



SEVENTH FRAMEWORK
PROGRAMME

Research Infrastructures

Deliverable 7.5

A grid-enabled SAXS/NMR/docking refinement protocol



Grant Agreement number:	RI-261572
Project acronym:	WeNMR
Project title:	WeNMR, A worldwide e-Infrastructure for NMR and structural biology
Deliverable type and status	Prototype – Public
Due Date	June, 2013 (M32)
Delivery Date	August, 2013 (M34)
Project coordinator name, title:	Prof. Alexandre M.J.J. Bonvin
Organization:	Utrecht University, The Netherlands
Tel:	+31 30 2533859
Fax:	+31 30 2537623
E-mail:	a.m.j.j.bonvin@uu.nl
Project website address:	www.wenmr.eu

Table of Contents

Table of Contents	2
Section 1: Summary of Deliverable	3
<i>Background</i>	3
<i>Goal</i>	3
Section 2: SAXS/NMR/docking refinement protocol	4
1. <i>Introduction to SAXS data</i>	4
2. <i>Introduction to CRY SOL and Chi (χ)</i>	4
3. <i>Incorporating Chi into HADDOCK's scoring function</i>	5
4. <i>Running SAXS-based HADDOCK Scoring on a Test Case</i>	6
5. <i>Reference</i>	10

Section 1: Summary of Deliverable

Background

As part of the efforts to integrate different structural biology methods into single analysis tools and pipelines, we aimed at developing a protocol for joint usage of SAXS and NMR data in docking. In particular, we set out to combine joint SAXS/NMR approaches with computational docking, which allows one to predict and assess protein-protein interfaces.

As described in more detail in **Deliverable 1.6** (the periodic report over period 2 of the WeNMR project), this involved the inclusion of SAXS data into the scoring stage of HADDOCK. That integration did result in improved selection of near-native (acceptable) solution for SAXS data in those cases for which the largest protein component has a prolate or oblate shape. A publication describing the integration of SAXS in HADDOCK (**Milestone 16**) was published in May 2013 in Acta Crystallogr D Biol Crystallogr ([doi: 10.1107/S0907444913007063](https://doi.org/10.1107/S0907444913007063)).

Goal

The goal of this deliverable was to implement the joint NMR/SAXS/Docking protocol in WeNMR and make it freely available through the WeNMR VRC. The protocol as described in the remaining of this deliverable is therefore also available on the WeNMR website: <http://www.wenmr.eu/wenmr/haddock-saxs-scoring>

Section 2: SAXS/NMR/docking refinement protocol

1. Introduction to SAXS data

Small Angle X-Ray Scattering (SAXS) allows characterization of biomolecular complexes under native conditions. It measures the intensity scattered by a protein sample at low scattering angles. From the scattering profile, size- and shape-related information, such as molecular weight, radius of gyration, maximum molecular dimension and a low-resolution 3D molecular envelope, can be extracted. (For more information on the SAXS data and the tools to interpret it, you can visit the related [WeNMR services](#) page.)

To this end, the SAXS information is typically translated into:

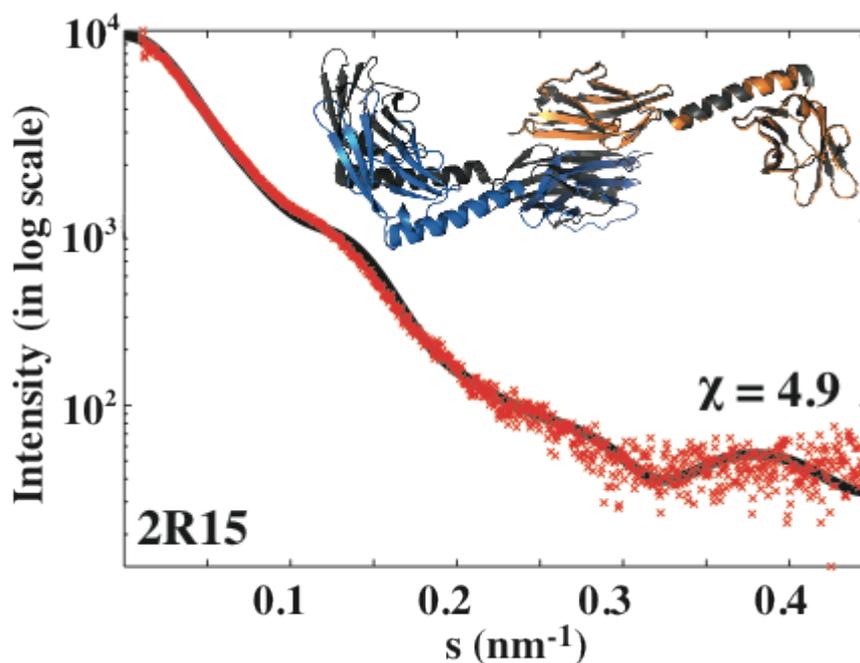
- (i) a molecular envelope, which can guide the construction of *ab initio* bead models or the docking of individual 3D structures,
- (ii) a restraining energy term, which can be used to refine structure directly against the SAXS curve (often in combination with orientational NMR restraints)
- (iii) a scoring term that calculates the discrepancy (Chi-or- χ) between the experimental scattering curve and the back-calculated one from the model.

In this document we focus on point (iii) and investigate the information-content of SAXS data and their usefulness in filtering (scoring) docking decoys.

2. Introduction to CRY SOL and Chi (χ)

To measure the quality of the fit between the model and the SAXS data, expressed as the discrepancy (χ), we use [CRY SOL](#), a freely available software to:

- (i) simulate the evaluation of the SAXS profile for macromolecules with known atomic structure (in PDB format).
- (ii) fit the simulated scattering profile to the experimental one obtained from SAXS measurements. This is done by minimizing the discrepancy (χ^2) value between two curves. If you are interested how this calculation is done, you can check CRY SOL's [reference paper](#).



An example CRY SOL output. The black line is the theoretical scattering curve calculated by CRY SOL (refers to the docking model-colored complex). The red crosses represent the experimental SAXS profile (measured for the black complex-PDB id: 2r15). The quality of the fit (χ) is reported as 4.9.

3. Incorporating Chi into HADDOCK's scoring function

In order to probe the scoring ability of the SAXS χ to discriminate near-native models, we re-ranked large set of docking models according to their fit to the experimental SAXS (χ) data. Our statistics revealed that χ -based SAXS scoring has a significant discriminative ability ([Karaca and Bonvin, Acta Cryst D, 2013](#)). However, the best results were achieved when χ is combined with the regular HADDOCK score with a weight of 50:

$$HADDOCK_{SAXS} = HADDOCKScore + 50 \cdot \chi_{SAXS}$$

where *HADDOCKScore* refers to different combinations of individual energy terms given different docking stages:

it0 (Rigid-body): $0.01E_{vdW} + 1.0E_{Elec} + 0.01E_{Restr} - 0.01E_{BSA} + 1.0E_{Desolv} + 0.1E_{Sym}$

it1 (Simulated Annealing): $1.0E_{vdW} + 1.0E_{Elec} + 0.1E_{Restr} - 0.01E_{BSA} + 1.0E_{Desolv} + 0.1E_{Sym}$

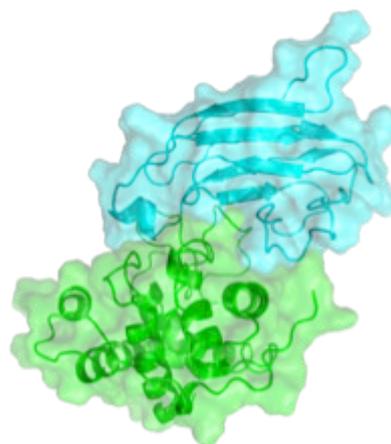
water (Solvent Refinement): $1.0E^{vdW} + 0.2E_{Elec} + 0.1E_{Restr} + 1.0E_{Desolv} + 0.1E_{Sym}$

Here, E_{vdW} is the van der Waals intermolecular energy, E_{Elec} the intermolecular electrostatic energy, E_{CM} the restraint energy, E_{Desolv} the empirical desolvation energy, BSA the buried surface area and, if present, E_{Sym} the symmetry restraint energy

4. Running SAXS-based HADDOCK Scoring on a Test Case

In this section, we'll guide you through a real CAPRI modeling example. CAPRI is abbreviated for Critical Assessment of PRediction of Interactions (<http://www.ebi.ac.uk/msd-srv/capri/capri.html>). It's a blind docking contest dedicated to the performance assessment of various docking methods.

This example belongs to [CAPRI round 27 \(Target 58\)](#). It's a lysozyme-lysozyme inhibitor complex (PDB id: [4g9s](#)). Since we are going to perform a scoring exercise, we'll use a pre-docked run that was submitted for performance assessment. (To generate the models, we used a special solvated docking protocol of HADDOCK. For more information on the protocol please see [Kastritis et al., Proteins, 2013](#)).



In order to be able to follow the next steps, you should download the tar-zipped archive **HADDOCK-SAXS-Scoring.tgz** available from the WeNMR web site and extract it under a Unix/Linux environment:

```
wget http://www.wenmr.eu/wenmr/files/HADDOCK-SAXS-Scoring.tgz
```

```
tar xvfz HADDOCK-SAXS-Scoring.tgz
```

After the archive is extracted, please go into the newly generated directory:

```
cd HADDOCK-SAXS-Scoring
```

Here, the docking statistics are provided in the *index.html* file. Please open it with any web-browser and toggle through it. In this way, you'll be able to see how many clusters (binding modes) are generated and which one ranks the best. There you can also download the top ranking structure belonging to each cluster and check how they look like compared to the

reference structure [4g9s](#). [If you do so, you'll see that Cluster6 (ranking #4) represents the binding mode that is closest to 4g9s.]

Now it's time run the SAXS scoring on the generated models. For this, you need an **experimental SAXS curve** (*T58_SAXS.dat* - provided in the directory) and two scripts (*HADDOCK-SAXS-Scoring.csh*, *re-rank-SAXS.csh*).

As a first step you need to run *HADDOCK-SAXS-Scoring.csh*. This script runs **Crysol** for calculating the fit of each model with respect to the experimental data. Therefore you need to have the executable of **Crysol** (available from <http://www.embl-hamburg.de/biosaxs/download.html>) for running this script. After getting **Crysol**, please change its hard-coded source directory (located in line #29) encoded in *HADDOCK-SAXS-Scoring.csh* to the directory, where your **Crysol** is located:

```
set crysol_dir = <enter here the location of the Crysol executable>
```

HADDOCK-SAXS-Scoring.csh takes two arguments, the first one is the *list of pdb's* to be analyzed and the second one is the *root name of the pdb*. In this case, you should run the following:

```
./HADDOCK-SAXS-Scoring.csh file.nam cluster &
```

This will take ~5 min. to complete. When the calculation finishes, you'll get the *saxs.fit* output file, in which the Chi values for each model are located. In order to incorporate these numbers into the conventional HADDOCK score, you should then run *re-rank-SAXS.csh*.

```
./re-rank-SAXS.csh
```

This script will output new scoring values, i.e. $HADDOCK_{SAXS}$ values. Now, it's time to check how the scoring gets affected after adding up the Chi value to it. For this, please do

```
./prepare-plot-stats.csh
```

As an output of this script, you'll get the stats file *HSSAXS_HS_SAXSFit_RMSD.list* that contains each scoring measure (Chi, HADDOCK, HADDOCKSAXS) and i-RMSD of a given model. Here i-RMSD denotes interface-RMSD and it measures how well the interface is predicted (since we had the Xtal structure, we were able to pre-calculate i-RMSDs for you). Ideally for an "acceptable" model, i-RMSD should be < 4Å.

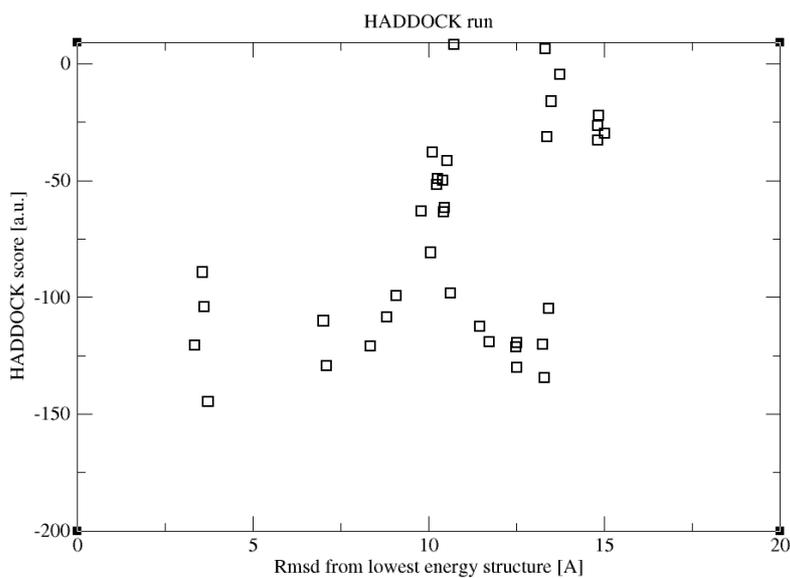
Now with any plotting tool, you can plot the columns of **HSSAXS_HS_SAXSFit_RMSD.list** and observe the distribution of each score with respect to i-RMSD. If you have [xmgrace](#) installed, you can also use the following commands to generate these plots:

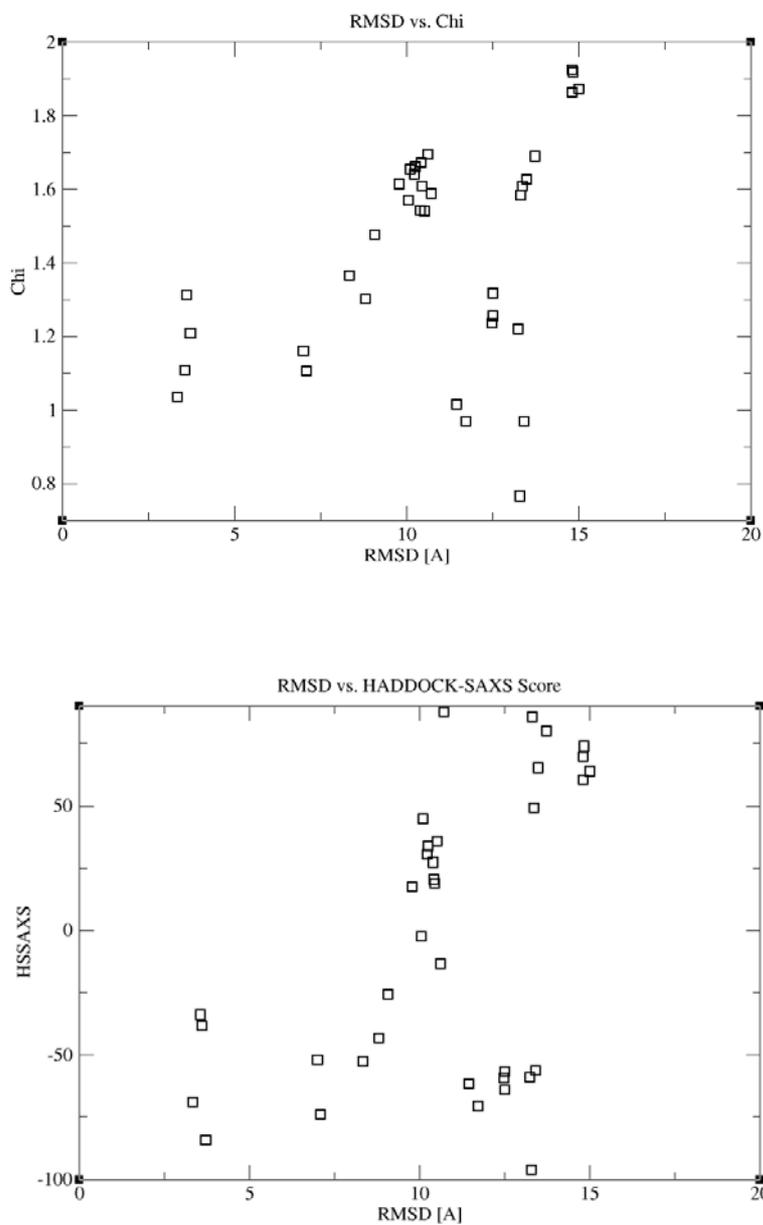
```
./make_rms-ene.csh 5 3 *_RMSD.list
./make_rms-HSSAXS.csh 5 2 *_RMSD.list
./make_rms-chi.csh 5 4 *_RMSD.list
```

Then just run xmgrace for each output file:

```
xmgrace rms-ene.xmgr
xmgrace rms-HSSAXS.xmgr
xmgrace rms-chi.xmgr
```

As a result, you'll get the following graphs:





Inspection of those plots reveals that the combination of HADDOCK and SAXS scoring provides the most funnel-like distribution. If you would like to check which structure ranks the first now and compare it with the conventional HADDOCK results, you can look through *HSSAXS_HS_SAXSFit.list*. In this file, the first column contains the pdbid, second one the $HADDOCK_{SAXS}$ score, the third one the HADDOCK score and fourth one the corresponding Chi value.

Now, let's see how the cluster ranking is changing, after calculating the energy statistics based on the combined HADDOCK_{SAXS} score:

```
#Clstrid meanE stdE
Cluster7 -71.2 17.7
Cluster1 -59.8 3.1
Cluster6 -56.4 24.3
Cluster4 -55.6 12.9
Cluster2 5.7 16.5
Cluster9 31.2 46.3
Cluster3 31.6 11.2
Cluster5 66.9 5.9
Cluster8 69.9 16.4
```

With this new scoring scheme, cluster6 (the "correct" binding mode) ranks 3rd, whereas it was ranking 4th before. Thus, we have a small improvement in scoring. The reason for not observing a dramatic change lies in the shape-dependent nature of the SAXS data: When two members of a complex are similar in shape and size, application of the SAXS data is not "that" discriminative. Please read our published article ([Karaca and Bonvin 2013](#)) for more information on this.

Ideally, you can use the scripts provided in this tutorial to calculate HADDOCK_{SAXS} score of any other HADDOCK model. In case you need more details on the scripts, you can consult the HADDOCK-SAXS-scriptorium page filed under the HADDOCK WeNMR [Documentation services](#).

5. Reference

- E. Karaca and A. M. J. J. Bonvin, [On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys](#), Acta Cryst. (2013). D69, 683-694, doi:10.1107/S0907444913007063