78, W79, L453, L454, E104, N108, S109, A110, P123, K124, E125, P138, G139, G140, Y141, Y142, N143, G144, G433, D148, G345, T333, G334, P335, N339, G340, D341, K342, G343, R344, G346, A357, S358, K359, D431, G432, W437, H438, S439, S450, G451, Q452, W455, D462.</p>	 <p><b>A:</b> E104, N108, S109, A110, P138, G139, G140, Y141, Y142, N143, G144, D148, D462. <b>B:</b> L453, L454, E104, N108, P123, K124, E125, S450, G451, Q452, W455, D462.</p>
<p><b>1EJA</b></p>	 <p><b>A:</b> A23, S26, A24, N25, I27, K60, S61, R62, I63, E70, H71, S116, R117, V118, L123, P124, R125, S127, C128, A129, A130, S146, T144, K145, S147, G148, S149, S150, Y151, P152, S153, S164, I176, T177, L185, G187, E186, N223, G188, K224, N202, G203, G219, C220, A221, K222, P225.</p>	 <p><b>B:</b> D30, L31, H32, C33, K34, V35, K36, C37, G40, E38, H39, A55, F41, K42, C54, K43, Y50, D44, D45, N46, G47, C48, E49, A51, C52, I53, D56, A57, P58, Q59.</p>	 <p><b>A:</b> G219. <b>B:</b> L31, H32, C33, K34, V35, K36, C37.</p>
<p><b>1EUV</b></p>	 <p><b>A:</b> G401, S402, L403, V404, P405, E406, L407, N408, R422, E423, N472, M491, T496, Q497, P596, L597, D598, F599.</p>	 <p><b>B:</b> P20, E21, T22, D68, G69, Q95, I96, G97, G98.</p>	 <p><b>A:</b> N472, M491. <b>B:</b> D68, G69, Q95, I96, G97, G98.</p>

Table 5 (continued)



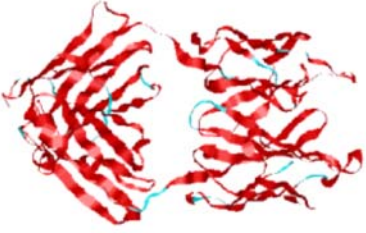
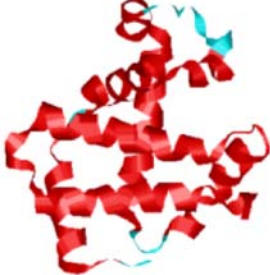


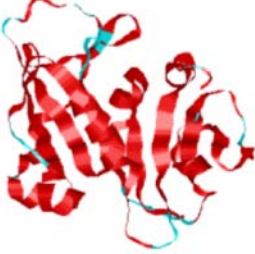
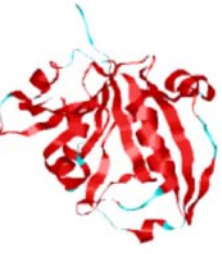
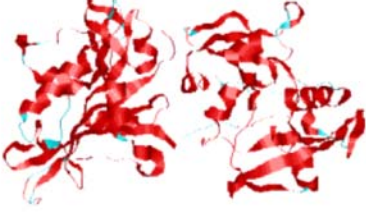
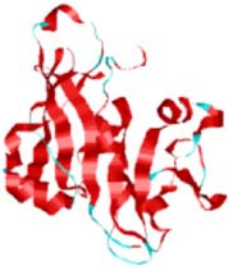

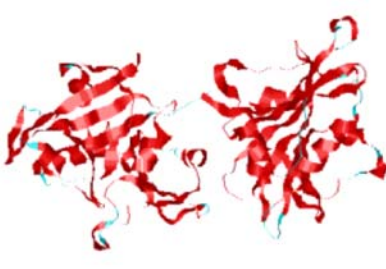

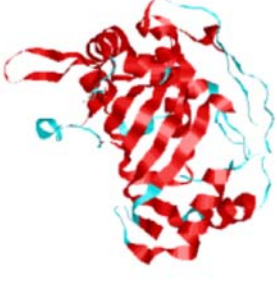

<b>1FLR</b>	 <p><b>A:</b> D1, V2, V3, M4, S61, G62, V99, P100, W101, G105, G106, Q161, N162, G163, S173, K174, D175.</p>	 <p><b>B:</b> E1, V2, K3, G8, G9, G10, G26, F27, T28, P41, E42, K43, G44, Y103, G104, S118, A119, K120, T121, P131, V132, T136, T137, G138, S139, S140, V141, L164, S165, S166, G167, P172, A173, P217, R218.</p>	 <p><b>B:</b> Y103, G104, P131, V132, P172, A173.</p>
<b>1I3E</b>	 <p><b>A:</b> G1, H2, N19, F45, D43, S44, G46, N47, S49, L48, S50, D80, L78, D79.</p>	 <p><b>B:</b> G1, H2, N19, N47, S49, L48, S50, D80, L78, D79.</p>	 <p><b>A:</b> N19, F45, G46, N47, S49, L48, S50, D80, L78, D79. <b>B:</b> N19, D80, L78, D79.</p>
<b>1IA1</b>	 <p><b>A:</b> L2, K3, P4, N5, G20, P26, W27, R28, L29, P46, N47, T48, P63, R67, P68, E82, N83, S95, D105, P138, S139, E144, M145, T147, P160, E174, D175, L173.</p>	 <p><b>B:</b> M1, L2, K3, P4, N5, P26, W27, R28, L29, P63, R67, P68, E82, N83, S95, V103, S104, D105, V106, P138, S139, E144, M145, P160, E174, D175, V172, L173.</p>	 <p><b>A:</b> L2, K3, P4, N5, S95, D105. <b>B:</b> S95.</p>

Table 5 (continued)

1IA2	 <p><b>A:</b> L2, K3, P4, N5, G20, P26, W27, R28, L29, T44, K45, P46, N47, T48, P63, R67, P68, E82, N83, S95, D105, P138, S139, E144, M145, T147, P160, E174, D175, D170, T171, V172, L173.</p>	 <p><b>B:</b> M1, L2, K3, P4, N5, P15, A16, L149, I19, G20, M25, P26, W27, R28, L29, K45, P46, N47, T48, R49, P63, F66, Q64, K65, R67, P68, E82, N83, V103, S104, D105, V106, K150, H137, P138, S139, E144, M145, F148, F151, P152, P160, V172, L173, E174, D175, G180, D181.</p>	 <p><b>A:</b> L2, K3, P4, N5, S95, D105. <b>B:</b> P15, A16, L149, I19, G20, M25, K45, P46, N47, T48, R49, F66, Q64, K65, K150, H137, F148, F151, P152, G180, D181.</p>
2TSC	 <p><b>A:</b> M1, G41, P43, L44, V45, T46, C50, T47, K48, R49, H51, L52, T78, I79, W80, G91, P92, P104, D105, G106, R126, R127, W133, N134, E137, V135, G136, A142, L143, A144, D156, G157, A235, V262, A260, E223, P226, L227, P228, R234, P236, E237, S238, Y242, Y252, D253, P254, P256, G257, I258, K259, P261, A263, I264.</p>	 <p><b>B:</b> M1, N19, D20, T24, R21, T22, G23, G25, F30, G31, H32, G41, F42, P43, L44, V45, T46, H255, R225, T47, K48, R49, C50, H51, L52, T78, D81, I79, W80, W83, E82, A84, A142, G91, S125, P123, D124, R127, R126, I128, E137, V135, G136, L138, M141, D139, K140, L143, A144, D156, G157, H212, N211, R222, S221, E223, P224, P226, L227, P228, K229, E250, L230, I231, E248, I232, K233, R234, A235, P236, I249, G251, Y252, D253, P254, P256, G257, I258, K259, A260, P261, V262, A263, I264.</p>	 <p><b>A:</b> P104, D105, G106, R126, R127, W133, N134, E137, V135, G136, D156, G157. <b>B:</b> N19, D20, T24, G25, F30, G31, H32, S125, P123, D124, R127, R126, I128, E137, V135, G136, D156, G157.</p>

**Table 6** Stability index (Si) for protein complexes

Complex (PDB Code)	Si
1A4Q	5.44200
1EJA	1.21600
1EUV	3.85600
1FLR	1.45900
1I3E	3.69250
1IA1	1.62450
1IA2	8.69150
2TSC	6.00850

quently contribute to diminishing the DOF in the network of interactions.

#### Broader validation of the methodology

RIGIX was applied to a broader spectrum of protein complexes for further validation of the methodology as well as to gain deeper insights into the characteristics of the proposed indices to quantitatively express structural rigidity or flexibility and to establish a function based on these that may enable the ranking of automatically generated protein complexes from isolated monomers—the PPI prediction problem. We selected eight other complexes besides those treated so far, and have performed the RIGIX calculation as explained in the preceding sections. The results are summarized in Table 5, where the structures treated are shown in ribbon model representation, on which flexible (cyan) and rigid (red) regions have been mapped according to RIGIX outputs. Under each model of the interacting monomers, the amino acids belonging to the flexible regions have been plotted, while those amino acids that change from structurally flexible to rigid are enumerated under the structure of the complex (the rightmost structure). In all cases, there is a decrease in the number of amino

acids that remain flexible in the complex as compared to the isolated subunits, and the number of amino acids changing their structural flexibility is proportional to their proximity to the interaction interface. Table 6 illustrates the overall rigidity index Si for these complexes. It can be noted that the larger the Si the more stable the complex is, in terms of both the number of interactions at the interface and the surface area of the interface.

Structural flexibility as an index to rank automatically generated protein complex decoys

Finally, application of changes in flexibility to score protein complex decoys generated by the docking module in MIAX is shown in Table 7. The table lists the PDB codes for the complex and the interacting subunits (ligand and receptor), the root mean squares (RMS) value for the best decoy output by MIAX, together with the ranking value from geometrical instances (G-rank) and the Flexibility Rank, which is the ranking value for the decoy based on comparison of the Si values for all the decoys generated by the system. The tendency is evidently towards a better scoring when flexibility loss or rigidity gain is taken into account than when not, as can be observed by the arrow (pointing upwards) in most of the re-ranked protein complex decoys (Table 7). This evaluation shows the suitability of the proposed indices to rank the decoys, and the more so when interfaces of the interacting subunits undergo conspicuous rearrangement upon interaction and complex formation.

#### Conclusions

We have adapted constraint counting methodology to analyze structural rigidity in bio-molecular complexes,

**Table 7** Scoring of automatically generated protein complex decoys using RIGIX. RMS Root mean square

Complex	Ligand	Receptor	RMS rank	G-rank	Flexibility rank
Lay7	1bta	1box	1.536	1,052	105 ↑
1brb	1bpi	1bra	4.321	3,727	1411 ↑
1bzx	1bpi	1bit	6.473	3,320	171 ↑
1slu	1ecy	1ane	4.650	2,950	86 ↑
1tpa	4pti	1az8	4.427	2,913	1704 ↑
1ugh	2ugi	1akz	2.00	209	627 ↓
2kai	1bpi	2pka	6.480	484	144 ↑
2pcf	1ag6	1ctm	6.00	2,715	184 ↑
2ptc	1bpi	1auj	2.829	1,370	456 ↑
2sic	3ssi	1st2	3.607	3,500	604 ↑
3btd	1bpi	1auj	2.829	1,766	636 ↑
3tgi	1bpi	1ane	3.043	231	59 ↑

and propose several indices to express quantitatively local as well as overall structural rigidity gain upon complex formation.

Interaction leads to flexibility loss of the monomers and an overall gain of rigidity in the complex system. The methodology intrinsically manipulates the network of inter- and intra-molecular atomic interactions that underlies the principles of bio-macro-molecular interaction and complex formation. We have expressed quantitatively several other interaction terms in addition to the hitherto considered weak forces set of interaction via which two biomolecules interact. Among them, we have demonstrated, in agreement with work reported in the literature, that C–O···H bonds play a critical role in the assessment of the local and global rigidity change upon complex formation. The methodology is not limited by the number of interacting monomers but can be applied to multimeric complexes without any substantial modification in the underlying algorithm as shown for the 1VGC and 1BBP, a trimer and tetramer, respectively.

Applying the methodology as a scoring scheme to rank higher automatically generated protein complex decoys, as illustrated in Table 7, indicates that in fact the rigidity/flexibility term is significant and it can be a major tool in the selection of complexed decoys with a close-to-native structure.

A more exhaustive statistical analysis of the effects of flexible domains in the function of proteins is now possible, thus laying the groundwork for establishing ranking schemes for automatic systems for prediction of the structure of protein complexes, like the system MIAX [7, 9].

Only protein–protein interactions were treated here, but generalization of the methodology to treat multimers composed of other different bio-molecules like RNAs or DNAs is under development.

## References

- Jacobs JD, Rader AJ, Kuhn LA, Thorpe MF (2001) Protein Flexibility Predictions Using Graph Theory. *Proteins: Struct Funct Genet* 44:150–165
- CAPRI: Critical Assessment of Prediction of Interactions. <http://www.ebi.ac.uk/msd-srv/capri/>
- Janin J (1997) Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins* 28:153–161
- Fischer D, Lin SL, Wolfson L, Nussinov R (1995) A geometry-based suite of molecular docking processes. *J Mol Biol* 248:459–477
- Weng Z, Vajd S, Delisi C (1996) Prediction of protein complexes using empirical free energy functions. *Protein Sci* 5:614–626
- Palma PN, Krippahl L, Wampler JE, Moura JGG (2000) Bigger: a new (soft) docking algorithm for predicting protein interactions. *Proteins* 39:372–384
- Del Carpio CA, Ichiishi E, Yoshimori A, Yoshikawa T (2002) MIAX: A new paradigm for modeling biomacromolecular interactions and complex formation in condensed phases. *Proteins: Struct Funct Genet* 48:696–732
- Katchalski-Katzir E, Shariv I, Eisenstein M, Friesema AA, Aflalo C, Vakser IA (1992) Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci USA* 89:2195–2199
- Del Carpio CA, Peissker T, Yoshimori A, Ichiishi E (2003) Docking unbound proteins with MIAX: a novel algorithm for protein–protein soft docking. *Genome Inform* 14:238–249
- Jacobs DJ (1998) Generic rigidity in three-dimensional bond-bonding networks. *J Phys A: Math Gen* 31:6653–6668
- Jacobs DJ, Hendrickson B (1997) An algorithm for two-dimensional rigidity percolation: the pebble game. *J Comput Phys* 147:346–365
- Hendrickson B (1992) Conditions for unique graph realizations. *SIAM J Comput* 21:65–84
- Moukarzel C (1996) An efficient algorithm for testing the generic rigidity of planar graphs. *J Phys A: Math Gen* 29:8079–8098
- Livesay DR, Dallakyan S, Wood GG, Jacobs DJ (2004) A flexible approach for understanding protein stability. *FEBS Lett* 576:468–476
- Jiang L, Lai L (2002) CH···O Hydrogen bonds at protein–protein interfaces. *J Biol Chem* 277(40):37732–37740
- Panigrahi SK, Desiraju GR (2007) Strong and weak hydrogen bonds in the protein–ligand interface. *Proteins: Struct Funct Bioinform* 67:128–141
- Morozov AV, Kortemme T, Tsemekhman K, Baker D (2004) Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc Natl Acad Sci USA* 101(11):6946–6951
- Senes A, Ubarretxena-Belandia I, Engelman DM (2001) The C $\alpha$ –H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci USA* 98(9):9056–9061
- Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285(4):1735–1747
- Brasseur R (1991) Differentiation of lipid-associating helices by use of three-dimensional molecular hydrophobicity potential calculations. *J Biol Chem* 266(24):16120–16127