

On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys

Ezgi Karaca and
Alexandre M. J. J. Bonvin*

Bijvoet Center for Biomolecular Research,
Faculty of Science, Padualaan 8,
3548 CH Utrecht, The Netherlands

Correspondence e-mail: a.m.j.j.bonvin@uu.nl

Received 22 October 2012

Accepted 13 March 2013

Scoring, the process of selecting the biologically relevant solution from a pool of generated conformations, is one of the major challenges in the field of biomolecular docking. A prominent way to cope with this challenge is to incorporate information-based terms into the scoring function. Within this context, low-resolution shape data obtained from either ion-mobility mass spectrometry (IM-MS) or SAXS experiments have been integrated into the conventional scoring function of the information-driven docking program *HADDOCK*. Here, the strengths and weaknesses of IM-MS-based and SAXS-based scoring, either in isolation or in combination with the *HADDOCK* score, are systematically assessed. The results of an analysis of a large docking decoy set composed of dimers generated by running *HADDOCK* in *ab initio* mode reveal that the content of the IM-MS data is of too low resolution for selecting correct models, while scoring with SAXS data leads to a significant improvement in performance. However, the effectiveness of SAXS scoring depends on the shape and the arrangement of the complex, with prolate and oblate systems showing the best performance. It is observed that the highest accuracy is achieved when SAXS scoring is combined with the energy-based *HADDOCK* score.

1. Introduction

Understanding how a single cell functions is the fundamental quest of life sciences. This can only be comprehensively addressed once the structure–function relationships of the biomolecular complexes occurring in that particular cell have been thoroughly explored. There are two main classical experimental techniques that can reveal the structure of the biomolecular complexes in atomistic detail: X-ray crystallography and NMR spectroscopy. Although these have helped immensely to shed light on the mechanical and functional world of biomolecules, they are faced with many challenges when the biomolecular systems under study become very large, comprise flexible or unstructured regions, exist in very tiny amounts, are membrane-associated and/or when their constituents interact only transiently (Putnam *et al.*, 2007; Ruotolo *et al.*, 2008). While these limitations hamper the acquisition of high-resolution information, low-resolution biochemical or biophysical data can often still be obtained. A limitation here is that most of the time these data are sparse and contain limited structural information compared with high-resolution methods. Therefore, computational modelling using integrative approaches such as macromolecular docking are needed to translate these sparse low-resolution data into useful structural information (Bonvin, 2006; Putnam *et al.*, 2007).

Low-resolution spatial information can be derived from a variety of biophysical experimental techniques comprising NMR (Melquiond & Bonvin, 2010; Madl *et al.*, 2011; Wang *et al.*, 2011), cryo-EM (Alber *et al.*, 2008; Lander *et al.*, 2012), MS (Rappsilber, 2011), IM-MS (Utrecht, Rose *et al.*, 2010; Jurneczko & Barran, 2011), EPR (Steinhoff, 2004), SAXS (Putnam *et al.*, 2007; Madl *et al.*, 2011), FRET (Brunger *et al.*, 2011) and/or biochemical experiments such as mutagenesis and chemical footprinting (Alber *et al.*, 2008; Melquiond & Bonvin, 2010; Garcia-Garcia *et al.*, 2012). The data obtained from these experiments can be integrated into a modelling procedure either during sampling, *i.e. a priori*, by restraining the conformational search space, or during scoring, *i.e. a posteriori*, by filtering or scoring the generated models based on the discrepancy between the experimentally measured structural properties and back-calculated properties. In this work, we focus on the latter and assess the applicability of low-resolution shape data obtained from either IM-MS or SAXS experiments for scoring decoys generated by macromolecular docking using the information-driven docking approach *HADDOCK* (Dominguez *et al.*, 2003; de Vries *et al.*, 2010). The reasons for first focusing on these experimental techniques are that they are rapid, effective and can be applied to a very broad mass range (Putnam *et al.*, 2007; Utrecht, Rose *et al.*, 2010; Jurneczko & Barran, 2011; Madl *et al.*, 2011).

IM-MS is a combination of two different spectrometric techniques, ion mobility and mass spectrometry, which work under gas-phase conditions. Coupling IM with MS provides information about the mass, subunit composition and collision cross-section (CCS) of the biomolecular complex under study (Ruotolo *et al.*, 2008; Politis *et al.*, 2010; Utrecht, Rose *et al.*, 2010). The CCS is defined as the rotationally averaged area of a molecule that is available to interact with the buffer gas, and contains low-resolution (one-dimensional) shape-related information (Utrecht, Rose *et al.*, 2010). It has been demonstrated that CCS values estimated for the lowest charge states measured under gas-phase conditions often correlate with those simulated from X-ray and NMR structures (Scarff *et al.*, 2008; Utrecht, Rose *et al.*, 2010; Zhou & Robinson, 2010; Benesch & Ruotolo, 2011). Thus far, CCS-guided modelling has been conducted by comparing experimental CCS values with those simulated from models. It has been used to understand various important biological phenomena such as virus capsid formation (Utrecht, Barbu *et al.*, 2010), aggregation of amyloid fibrils (Bernstein *et al.*, 2009), protein-folding pathways (Ruotolo *et al.*, 2007) and potential conformational changes (Politis *et al.*, 2010). Recently, D'Abramo and coworkers have used CCS data to score complexes generated by docking (D'Abramo *et al.*, 2009). They tested their CCS-based scoring function on available CAPRI (Critical Assessment of PRediction of Interactions) targets and observed that the probability of selecting a near-native solution was significantly higher than a random selection.

SAXS allows characterization of biomolecular complexes under native conditions (Putnam *et al.*, 2007; Mertens & Svergun, 2010). It measures the intensity scattered by a protein sample at low scattering angles. From the scattering

profile, size-related and shape-related information can be extracted, such as the molecular weight, radius of gyration, maximum molecular dimension and a low-resolution three-dimensional molecular envelope (Putnam *et al.*, 2007; Buey *et al.*, 2009; Madl *et al.*, 2011). The fact that SAXS experiments can be conducted in a rapid manner has recently made it possible to run them in a high-throughput manner (Hura *et al.*, 2009). Over the years, SAXS has been used to shed light on a variety of challenging vital biological processes such as folding–unfolding events, conformational changes and oligomerization processes (Putnam *et al.*, 2007; Mertens & Svergun, 2010; Madl *et al.*, 2011). To this end, the SAXS information has typically been translated into (i) molecular envelopes, which can guide the construction of *ab initio* bead models or the docking of individual three-dimensional structures (Svergun *et al.*, 2001; Buey *et al.*, 2009; Mertens & Svergun, 2010), (ii) a restraining energy term, which can be used to refine a structure directly against the SAXS curve (often in combination with orientational NMR restraints; Grishaev *et al.*, 2005; Förster *et al.*, 2008; Madl *et al.*, 2011), or (iii) a scoring term that calculates the discrepancy between the experimental scattering curve and the back-calculated curve from the model (Filgueira de Azevedo *et al.*, 2003; Sondermann *et al.*, 2005; Covaceuszach *et al.*, 2008). SAXS-based scoring has recently been incorporated into two *ab initio* docking methods, *pyDock* (Pons *et al.*, 2010) and *PatchDock* (Schneidman-Duhovny *et al.*, 2010, 2012), revealing that SAXS integrated scoring improves the accuracy of model selection significantly.

Here, we present a thorough and systematic investigation of the information content of both CCS and SAXS data and their usefulness in filtering docking decoys based on a benchmark of 176 complexes (Docking Benchmark 4.0; Hwang *et al.*, 2010). For this we used *HADDOCK*, an information-driven docking program that allows the inclusion of various types of sparse experimental data to drive the modelling of biomolecular complexes (Dominguez *et al.*, 2003; de Vries *et al.*, 2010). The docking procedure in *HADDOCK* is composed of three stages: initial docking by rigid-body energy minimization (it0), semi-flexible refinement in torsion-angle space (it1) and final refinement in explicit solvent (water). The binding mode of the complex is roughly determined during it0 and the top-scoring models (typically the top 200–400, ranked according to the *HADDOCK* score) are then selected for further refinement. In this study, we demonstrate that this selection process plays a critical role in the accuracy of the final models. For this, we construct two different scoring functions by combining CCS information from IM-MS or SAXS profiles with the conventional *HADDOCK* score, and evaluate their performance on docking decoys obtained by running *HADDOCK* in *ab initio* mode. We analyze the strengths and weaknesses of each data type within the context of macromolecular docking and reveal that the highest accuracy in scoring can be obtained when these terms are combined with the conventional *HADDOCK* score. Last but not least, this study also defines the baseline performance of *HADDOCK* both for *ab initio* docking and for the inclusion of shape data as a filter.

2. Materials and methods

2.1. Running HADDOCK in *ab initio* mode

All complexes were docked using the *ab initio* mode of HADDOCK 2.2 with only centre-of-mass (CM) restraints. CM restraints are distance restraints defined between the centres of mass of the molecules. They are automatically calculated from the dimensions of each component along its principal x , y , z axis (d_x , d_y , d_z) as

$$d_{\text{CM}} = \frac{\sum_{i=1}^{N \leq 6} (d_{x,i} + d_{y,i} + d_{z,i} - d_{\text{max},i})}{4}, \quad (1)$$

where N is the number of constituents that are docked, d_x , d_y and d_z are the dimensions along each principal axis and d_{max} is the longest dimension (the latter is subtracted to ensure the generation of tight restraints). From this, a distance restraint is defined with an upper distance bound set to $(d_{\text{CM}} + 1)$ Å and a lower bound equal to 0.

For each docking, 10 000 rigid-body structures were generated with $n_{\text{trials}} = 5$ (meaning that five docking trials were performed and the best solution was kept) starting from random orientations. For each solution, the 180° rotated solution was also automatically sampled, resulting in the effective sampling of 100 000 decoys, of which only 10 000 were written to disc. The number of structures for subsequent refinement (it1, water) was set to 400. In the case of symmetrical multimers, proper noncrystallographic symmetry (NCS) and symmetry restraints were used as described previously (Karaca *et al.*, 2010). The solutions were ranked at the end of each docking stage according to the following HADDOCK scores:

$$\text{it0: } 0.01E_{\text{vdw}} + 1.0E_{\text{Elec}} + 0.01E_{\text{CM}} - 0.01\text{BSA} + 1.0E_{\text{Desolv}} + 0.1E_{\text{Sym}}, \quad (2)$$

$$\text{it1: } 1.0E_{\text{vdw}} + 1.0E_{\text{Elec}} + 0.1E_{\text{CM}} - 0.01\text{BSA} + 1.0E_{\text{Desolv}} + 0.1E_{\text{Sym}}, \quad (3)$$

$$\text{water: } 1.0E_{\text{vdw}} + 0.2E_{\text{Elec}} + 0.1E_{\text{CM}} + 1.0E_{\text{Desolv}} + 0.1E_{\text{Sym}}, \quad (4)$$

where E_{vdw} is the van der Waals intermolecular energy, E_{Elec} is the intermolecular electrostatic energy, E_{CM} is the distance-restraint energy, E_{Desolv} is an empirical desolvation energy term (Fernández-Recio *et al.*, 2004), BSA is the buried surface area and, where present, E_{Sym} is the symmetry-restraint energy. The nonbonded interactions (E_{vdw} and E_{Elec}) were calculated with an 8.5 Å cutoff using OPLS parameters (Jorgensen & Tirado-Rives, 1988). The final models were clustered based on the pairwise ligand interface r.m.s.d. with a minimum cluster size of four and an r.m.s.d. cutoff of 7.5 Å. The resulting clusters were ranked based on the average score of their top four members.

2.2. Generation of synthetic CCS values

The Leeds method (Smith *et al.*, 2008) was used to generate synthetic CCS values for the native complexes and for the

docking models. The Leeds method is a Monte Carlo approach that estimates the area of the protein with which the buffer gas can collide. It is the fastest of the current CCS simulation methods (Shvartsburg & Jarrold, 1996; Wyttenbach *et al.*, 1997, 2000; Shvartsburg *et al.*, 2001) and its predictions have been shown to be in good agreement (with an ~7% difference) with experimental values (Smith *et al.*, 2008; Uetrecht, Rose *et al.*, 2010). During simulations, the default Leeds settings were kept with the choice of helium as the buffer gas. In order to measure the discrepancy between the CCS values of the native complex (CCS_{Refc}) and the model (CCS_{Mod}), the following fit term was defined,

$$\text{CCS}_{\text{Fit}} = \text{abs}(\text{CCS}_{\text{Refc}} - \text{CCS}_{\text{Mod}}) / \text{CCS}_{\text{Refc}}. \quad (5)$$

2.3. Generation of synthetic SAXS curves

A commonly used method, CRY SOL (Svergun *et al.*, 1995), was used to simulate the SAXS curves of the native complexes and of the docking models. Synthetic SAXS data were simulated for the momentum-transfer (s) range 0.005–0.5 Å⁻¹ with default parameters, except for the maximum order of harmonics and the number of data points generated, which were set to 18 and 256, respectively. In order to measure the impact of SAXS data on docking in a realistic manner, noise was added to the data to mimic the experimental error, as in previously published examples (Williamson *et al.*, 2008; Blobel *et al.*, 2009; Bernadó, 2010).

(i) The error-to-intensity ratio (k_{exp}) was calculated from the data points of a good-quality experimental SAXS curve (measured from a sample at a concentration of >10 mg ml⁻¹; Tobias Madl, personal communication) by

$$k_{\text{exp}}(s) = \sigma_{\text{exp}}(s) / I_{\text{exp}}(s). \quad (6)$$

(ii) A second-order Gaussian error function was fitted to the distribution of k_{exp} as a function of the momentum transfer s in order to simulate the error-to-intensity ratio (k_{sim}),

$$k_{\text{sim}}(s) = 0.22 \exp \left\{ \left[-\frac{(s - 0.42)}{0.20} \right]^2 \right\} + 0.06 \exp \left\{ \left[-\frac{(s - 0.24)}{0.08} \right]^2 \right\}. \quad (7)$$

For a realistic error estimation, k_{sim} was randomly chosen within the confidence interval (95%) of the Gaussian error function (by using the *rand* function of MATLAB v.7.8.0.347; for details, see Supplementary Material¹). This ratio was then used to scale the intensity predicted by CRY SOL according to

$$\sigma_{\text{sim}}(s) = I_{\text{CRY SOL}}(s) k_{\text{sim}}(s). \quad (8)$$

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: IC5089). Services for accessing this material are described at the back of the journal.

To measure the fit between the SAXS curves of the native complex and the model, the discrepancy value χ (calculated by *CRY SOL*) was used,

$$\chi = \left\{ \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(s_i) - cI(s_i)}{\sigma(s_i)} \right]^2 \right\}^{1/2}, \quad (9)$$

where N_p is the total number of experimental data points, $\sigma(s_i)$ is the experimental error, c is a scaling factor, $I_e(s_i)$ is the experimental intensity and $I(s_i)$ is the calculated intensity (Svergun *et al.*, 1995). During fitting, the constant subtraction option was used to improve the fit (Svergun *et al.*, 1995; Petoukhov *et al.*, 2012).

2.4. Construction of the new scoring function

The CCS and SAXS fit terms defined above (equations 5 and 9) were individually incorporated into the standard *HADDOCK* score. Their optimum weight was determined by exploring a range between 0 and 500 and selecting the weight that resulted in the maximum number of benchmark cases with at least one hit (i-r.m.s.d. ≤ 4 Å; see §2.7) in their top 400. This optimization was performed using the *it0* structures generated for the 176 cases of Docking Benchmark 4.0 (Hwang *et al.*, 2010; see below) by running *HADDOCK* in *ab initio* mode. As a result, the following scoring functions were defined:

$$HADDOCK_{CCS} = HADDOCK \text{ score} + 50 \times CCS_{Fit}, \quad (10)$$

$$HADDOCK_{SAXS} = HADDOCK \text{ score} + 50 \times \chi. \quad (11)$$

In order to define the baseline performance, a random selection of models for various top X categories was performed. A near-native solution (or hit; see §2.7) in the top X can be obtained if there are $\geq 10\,000/X$ hits generated for that particular case. For each top X category, an enrichment factor was calculated by averaging over all complexes the ratio of the number of hits selected by the combined scoring function (either $HADDOCK_{CCS}$ or $HADDOCK_{SAXS}$) to the number of hits selected by random selection.

2.5. Docking benchmarks

The first benchmark used in this study is Docking Benchmark 4.0 (Hwang *et al.*, 2010), which consists of 176 unbound–unbound cases composed of 52 enzyme–inhibitor complexes, 25 antibody–antigen complexes and 99 other types of complexes. As SAXS and IM-MS are sensitive to mass change, the number of residues in the unbound states was matched to that in the native complex (*i.e.* missing parts in the structures of the complexes were removed from the structures of the free forms).

The second benchmark is composed of multimeric complexes with more than two components. It is an extension of the previously published benchmarks (Karaca *et al.*, 2010; Mashiach-Farkash *et al.*, 2011) and contains nine cases: four homotrimers, two homotetramers and three homopentamers with C_3 , D_2 and C_5 symmetry, respectively (Table 1). For six of them (PDB entries 1a3f, 1qu9, 1ous, 1vim, 1vpn and 1c4q),

Table 1

Properties of the symmetric multimer benchmark.

| PDB code | Complex type | No. of amino acids per monomer | Shape of each chain/anisotropy value |
|--|----------------|--------------------------------|--------------------------------------|
| 1js0† (Liu <i>et al.</i> , 2002) | C_3 trimer | 124 | Prolate/2.2 |
| 1qu9‡ (Volz, 1999) | C_3 trimer | 128 | Prolate/0.1 |
| 1a3f‡ (Segelke <i>et al.</i> , 1998) | C_3 trimer | 137 | Prolate/1.5 |
| 1urz† (Bressanelli <i>et al.</i> , 2004) | C_3 trimer | 400 | Prolate/19 |
| 1ous‡ (Loris <i>et al.</i> , 2003) | D_2 tetramer | 114 | Prolate/2.8 |
| 1vim‡ (Badger <i>et al.</i> , 2005) | D_2 tetramer | 200 | Prolate/0.7 |
| 1b0c† (Hamiaux <i>et al.</i> , 2000) | C_5 pentamer | 58 | Prolate/3.6 |
| 1c4q‡ (Chen <i>et al.</i> , 1999) | C_5 pentamer | 69 | Spherical/0.0 |
| 1vpn‡ (Stehle & Harrison, 1997) | C_5 pentamer | 289 | Prolate/4.4 |

† Docking was started with separated components of the crystal structure. ‡ Docking was started with free forms of the monomers.

docking was started with the separated components of the crystal structure ('bound docking') as the unbound coordinates were not available.

2.6. Classification of the cases according to their shape

We classified the shape anisotropy (SA) of the various systems in order to study the shape-dependency of the docking and scoring results. Since in a real case the structure of the complex is not known, this classification was performed on the largest component of each complex. If the partners were similar in size, the one with the largest shape anisotropy was considered. The shape anisotropy was calculated from the eigenvalues (λ_i) of the molecule's gyration tensor (\mathbf{G} ; Dima & Thirumalai, 2004; Rawat & Biswas, 2011),

$$SA = \frac{\prod_{i=1}^3 (\lambda_i - \bar{\lambda})}{\bar{\lambda}^3}, \quad (12)$$

where $\bar{\lambda}$ is the mean of the eigenvalues. For a rod-like prolate protein the shape anisotropy would be larger than zero (as $\lambda_1 \gg \lambda_2 \simeq \lambda_3$), for a disc-like oblate protein it would be smaller than zero (as $\lambda_1 \ll \lambda_2 \simeq \lambda_3$) and for an isotropic sphere-like protein it would be equal to zero (as $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$).

2.7. Assessment of the quality of the docking models

In the rigid-body docking stage, a (near-native) solution with i-r.m.s.d. ≤ 4.0 Å was considered to be a 'hit'. A benchmark case containing at least one hit within the top 400 was considered to be 'successful'. After water refinement, the models were evaluated based on the CAPRI criteria (Méndez *et al.*, 2003),

(i) acceptable prediction (one star): i-r.m.s.d. ≤ 4 Å or l-r.m.s.d. ≤ 10 Å;

(ii) good prediction (two stars): i-r.m.s.d. ≤ 2 Å or l-r.m.s.d. ≤ 5 Å;

(iii) high-quality prediction (three stars): i-r.m.s.d. ≤ 1 Å or l-r.m.s.d. ≤ 1 Å;

where i-r.m.s.d. refers to the interface r.m.s.d. and l-r.m.s.d. to the ligand r.m.s.d. calculated over the backbone atoms of the ligand (rigid component) after fitting on the backbone atoms

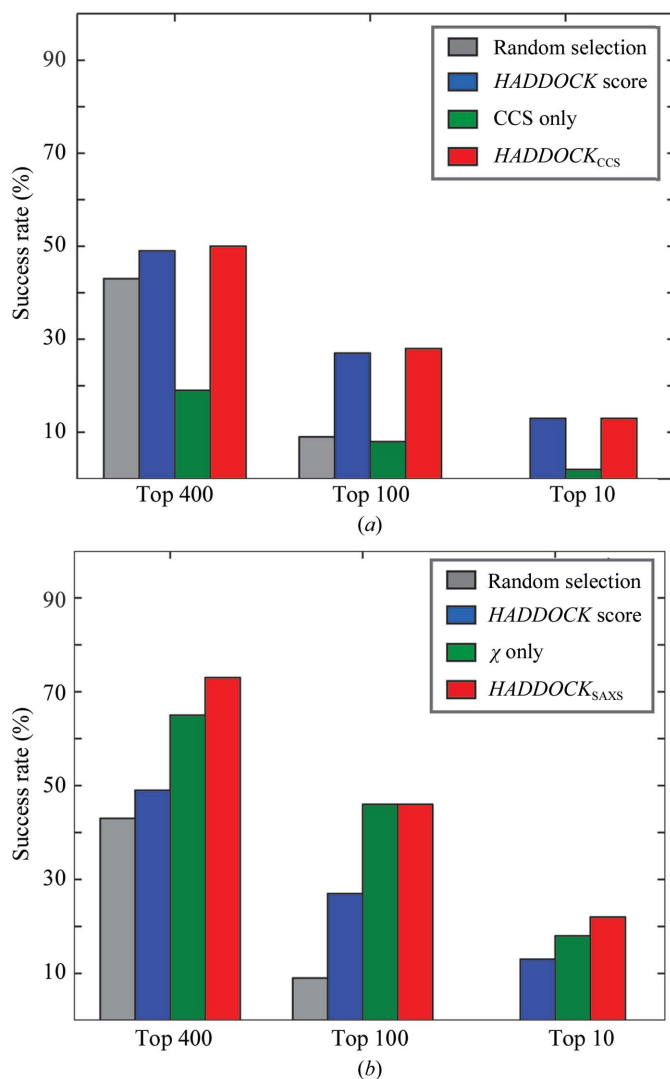


Figure 1

Success rates of the various scoring functions. The success rates were calculated on Docking Benchmark 4.0 cases for which *HADDOCK ab initio* could generate at least one hit (138 cases). For each top-ranking category, the scoring performance is presented in the order (a) random selection, *HADDOCK*, CCS_{Fit} only and *HADDOCK*_{CCS} scores and (b) random selection, *HADDOCK*, χ_{SAXS} only and *HADDOCK*_{SAXS} scores.

of the receptor. The cluster score was calculated by averaging the *HADDOCK* score of its top four members. A cluster was considered to be a near-native cluster if one of its top four members was of corresponding quality.

3. Results and discussion

3.1. The performance of *HADDOCK ab initio* is mainly limited by scoring rather than sampling

HADDOCK was run in *ab initio* mode to dock the unbound structures of Docking Benchmark 4.0 (see §§2.1 and 2.5). At the rigid-body docking stage at least one hit (a near-native model among the 10 000 rigid-body docking models) was obtained for 78% of the whole benchmark, corresponding to 138 cases out of 176. Within these 138 successful cases, at least one near-native model could be ranked in the top 400 for 49%

of the cases, in the top 100 for 27% of the cases and in the top 10 for 13% of the cases. These statistics reveal a large success-rate difference between sampling and scoring: considering the 138 successful cases, the overall success rate for the top 400 fell upon scoring by 50% (from 100 to 49%). This indicates that the success rate of *HADDOCK ab initio* is mainly limited by scoring rather than sampling (de Vries & Bonvin, 2011). A possible solution to overcome this problem is to incorporate external information into the scoring function (Brunger *et al.*, 1999; Robinson *et al.*, 2007; Alber *et al.*, 2008; Melquiond & Bonvin, 2010; Muradov *et al.*, 2010). For this purpose, we developed and tested two different scoring functions that integrate collision cross-section information from IM-MS and the scattering profile from SAXS experiments.

3.2. CCS does not discriminate between different docking poses of dimers

In order to evaluate the effect of CCS on the *HADDOCK* scoring function, the set of 138 successful complexes with at least one near-native solution were ranked according to (i) a random selection, (ii) the standard *HADDOCK* score (2), (iii) CCS_{Fit} only (5) and (iv) the optimized combined *HADDOCK*_{CCS} score (10). The corresponding scoring performances are depicted in Fig. 1(a). This analysis revealed that the success rate of the CCS_{Fit} -based ranking was either significantly worse than the random selection (as in the top 400 category) or similar to it. This implies that for this particular benchmark set the CCS data do not contain any information that would allow discrimination between different docking poses. Accordingly, the incorporation of the CCS_{Fit} term into the standard *HADDOCK* score did not result in any improvement.

Previously, Ruotolo and coworkers demonstrated that the simulated CCS values for a complex composed of 8 kDa subunits (~75 amino acids) start to diverge significantly from each other only at large subunit numbers (varying between eight and 12 subunits; see Fig. 1a of Ruotolo *et al.*, 2008). This observation was further supported by the findings of Politis and coworkers (Fig. 3a of Politis *et al.*, 2010) and Pukala and coworkers (Fig. 2 of Pukala *et al.*, 2009). In our case, the lack of sensitivity of the CCS_{Fit} term could be attributed to the fact that Docking Benchmark 4.0 only consists of dimers of rather limited size (between 7 and 90 kDa). To investigate the subunit- and mass-dependency of CCS, we scored the *it0* models of PDB entries 1vim and 1vpn (Table 1), which are the two largest complexes of our multimer benchmark, using CCS_{Fit} and *HADDOCK*_{CCS} terms. In the case of 1vim, both CCS_{Fit} and *HADDOCK*_{CCS} ranked all of the near-native solutions within the top 10, whereas for 1vpn they could not rank any hit even within the top 400. Although these results were not conclusive, they suggest that the CCS data do carry potential for the study of large macromolecular complexes.

3.3. *HADDOCK*_{SAXS} improves the scoring of rigid-body docking models

In order to assess the scoring ability of *HADDOCK*_{SAXS}, the success rates of (i) a random selection, (ii) the standard

Table 2

Enrichment factors of various scoring functions compared with a random selection.

The enrichment factor is calculated as the ratio of the number of selected near-native solutions (for the given top category) compared with a random selection of the same number of models. (This measure could only be calculated for the top 400 and top 100 categories, as for the lower categories, owing to the limited number of hits, a random selection was unable to select any hit.)

| | <i>HADDOCK</i> | χ_{SAXS} only | <i>HADDOCK</i> _{SAXS} |
|---------|----------------|---------------------------|--------------------------------|
| Top 400 | 4.5 | 5.7 | 7.9 |
| Top 100 | 9.4 | 14.5 | 16.8 |

HADDOCK score, (iii) the SAXS only χ -score (χ_{SAXS}) and (iv) the combined *HADDOCK*_{SAXS} score were compared (Fig. 1*b*). The success-rate statistics revealed that the χ_{SAXS} score has a significant discriminative ability (even higher than the standard *HADDOCK* score). However, among all of the scoring functions, the combination of χ_{SAXS} and *HADDOCK* score performed the best, especially for the top 400 category (Fig. 1*b*), with a 24% improvement in success rate (from 49 to 73% measured over the set of 138 complexes having at least one hit). We also calculated the enrichment factor of each scoring function compared with a random selection (Table 2; see §2.4). On average *HADDOCK*_{SAXS} scoring could enrich the number of hits selected per case by eightfold for the top 400 category and by 17-fold for the top 100 category, showing the best performance and almost doubling the performance of the *HADDOCK* score.

It has already been demonstrated that SAXS can distinguish better between different conformations of the same complex if the anisotropy of the constituents of the complex is high; namely, if their shape is far from spherical (Pons *et al.*, 2010; Madl *et al.*, 2011; Gabel, 2012). To explore this point, we classified the successful set of 138 complexes into three sets: prolate (90 cases), oblate (25 cases) and spherical (23 cases). This classification was based on the largest constituent of each complex and not the complex itself, so that this measure can potentially be used as a predictor of the impact of SAXS scoring in real cases where the structure of the complex is still unknown. The performance of the various scoring functions was analyzed for each class separately, revealing that χ_{SAXS} scoring performs significantly better for anisotropic (66% for prolate, 72% for oblate) complexes than for spherical (57%) complexes (Fig. 2). The same can be observed for the combined *HADDOCK*_{SAXS} score, although with an increased overall performance (74% for prolate, 76% for oblate and 65% for spherical).

When the same set was classified according to biochemical function, the most pronounced increase in scoring accuracy was seen for antibody–antigen complexes: *HADDOCK*_{SAXS} could rank at least one hit in the top 400 for all of the antibody–antigen cases (100% success rate), whereas the regular *HADDOCK* score could only perform the same for half of them. This impressive improvement is related to the shape-dependency of SAXS scoring since all of the antibody–antigen complexes are highly anisotropic (prolate). Compared with

the antibody–antigen complexes, the success rate of *HADDOCK*_{SAXS} decreased to 76% for enzyme–inhibitor complexes and to 65% for the ‘other’ category.

3.4. Impact of flexible refinement

The top 400 models that were re-ranked according to the *HADDOCK*_{SAXS} score were subjected to *HADDOCK*’s two-step semi-flexible refinement protocol (it1 and water), which typically allows the modelling of modest interfacial (~1–2.5 Å) side-chain and backbone conformational rearrangements (de Vries *et al.*, 2007). This flexible refinement resulted in a success-rate increase of 2% compared with the top 400 of the initial rigid-body docking (73%; see Supplementary Fig. S2). Among the 103 cases with at least one hit, 56% contained one-star quality, 42% contained two-star quality solutions and only 2% contained three-star quality solutions. Comparison of the performance of *HADDOCK* and *HADDOCK*_{SAXS} revealed that *HADDOCK*_{SAXS} improves the success rate of the high-ranking top categories (Fig. 3*a*). All of the above statistics were based on single-structure ranking; however, in the final stage of *HADDOCK* the water-

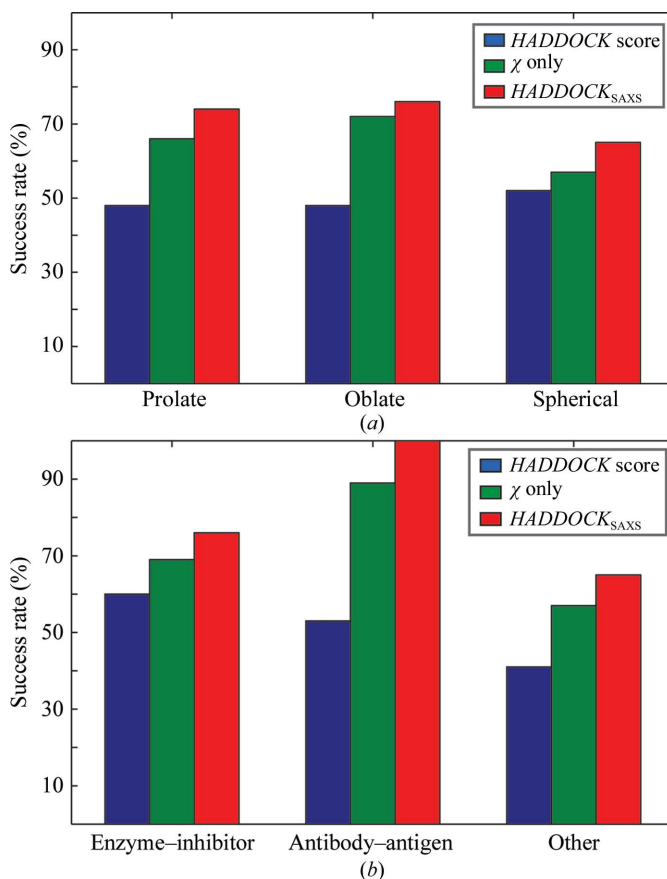


Figure 2

The success rates of various scoring functions classified according to the shape and function of the complexes. The success rates presented here are for the top 400 category calculated on the 138 successful cases of the benchmark. The successful docking decoys are classified according to (a) their shape determinants (prolate, 90 cases; oblate, 25 cases; spherical, 23 cases) and (b) their biochemical functions (enzyme–inhibitor, 45 cases; antibody–antigen, 19 cases; other, 74 cases).

Table 3

Rigid-body docking (*it0*) performance of *HADDOCK* and the scoring performance of *HADDOCK*_{SAXS} on the symmetric multimer benchmark.

The second column refers to the total number of hits generated, the third column to the number of hits ranked within the top 400 using the regular *HADDOCK* score (4) and the last column to the number of hits ranked using the combined *HADDOCK*_{SAXS} score (11). A solution is considered to be a hit if it is within 4.0 Å i-r.m.s.d. or 10 Å l-r.m.s.d. of the native solution (see §2.7). The last column corresponds to the enrichment factor, calculated as the ratio of the number of hits selected by combined *HADDOCK*_{SAXS} scoring to the number of hits selected by regular *HADDOCK* scoring.

| PDB code | No. of hits generated | No. of hits ranked by <i>HADDOCK</i> (% of total hits) | No. of hits ranked by <i>HADDOCK</i> _{SAXS} (% of total hits) | Enrichment factor compared with the standard <i>HADDOCK</i> score |
|----------|-----------------------|--|--|---|
| 1js0 | 90 | 4 (4%) | 88 (98%) | 24.5 |
| 1qu9 | 409 | 368 (92%) | 366 (91%) | 1.0 |
| 1a3f | 418 | 112 (28%) | 273 (68%) | 2.4 |
| 1urz | — | — | — | — |
| 1ous | 1 | 0 | 0 | 0 |
| 1vim | 8 | 8 (100%) | 8 (100%) | 1 |
| 1b0c | 66 | 0 | 30 (45%) | ∞ |
| 1c4q | 231 | 217 (94%) | 212 (92%) | 1.4 |
| 1vpn | 8 | 5 (62%) | 5 (62%) | 1 |

refined solutions are clustered and the putative binding mode is selected according to cluster statistics (see §2.7) rather than individual ranking. If the same set is considered, the success rate of *HADDOCK*_{SAXS} for the top 10 category is 45% in the case of individual ranking, whereas this success rate increases to 54% for the ranking based on cluster statistics. Cluster-based scoring is thus more efficient in fishing out the relevant conformation than individual ranking (Fig. 3*b*). Note again that these numbers refer to the scoring performance of *HADDOCK* run in *ab initio* mode without any additional information apart from the SAXS term used in scoring. As an alternative to *HADDOCK ab initio*, docking can also be driven by bioinformatic predictions. In this case, the scoring performance on a smaller benchmark was shown to double from 27 to 58% after water refinement (de Vries & Bonvin, 2011). The combination of the two (the SAXS scoring of bioinformatics-driven docking predictions) still remains to be investigated.

3.5. *HADDOCK*_{SAXS} improves the ranking of symmetric multimers, especially for anisotropic multimers

In order to assess the impact of SAXS scoring on larger systems, we run *HADDOCK* in *ab initio* mode on a symmetric multimer benchmark of nine complexes composed of trimers to pentamers (Table 1). This benchmark is an extension of our previously reported multibody docking work (Karaca *et al.*, 2010). *HADDOCK* generated a substantial number of hits for five cases (Table 3). However, their number decreased significantly for larger system sizes (as for PDB entries 1ous, 1vim and 1vpn; Table 3) or in cases where significant conformational changes are observed upon binding, such as PDB entry 1urz, in which the induced backbone conformational

change is 4.3 Å. For these, very few (or none in the latter case) near-native solutions could be sampled.

Irrespective of the docking difficulty, *HADDOCK*_{SAXS} could rank at least one hit within the top 400 for seven of the eight cases, which translates into a success rate of 88%. As observed previously for dimers, the scoring performance is also significantly better in the case of multimers for systems with high anisotropy (PDB entries 1a3f, 1js0 and 1b0c) than those with low anisotropy (PDB entries 1qu9 and 1c4q) (Table 1). At the end of the water refinement, a top-ranking high-accuracy three-star solution could be generated for four cases, a medium-quality two-star solution for two cases and an acceptable one-star solution for one case (Table 4, Fig. 4). Moreover, all of these high-quality solutions were populated in top-ranking clusters.

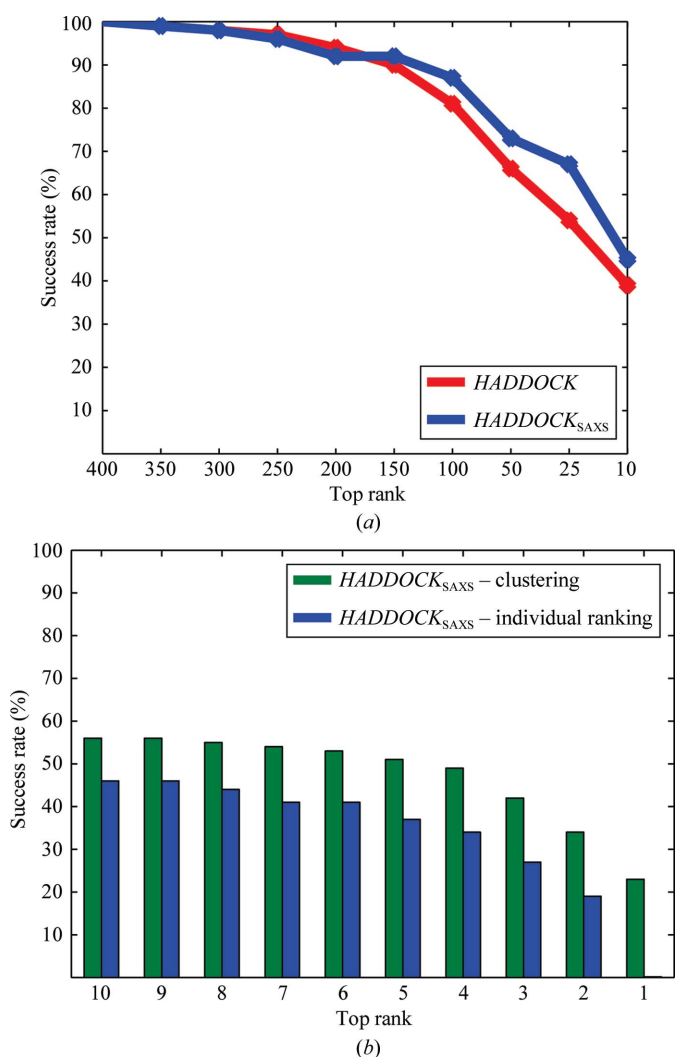


Figure 3
(a) Comparison of the performance of *HADDOCK* scoring (blue) and *HADDOCK*_{SAXS} scoring (red) at the end of the water stage. The success rates were calculated for different top-ranking categories (between top 400 and top 10). (b) Comparison of the performance of individual-structure (blue) and cluster-based ranking (green) at the end of the water stage. The illustrated statistics were calculated for different top-ranking categories using *HADDOCK*_{SAXS}. In both cases, the success rates were calculated over the Docking Benchmark 4.0 cases for which *HADDOCK* could generate at least one hit *ab initio* (103 cases).

Table 4

Scoring performance of *HADDOCK*_{SAXS} on the symmetric multimer benchmark at the end of the flexible refinement (water) stage.

| | Single-structure scoring (quality†/rank) | Cluster-based scoring (quality†/rank) |
|------|---|--|
| 1js0 | */1, **/14 | */1, **/2 |
| 1qu9 | ***/1 | ***/1 |
| 1a3f | ***/1 | ***/1 |
| 1urz | — | — |
| 1ous | — | — |
| 1vim | */1 | */1 |
| 1b0c | */7 | */1, **/2 |
| 1c4q | ***/1 | ***/1 |
| 1vpn | ***/1 | Not clustered |

† For the definition of quality, see §2.7.

3.6. Application examples with experimental SAXS profiles

We tested the performance of *HADDOCK*_{SAXS} on three cases (PDB entries 3k3k, 2r15 and 1o6s) for which experimental SAXS profiles were available (Table 5). The experimental SAXS curve of 3k3k was taken from the BIOSIS

database (Hura *et al.*, 2009), while the other two were kindly provided by Dmitri Svergun (Pons *et al.*, 2010). *HADDOCK ab initio* and *HADDOCK*_{SAXS} were used for docking and scoring, respectively, in the same way as described above. The data and results are summarized in Tables 5 and 6.

PDB entry 3k3k is an abscisic acid receptor that forms an asymmetric homodimer (Nishimura *et al.*, 2009). Docking of 3k3k was started from the isolated chains of 3k3k. At the end of it0, *HADDOCK* could rank 63% of the hits generated within the top 400, whereas this decreased to 44% upon *HADDOCK*_{SAXS} scoring (Table 6a). This decrease in accuracy arose from the fact that the χ _{SAXS} value calculated for a degenerate binding mode was smaller than for the native solution (Figs. 5a and 5b). Recently, Gabel *et al.* (2012) remarked that when the monomers of a homodimer are oriented in a ‘side-by-side’ fashion, as in this case, the SAXS curves simulated from different side-by-side orientations of the same complex fit equally well to the experimental curve. This degeneracy issue is less problematic when the monomers are arranged in an ‘end-to-end’ (rod-like) fashion. Despite

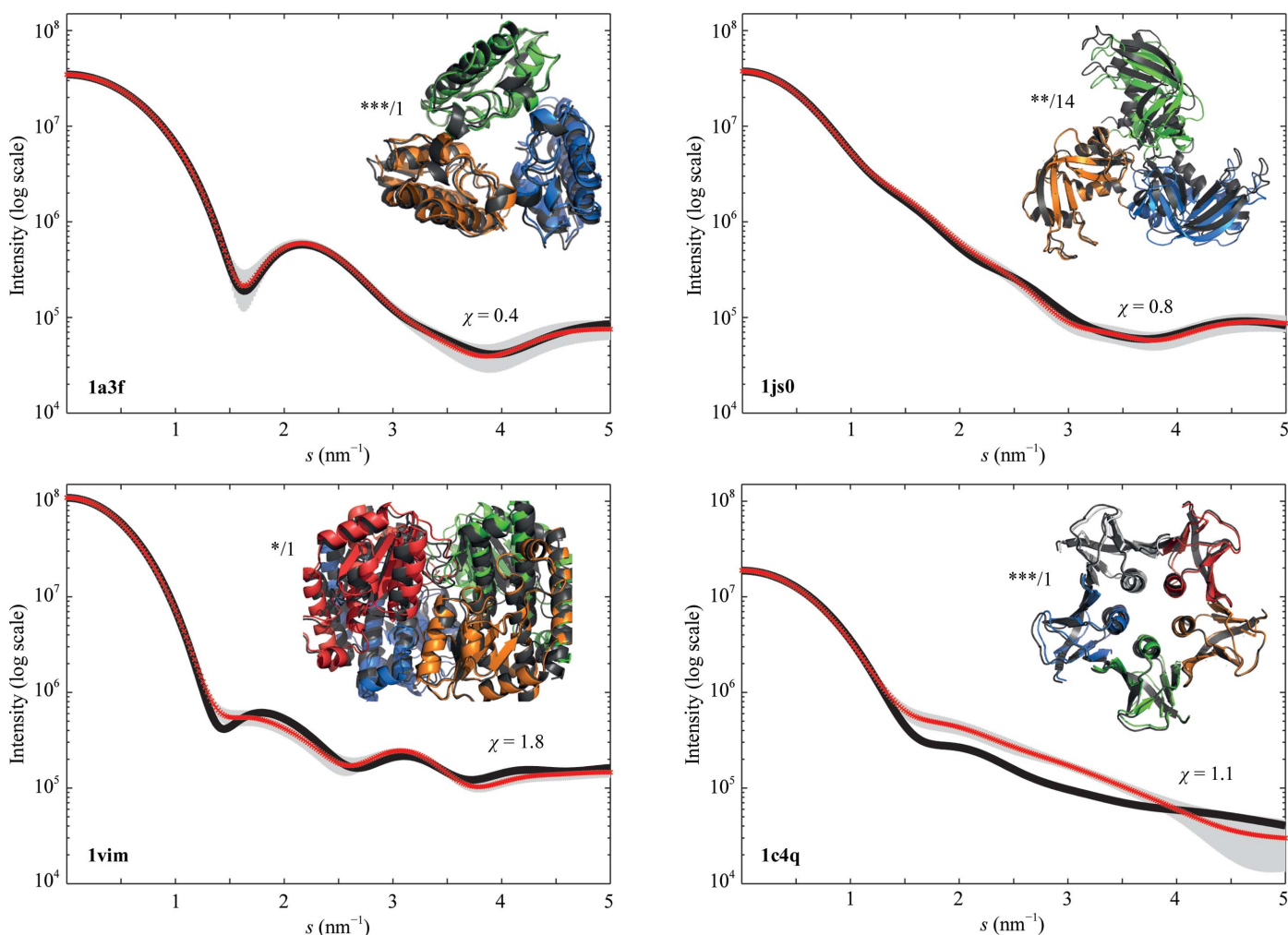


Figure 4

Illustration of the best models ranked by *HADDOCK*_{SAXS}. For each case, the fit between the simulated scattering curve of the reference crystal structure (red crosses with corresponding errors shown as grey lines) and that of the best *HADDOCK* model (black continuous lines) is illustrated. The structure of the respective best *HADDOCK* model (coloured) superimposed on top of the crystal structure (black) is depicted, together with their associated quality and rank.

Table 5

Test benchmark with experimental SAXS curves.

| PDB code | Complex type | No. of amino acids per monomer | Shape of each chain/anisotropy value |
|--|--------------|--------------------------------|--------------------------------------|
| 3k3k [†] (Nishimura <i>et al.</i> , 2009) | Homodimer | 211 | Prolate/0.8 |
| 2r15 [†] (Pinotsis <i>et al.</i> , 2008) | Homodimer | 212 | Prolate/26.8 |
| 1o6s [‡] (Schubert <i>et al.</i> , 2002) | Dimer | 466/105 | Prolate/11.3 |

[†] Docking was started with separated components of the crystal structure. [‡] The docking was started with free forms of the monomers: 1o6t (Schubert *et al.*, 2002) and chain *B* of 1ff5 (Pertz *et al.*, 1999).

the degeneracy problem, the contribution of the *HADDOCK* score ensured that a biologically meaningful high-accuracy cluster was ranked at top at the end of the flexible refinement (Fig. 5*b*).

PDB entry 2r15 is a complex of an end-to-end myomesin fragment composed of two highly prolate domains (My12 and My13; Table 5; Pinotsis *et al.*, 2008). My12 was used as the starting structure for both chains. During rigid-body docking, only two near-native models could be generated, which were successfully selected by *HADDOCK*_{SAXS} for further refinement. At the end of the flexible refinement one of these near-native solutions was ranked within the top 10 (Table 6*b*). This scoring success (the successful selection of two out of 10 000 models) could be achieved since the monomers of 2r15 are highly prolate and are oriented in an end-to-end fashion in the complex (Fig. 5*c*).

The last example, PDB entry 1o6s, is a prolate complex of internalin (In1A) with the N-terminal domain of human E-cadherin (hEC1; Table 5; Schubert *et al.*, 2002). Docking was started with the free forms of In1A (PDB entry 1o6t; Schubert *et al.*, 2002) and hEC1 (PDB entry 1ff5, chain *B*; Pertz *et al.*, 1999). During rigid-body docking 25 near-native models were generated, none of which were ranked within the top 400 using the regular *HADDOCK* score. In contrast, *HADDOCK*_{SAXS} successfully selected 80% of these near-native models for further refinement, which ended up in the top-ranking one-star quality cluster at the end of flexible refinement (Table 6, Fig. 5*d*).

3.7. Comparison to related methods

SAXS-based rigid-body modelling can be a promising alternative to docking in the presence of experimental SAXS data. In order to test this, we ran *SASREF* (Petoukhov & Svergun, 2005), a commonly used rigid-body modelling tool, on the unbound structures of Docking Benchmark 4.0 after randomizing their orientation (an important step since the unbound structures of the benchmark are already superimposed onto their respective complex structures). In each case, ten *SASREF* models were generated by following a previously described protocol (Pons *et al.*, 2010). *SASREF* could produce at least one hit for 12 cases out of 176, leading to a success rate of 7%. If only the 138 cases for which *HADDOCK* could generate at least one acceptable solution are considered, the success rate of *SASREF* for the top 10 category stays the same (7%) compared with 41% in the case

Table 6Scoring performance of *HADDOCK*_{SAXS}.

(*a*) Rigid-body docking (it0) performance of *HADDOCK* and scoring performance of *HADDOCK*_{SAXS} on the benchmark with experimental SAXS curves. The second column refers to the total number of hits generated, the third column to the number of hits ranked within the top 400 by *HADDOCK* and the last column to the number of hits ranked by *HADDOCK*_{SAXS}.

| PDB code | No. of hits generated out of 10 000 models | No. of hits ranked by <i>HADDOCK</i> in the top 400 (% of total hits) | No. of hits ranked by <i>HADDOCK</i> _{SAXS} in the top 400 (% of total hits) |
|----------|--|---|---|
| 3k3k | 117 | 74 (63%) | 52 (44%) |
| 2r15 | 2 | 0 (0%) | 2 (100%) |
| 1o6s | 25 | 0 (0%) | 20 (80%) |

(*b*) Scoring performance of *HADDOCK*_{SAXS} at the end of the flexible refinement (water) stage.

| PDB code | Single-structure scoring (quality [†] /rank) | Cluster-based scoring (quality [†] /rank) |
|----------|---|--|
| 3k3k | ***/17 | ***/1 |
| 2r15 | */6, **/12 | — |
| 1o6s | */32 | */1 |

[†] For the definition of quality, see §2.7.

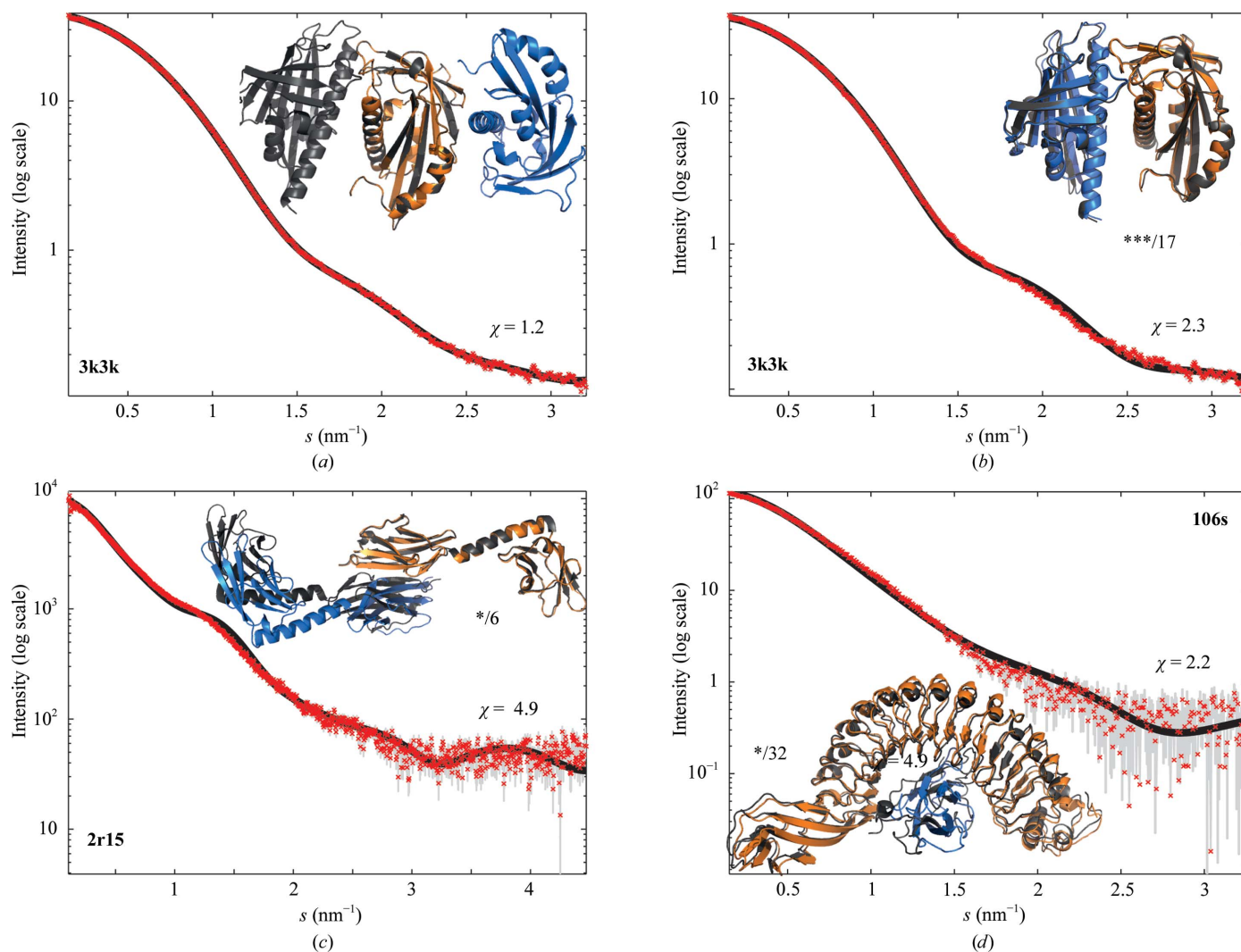
of *HADDOCK*_{SAXS} (the outcome of the water-clustering statistics).

Other methods that rescore docking models based on combined scoring (a docking score and an SAXS term) include *pyDockSAXS* (Pons *et al.*, 2010) and *FoXSDock* (Schneidman-Duhovny *et al.*, 2010, 2012). The performance of *pyDockSAXS* has been demonstrated using 70 cases in Docking Benchmark 2.0 (Mintseris *et al.*, 2005), a subset of the benchmark used by *FoXSDock* and *HADDOCK*_{SAXS}. With this common benchmark, the success rates of each method were determined to be (in increasing order) 43% for *pyDockSAXS*, 48% for *HADDOCK*_{SAXS} and 63% for *FoXSDock*. Here, it is important to note that these numbers also depend on the number of hits generated per case. A more reliable comparison could be made if the same set of structures was scored using these three scoring functions.

4. Conclusions

In this work, we have demonstrated that in the absence of external information the accuracy of docking models obtained by running *HADDOCK* in *ab initio* mode is limited by scoring rather than by sampling. However, the inaccurate nature of conventional scoring functions can be improved by incorporating information-based terms. Within this context, we have evaluated the impact of integrating CCS from IM-MS and the scattering profile from SAXS experiments into the *HADDOCK* scoring function.

Our analysis has revealed that the information content of CCS data is of too low resolution to distinguish correct models in a large docking decoy set composed of dimers. Despite this failure to distinguish near native docking poses for rather small dimeric assemblies, in principle CCS data have the potential to add useful information for large macromolecular


Figure 5

The performance of $HADDOCK_{SAXS}$ on cases with experimental SAXS profiles. For each case, the fit between the experimental scattering curve (red crosses with corresponding errors in grey) and the simulated curve of the $HADDOCK$ model (black lines) is depicted together with the superimposed $HADDOCK$ model (coloured) on top of the crystal structure (black). Quality and rank are indicated next to each model. (a, b) When the monomers of a homodimer are oriented in a side-by-side fashion, as in the 3k3k case, the SAXS curve simulated from a degenerate binding mode (a) could fit better to the reference SAXS curve than the correct binding mode (b). (c) For 2r15, although only two near-native models were sampled during it0, one of them could be successfully ranked within the top 10 with $HADDOCK_{SAXS}$ after water refinement. (d) $HADDOCK_{SAXS}$ could select 80% of the near-native 106s models for further refinement, which corresponds to a top-ranking one-star quality cluster after water refinement.

complexes. Indeed, various examples in the literature have underpinned the size-dependency of CCS, indicating that CCS-based scoring starts to become discriminative at high subunit numbers (Ruotolo *et al.*, 2008; Pukala *et al.*, 2009; Politis *et al.*, 2010). In order to test this, we re-ranked the rigid-body docking models of the two largest assemblies of our multimer benchmark: PDB entries 1vim and 1vpn (Table 1). As a result, CCS_{Fit} and $HADDOCK_{CCS}$ worked well for 1vim (both of them ranked all of the near-native solutions at the top), whereas for 1vpn they could not select any hit for further refinement. This analysis suggests that there is room for further improvement, especially for large assemblies.

In contrast to the CCS data, the results presented here for SAXS scoring reveal a significant improvement in performance, especially when combined with an energy-based

scoring function such as the classical $HADDOCK$ score. However, the scoring performance depends on the shape and arrangement of the complex. The SAXS scoring term should preferentially be combined with conventional energy-based scoring functions in order to ensure the selection of physically relevant interfaces. Any additional source of information that can be included in the sampling phase will also increase the coverage of near-native solutions. $HADDOCK$ is an excellent candidate for this purpose, with its built-in ability to incorporate various sources of data that can provide information on, for example, binding interfaces, relative orientation of molecules or residue–residue distances. The $HADDOCK_{CCS}$ and $HADDOCK_{SAXS}$ scoring functions and associated scripts will be part of a future release of $HADDOCK2.2$ and are freely available upon request.

This work was supported by the Netherlands Organization for Scientific Research (NWO; VICI grant 700.56.442 to AMJJB) and the European Community (FP7 FP7 e-Infrastructure 'WeNMR' project, grant No. 21301). Support by the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands (*via* the Dutch BiG Grid project), Portugal, Spain, UK, South Africa and Taiwan and the Latin America GRID infrastructure *via* the Gisela project is acknowledged for the use of web portals, computing and storage facilities. We thank Dr Tobias Madl (Helmholtz Zentrum München, Technische Universität München) and Dr Dmitri Svergun (EMBL Hamburg) for providing experimental SAXS data, Dr Tobias Madl and Dr Ioana Barbu (Utrecht University) for helpful discussions about integrating SAXS and CCS data into modelling, respectively, and Dr Adrien Melquiond, Dr Christophe Schmitz and Mikael Trellet (Utrecht University) for their help in the development of scoring optimization.

References

- Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. (2008). *Annu. Rev. Biochem.* **77**, 443–477.
- Badger, J. *et al.* (2005). *Proteins*, **60**, 787–796.
- Benesch, J. L. & Ruotolo, B. T. (2011). *Curr. Opin. Struct. Biol.* **21**, 641–649.
- Bernadó, P. (2010). *Eur. Biophys. J.* **39**, 769–780.
- Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condrón, M. M., Bitan, G., Teplow, D. B., Shea, J. E., Ruotolo, B. T., Robinson, C. V. & Bowers, M. T. (2009). *Nature Chem.* **1**, 326–331.
- Blobel, J., Bernadó, P., Svergun, D. I., Tauler, R. & Pons, M. (2009). *J. Am. Chem. Soc.* **131**, 4378–4386.
- Bonvin, A. M. (2006). *Curr. Opin. Struct. Biol.* **16**, 194–200.
- Bressanelli, S., Stiasny, K., Allison, S. L., Stura, E. A., Duquerroy, S., Lescar, J., Heinz, F. X. & Rey, F. A. (2004). *EMBO J.* **23**, 728–738.
- Brunger, A. T., Adams, P. D. & Rice, L. M. (1999). *Prog. Biophys. Mol. Biol.* **72**, 135–155.
- Brunger, A. T., Strop, P., Vrljic, M., Chu, S. & Weninger, K. R. (2011). *J. Struct. Biol.* **173**, 497–505.
- Buey, R. M., Chacón, P., Andreu, J. M. & Fernando Díaz, J. (2009). *Applications of Synchrotron Light to Scattering and Diffraction in Materials and Life Sciences*, edited by T. A. Ezquerra, M. C. García-Gutiérrez, A. Nogales & M. Gomez, pp. 245–263. Berlin, Heidelberg: Springer.
- Chen, Y. W., Stott, K. & Perutz, M. F. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 1257–1261.
- Covaceuszach, S., Cassetta, A., Konarev, P. V., Gonfloni, S., Rudolph, R., Svergun, D. I., Lamba, D. & Cattaneo, A. (2008). *J. Mol. Biol.* **381**, 881–896.
- D'Abramo, M., Meyer, T., Bernadó, P., Pons, C., Recio, J. F. & Orozco, M. (2009). *J. Chem. Theory Comput.* **5**, 3129–3137.
- Dima, R. I. & Thirumalai, D. (2004). *J. Phys. Chem. B*, **108**, 6564–6570.
- Dominguez, C., Boelens, R. & Bonvin, A. M. (2003). *J. Am. Chem. Soc.* **125**, 1731–1737.
- Fernández-Recio, J., Totrov, M. & Abagyan, R. (2004). *J. Mol. Biol.* **335**, 843–865.
- Filgueira de Azevedo, W., dos Santos, G. C., dos Santos, D. M., Olivieri, J. R., Canduri, F., Silva, R. G., Basso, L. A., Renard, G., da Fonseca, I. O., Mendes, M. A., Palma, M. S. & Santos, D. S. (2003). *Biochem. Biophys. Res. Commun.* **309**, 923–928.
- Förster, F., Webb, B., Krukenberg, K. A., Tsuruta, H., Agard, D. A. & Sali, A. (2008). *J. Mol. Biol.* **382**, 1089–1106.
- Gabel, F. (2012). *Eur. Biophys. J.* **41**, 1–11.
- García-García, J., Bonet, J., Guney, E., Fornes, O., Planas, J. & Oliva, B. (2012). *Mol. Informatics*, **31**, 342–362.
- Grishae, A., Wu, J., Trehella, J. & Bax, A. (2005). *J. Am. Chem. Soc.* **127**, 16621–16628.
- Hamiaux, C., Pérez, J., Prangé, T., Veessler, S., Riès-Kautt, M. & Vachette, P. (2000). *J. Mol. Biol.* **297**, 697–712.
- Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L. II, Tsutakawa, S. E., Jenney, F. E. Jr, Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W. & Tainer, J. A. (2009). *Nature Methods*, **6**, 606–612.
- Hwang, H., Vreven, T., Janin, J. & Weng, Z. (2010). *Proteins*, **78**, 3111–3114.
- Jorgensen, W. L. & Tirado-Rives, J. (1988). *J. Am. Chem. Soc.* **110**, 1657–1666.
- Jurneczko, E. & Barran, P. E. (2011). *Analyst*, **136**, 20–28.
- Karaca, E., Melquiond, A. S., de Vries, S. J., Kastrius, P. L. & Bonvin, A. M. (2010). *Mol. Cell. Proteomics*, **9**, 1784–1794.
- Lander, G. C., Saibil, H. R. & Nogales, E. (2012). *Curr. Opin. Struct. Biol.* **22**, 627–635.
- Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2002). *Protein Sci.* **11**, 371–380.
- Loris, R., Tielker, D., Jaeger, K. E. & Wyns, L. (2003). *J. Mol. Biol.* **331**, 861–870.
- Madl, T., Gabel, F. & Sattler, M. (2011). *J. Struct. Biol.* **173**, 472–482.
- Mashiach-Farkash, E., Nussinov, R. & Wolfson, H. J. (2011). *Proteins*, **79**, 2607–2623.
- Melquiond, A. S. J. & Bonvin, A. M. J. J. (2010). *Protein-Protein Complexes: Analysis, Modeling and Drug Design*, edited by M. Zacharias, pp. 183–209. London: Imperial College Press.
- Méndez, R., Leplae, R., De Maria, L. & Wodak, S. J. (2003). *Proteins*, **52**, 51–67.
- Mertens, H. D. & Svergun, D. I. (2010). *J. Struct. Biol.* **172**, 128–141.
- Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R. & Weng, Z. (2005). *Proteins*, **69**, 511–520.
- Muradov, D., Kobe, B., Dixon, E. N. & Huber, T. (2010). *Introduction to Protein Structure Prediction: Methods and Algorithms*, edited by H. Rangwala & G. Karypis, pp. 265–277. Hoboken: John Wiley & Sons.
- Nishimura, N., Hitomi, K., Arvai, A. S., Rambo, R. P., Hitomi, C., Cutler, S. R., Schroeder, J. I. & Getzoff, E. D. (2009). *Science*, **326**, 1373–1379.
- Pertz, O., Bozic, D., Koch, A. W., Fauser, C., Brancaccio, A. & Engel, J. (1999). *EMBO J.* **18**, 1738–1747.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.
- Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237–1250.
- Pinotsis, N., Lange, S., Perriard, J.-C., Svergun, D. I. & Wilmanns, M. (2008). *EMBO J.* **27**, 253–264.
- Politis, A., Park, A. Y., Hyung, S.-J., Barsky, D., Ruotolo, B. T. & Robinson, C. V. (2010). *PLoS One*, **5**, e12080.
- Pons, C., D'Abramo, M., Svergun, D. I., Orozco, M., Bernadó, P. & Fernández-Recio, J. (2010). *J. Mol. Biol.* **403**, 217–230.
- Pukala, T. L., Ruotolo, B. T., Zhou, M., Politis, A., Stefanescu, R., Leary, J. A. & Robinson, C. V. (2009). *Structure*, **17**, 1235–1243.
- Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.
- Rappsilber, J. (2011). *J. Struct. Biol.* **173**, 530–540.
- Rawat, N. & Biswas, P. (2011). *Phys. Chem. Chem. Phys.* **13**, 9632–9643.
- Robinson, C. V., Sali, A. & Baumeister, W. (2007). *Nature (London)*, **450**, 973–982.
- Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S.-J. & Robinson, C. V. (2008). *Nature Protoc.* **3**, 1139–1152.
- Ruotolo, B., Hyung, S., Robinson, P., Giles, K., Bateman, R. & Robinson, C. (2007). *Angew. Chem. Int. Ed.* **46**, 8001–8004.
- Scarff, C. A., Thalassinou, K., Hilton, G. R. & Scrivens, J. H. (2008). *Rapid Commun. Mass Spectrom.* **22**, 3297–3304.

- Schneidman-Duhovny, D., Hammel, M. & Sali, A. (2010). *J. Struct. Biol.* **173**, 461–471.
- Schneidman-Duhovny, D. *et al.* (2012). *Bioinformatics*, **28**, 3282–3289.
- Schubert, W. D., Urbanke, C., Ziehm, T., Beier, V., Machner, M. P., Domann, E., Wehland, J., Chakraborty, T. & Heinz, D. W. (2002). *Cell*, **111**, 825–836.
- Segelke, B. W., Nguyen, D., Chee, R., Xuong, N. H. & Dennis, E. A. (1998). *J. Mol. Biol.* **279**, 223–232.
- Shvartsburg, A. A., Hudgins, R. R., Dugourd, P. & Jarrold, M. F. (2001). *Chem. Soc. Rev.* **30**, 26–35.
- Shvartsburg, A. A. & Jarrold, M. F. (1996). *Chem. Phys. Lett.* **261**, 86–91.
- Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. & Ashcroft, A. E. (2008). *Eur. J. Mass Spectrom.* **15**, 113–130.
- Sondermann, H., Nagar, B., Bar-Sagi, D. & Kuriyan, J. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 16632–16637.
- Stehle, T. & Harrison, S. C. (1997). *EMBO J.* **16**, 5139–5148.
- Steinhoff, H. J. (2004). *Biol. Chem.* **385**, 913–920.
- Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
- Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E. & Heck, A. J. (2010). *Nature Chem.* **3**, 126–132.
- Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K. & Heck, A. J. (2010). *Chem. Soc. Rev.* **39**, 1633–1655.
- Volz, K. (1999). *Protein Sci.* **8**, 2428–2437.
- Vries, S. J. de & Bonvin, A. M. J. J. (2011). *PLoS One*, **6**, e17695.
- Vries, S. J. de, van Dijk, M. & Bonvin, A. M. J. J. (2010). *Nature Protoc.* **5**, 883–897.
- Vries, S. J. de, van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. & Bonvin, A. M. J. J. (2007). *Proteins*, **69**, 726–733.
- Wang, X., Lee, H.-W., Liu, Y. & Prestegard, J. H. (2011). *J. Struct. Biol.* **173**, 515–529.
- Williamson, T. E., Craig, B. A., Kondrashkina, E., Bailey-Kellogg, C. & Friedman, A. M. (2008). *Biophys. J.* **94**, 4906–4923.
- Wytenbach, T., Helden, G., Batka, J. J., Carlat, D. & Bowers, M. T. (1997). *J. Am. Soc. Mass Spectrom.* **8**, 275–282.
- Wytenbach, T., Witt, M. & Bowers, M. T. (2000). *J. Am. Chem. Soc.* **122**, 3458–3464.
- Zhou, M. & Robinson, C. V. (2010). *Trends Biochem. Sci.* **35**, 522–529.