FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles

Dina Schneidman-Duhovny[1,*], Michal Hammel[2], John A. Tainer[2,3] and Andrej Sali[1,*]

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry and California Institute for Quantitative Biosciences (QB3), University of California at San Francisco, CA 94143, USA, [2]Molecular Biophysics & Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA and [3]Department of Molecular and Cellular Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

ABSTRACT

Small Angle X-ray Scattering (SAXS) is an increasingly common and useful technique for structural characterization of molecules in solution. A SAXS experiment determines the scattering intensity of a molecule as a function of spatial frequency, termed SAXS profile. Here, we describe three web servers for modeling atomic structures based on SAXS profiles. FoXS (Fast X-Ray Scattering) rapidly computes a SAXS profile of a given atomistic model and fits it to an experimental profile. FoXSDock docks two rigid protein structures based on a SAXS profile of their complex. MultiFoXS computes a population-weighted ensemble starting from a single input structure by fitting to a SAXS profile of the protein in solution. We describe the interfaces and capabilities of the servers (salilab.org/foxs), followed by demonstrating their application on Interleukin-33 (IL-33) and its primary receptor ST2.

INTRODUCTION

SAXS has become a widely used technique for structural characterization of molecules in solution ([1]). A key strength of SAXS is that it is informative about the shapes of macromolecules as large as 1000 Å across at near physiological conditions, in the 10 to 50 Å resolution range. The experiment is performed with ∼15 μl of the sample at the concentration of ∼1.0 mg/ml. It usually takes only a few minutes on a well-equipped synchrotron beam line ([1,2]) and can be conducted for a range of conditions ([3]). The SAXS profile of a macromolecule, $I(q)$, is computed by subtracting the SAXS profile of the buffer from the SAXS profile of the macromolecule in the buffer. The profile can be converted into an approximate distribution of pairwise atomic distances of the macromolecule (i.e. the pair-distribution function) *via* a Fourier transform.

Computational approaches for modeling a macromolecular structure based on its SAXS profile can be classified based on the system representation into *ab initio* and atomic resolution modeling methods ([4,5]). On the one hand, the *ab initio* methods search for coarse 3-dimensional shapes represented by dummy atoms (beads) that fit the experimental profile ([6–8]). On the other hand, atomic resolution modeling approaches generally rely on an all atom representation to search for models that fit the computed SAXS profile to the experimental one ([9]). Therefore, atomic resolution modeling can be used only if an approximate structure or a comparative model of the studied molecule or its components are available. With the increasing number of structures in the Protein Data Bank (PDB) ([10]) that can serve as templates for comparative modeling of a large number of sequences ([11]), we have focused our own efforts on atomic resolution modeling with SAXS profiles ([12–15]).

SAXS-based atomic modeling can be used in a wide range of applications, such as comparing solution and crystal structures, modeling of a perturbed conformation (e.g. modeling active conformation starting from non-active conformation), structural characterization of flexible proteins, assembly of multi-domain proteins starting from single domain structures, assembly of multi protein complexes, fold recognition and comparative modeling, modeling of missing regions in the high-resolution structure and determination of biologically relevant states from the crystal ([16–18]). Several software packages and webservers are available for some of these tasks, including ATSAS ([19]), pyDockSAXS ([20,21]) and ClusPro ([22]). Here, we describe how our tools can be used to facilitate addressing several of these questions (Figure [1]).

---

*To whom correspondence should be addressed. Tel: +1 415 514 4227; Fax: +1 415 514 4231; Email: dina.schneidman@mail.huji.ac.il
Correspondence may also be addressed to Andrej Sali. Email: sali@salilab.org
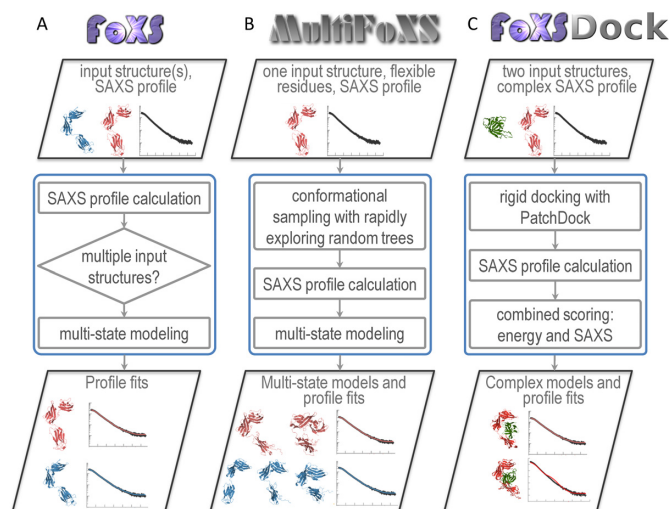
Figure 1. Overview of the input and output of the three servers: (**A**) FoXS, (**B**) MultiFoXS and (**C**) FoXSDock.

COMPUTING SAXS PROFILE WITH FoXS

Rapid and accurate computation of the SAXS profile of a given atomic structure and its comparison with the experimental profile is a basic component in any SAXS-based atomic modeling. FoXS is a webserver that performs this task (13,23). Over the past few years, FoXS webserver was upgraded by adding new multi-profile fitting functionality and interactive output visualization for profiles and structures, as we describe below.

The profiles are calculated using the Debye formula (24) and fitted to the experimental data with adjustment of the excluded volume (c_1) and hydration layer density (c_2) parameters. The fit score is computed by minimizing the χ function with respect to c, c_1 and c_2:

$$\chi = \sqrt{\frac{1}{S}\sum_{i=1}^{S}\left(\frac{I_{\exp}(q_i) - c\,I(q_i, c_1, c_2)}{\sigma(q_i)}\right)^2} \quad (1)$$

where $I_{\exp}(q)$ and $I(q)$ are the experimental and computed profiles, respectively, $\sigma(q)$ is the experimental error of the measured profile, S is the number of points in the profile and c is the scale factor.

The input to the server is a structure file in the PDB format and an experimental SAXS profile. The output is the computed profile fitted to the experimental one. The server displays the profile fit along with the residuals on the left side of the window. The input structure is displayed in a JS-mol window (25). The user can zoom on different parts of the fit plot, download the fit plot file and rotate the structure in the JSmol window (Supplementary Figures S1, S3). A summary of fit parameters (χ, c_1 and c_2) is given under the two windows.

If the user uploads multiple structures (in a zip file), the server computes the profile for each structure and fits it to the experimental data. The output is again displayed in the profile fit and structure windows, followed by a table with the fit summary for each structure. The user can use show/hide button to display or hide the computed profile

and the corresponding structure (Supplementary Figures S2, S4). The rest of the output is described in the following section.

FITTING MULTIPLE STRUCTURES TO A SINGLE PROFILE

The unique capability of FoXS webserver is a possibility to account for multiple states contributing to a single observed SAXS profile. Multiple states can correspond to conformational heterogeneity (multiple conformations of the same protein or complex) and/or compositional heterogeneity (varying contents of protein and ligand molecules in the system). We have developed a scoring function and an enumeration procedure to compute multi-state models based on a SAXS profile. The score of a multi-state model is:

$$\chi = \sqrt{\frac{1}{S}\sum_{i=1}^{S}\left(\frac{I_{\exp}(q_i) - c\sum_{n} w_n I_n(q_i, c_1, c_2)}{\sigma(q_i)}\right)^2} \quad (2)$$

where $I_n(q, c_1, c_2)$ and w_n are the computed profile and the corresponding weight, respectively, for each of the N states in the model; this equation minimizes data overfitting by using a single set of c_1 and c_2 values for all N states. The enumeration of multi-state models (subsets of conformations and their weights) of size N (typically, $1 < N < 5$), such that the corresponding sum of weighted SAXS profiles fits the experimental SAXS data, is performed iteratively using a branch-and-bound method (26).

Given multiple input structures, FoXS by default enumerates multi-state models and fits them to the data (Figure 1). The server output page contains a link to the multi-state model page above the display windows. The top of the multi-state model page (Supplementary Figure S5) shows the bar plot (left) and the fit plot (right). The bar plot shows the χ values for the best scoring N-state models (for each N). The error bar indicates the range of χ values for the top 100 multi-state models. The fit plot is similar to the fit plot for single structures, displaying weighted profiles for $N > 1$. The best scoring N-state model for each N is shown below the bar and fit plots. For each model, a summary of fit parameters (χ, c_1 and c_2) is given, followed by the display of the structural states in JSmol windows (one window per state) and their corresponding weights. The user can download a weighted profile for a multi-state model.

PROTEIN–PROTEIN DOCKING WITH FoXSDock

While many structures of single protein components are increasingly available, structural characterization of their complexes remains challenging. Methods for modeling assembly structures from individual components frequently suffer from large errors, due to protein flexibility and inaccurate scoring functions. SAXS profile of the complex can significantly improve the success rate of protein–protein docking (14,15). FoXSDock is a web server for protein–protein docking restrained by a SAXS profile of the complex; the input to FoXSDock is the structures of the docked proteins in the PDB format and a SAXS profile of their complex. The output is a list of complex models computed *via* rigid docking (27,28) sorted by a combined SAXS

and statistical potential (energy) scores (29) (Supplementary Figure S6). To calculate the combined score, SAXS χ scores and statistical potential scores are normalized with respect to all docking models. The combined score is the sum of the normalized Z-scores. The normalization of the scores allows us to avoid the use of weights for the terms of the combined score. It is possible that the sample for the SAXS experiment contained a mixture of monomers and the complex. Therefore, FoXSDock can optionally rely on a multi-state weighted scoring function (Equation 2). Currently FoXSDock is the only SAXS-based docking method with multi-state scoring function.

SAMPLING STRUCTURAL STATES WITH MultiFoXS

MultiFoXS addresses conformational heterogeneity in solution by relying on a SAXS profile. The input is a single atomic structure (or a comparative model), a list of flexible residues and a SAXS profile for the protein. The server proceeds in two steps. In the first step, it samples the input structure by exploring the space of the φ and ψ main chain dihedral angles of the user-defined flexible residues with a Rapidly exploring Random Trees (RRTs) algorithm (30–33). The second step is identical to the multi-state FoXS protocol: a SAXS profile is calculated for each sampled conformation, followed by enumeration of the best-scoring multi-state models using the multi-state scoring function (Equation 2). The output is similar to the FoXS multi-state model page, with an additional plot displaying the radius of gyration (Rg) distribution for the top 100 best scoring N-state models (Figure 2).

APPLICATION TO ST2 AND ST2-IL33

We now demonstrate the three servers by analyzing the experimental SAXS profiles of the primary Interleukin-33 (IL-33) receptor ST2 and its complex with IL33 (BIOISIS ST2ILP) (34).

FoXS

The X-ray structure of the ST2-IL33 complex is available (34) (PDB 4kc3). Therefore, we first compared the X-ray structure to the SAXS profile using FoXS (Supplementary Figure S1). The computed profile does not fit the experimental data within the noise ($\chi = 3.4$). We hypothesized that the unresolved loops and the C-terminal Histidine tag in the X-ray structure explain the difference. Using MODELLER v9.8 (11), we added the missing atoms and compared 10 alternative models to the SAXS profile (Supplementary Figure S2). The resulting models have a significantly better fit than the X-ray structure, with the best χ value of 1.6, which is within the experimental noise (35). The ST2 chain extracted from the complex does not fit the experimental profile either (Supplementary Figure S3, $\chi = 7.7$). In this case, addition of missing atoms improved the fit only slightly (Supplementary Figure S4, $\chi = 5.5$), in contrast to the ST2-IL33 complex. Therefore, we used MultiFoXS for computing a multi-state model of ST2.
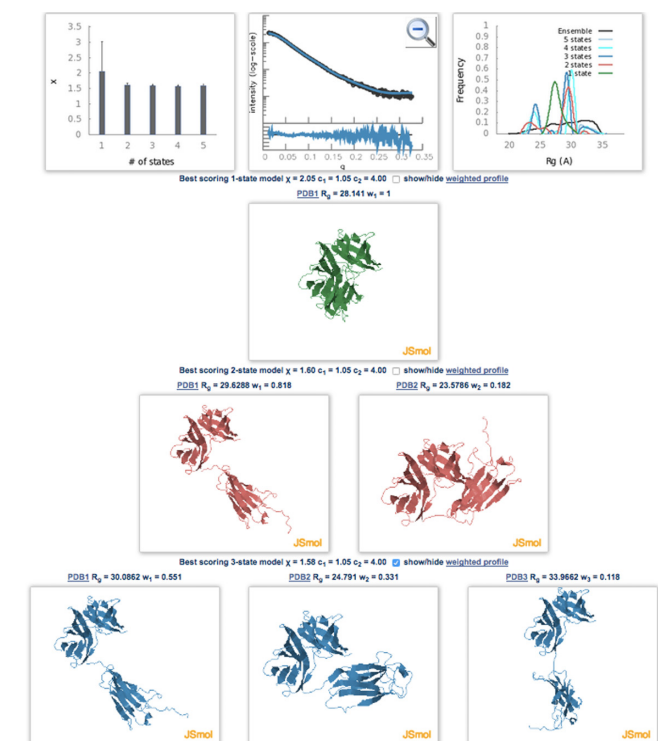

Figure 2. MultiFoXS output page for ST2. The bar, fit and Rg distribution plots are at the top. The top scoring one-, two- and three-state models are shown in green, red and blue, respectively.

MultiFoXS

The input was the ST2 model and the experimental SAXS profile. ST2 consists of three immunoglobulin-like domains (D1–D3). Based on previous studies, we defined the linker between the D2 and D3 domains as flexible, as we did the C-terminal Histidine tag. The server sampled over 10 000 conformations, calculated their SAXS profiles, enumerated and scored multi-state models. The χ value improved significantly even for a single-state model ($\chi = 2.1$) (Figure 2) compared to the X-ray structure (Supplementary Figures S3, S4). The fit is even better with two- or three-state models ($\chi = 1.6$), as expected. To estimate the number of states in solution, we examined the Rg distribution (Figure 2, top right corner). The Rg distribution in the initial pool of 10 000 conformations is almost uniform (black line). The top scoring one-state models (green line) have Rg in the range of 26–30 Å. For two- and three-state models (red and blue lines, respectively), the Rg distribution has two peaks: one at 23–25 Å with the weight of 0.3 and the other at 28–31 Å with the weight of 0.6. For models with 3 or more states, there is also a peak of more open conformations at 31–35 Å with the weight of 0.1. The conformations from the most populated peak (Rg in the range of 28–31 Å) resemble the IL33 binding conformation. Therefore, based on MultiFoXS results, we conclude that ST2 exists in multiple states in solution, corresponding to a wide range of open and closed conformations. Upon IL33 binding, there is a population shift to the IL33 binding conformation.

FoXSDock

The input was the ST2 model, the IL33 NMR structure (PDB 2kll) and the SAXS profile of the complex. FoXS-Dock produced a list of complex models (Supplementary Figure S6A). The models illustrate the benefit of docking restrained by a SAXS profile: The model with the best combined SAXS and energy score has a relatively low interface-RMSD deviation from the X-ray structure of 3.7 Å, while the model ranked as top scoring by the energy score alone has a much larger interface-RMSD error of 14.1 Å.

APPLICATION TO POLYNUCLEOTIDE KINASE WITH DNA SUBSTRATE

Mammalian polynucleotide kinase (mPNK) is a DNA repair enzyme (36) consisting of three functional domains: 5′ DNA kinase and 3′ -phosphatase domains (closely associated into the catalytic segment PK), and an N-terminal FHA (Forkhead-associated) domain. We analyze the solution conformations of mPNK and its interaction complex with the DNA substrate using the SAXS profiles of full-length mPNK and PK-DNA complex, respectively (37).

FoXS

The X-ray structure of mPNK is available (36) (PDB 1yj5). Although the computed profile has a reasonable fit to the experimental data (Supplementary Figure S7, $\chi = 2.2$), we used MultiFoXS to check whether multi-state models for mPNK can improve the fit quality.

MultiFoXS

We defined the linker residues (111–139) between the catalytic segment and FHA domains as flexible. The server sampled over 10 000 conformations, calculated their SAXS profiles, enumerated and scored multi-state models. The χ value improved slightly even for a single-state model ($\chi = 2.0$) (Supplementary Figure S8) compared to the X-ray structure (Supplementary Figure S7). The fit is significantly better with multi-state models with two or more states ($\chi < 1.5$). We observe two major peaks in the *Rg* distribution: one at 34 Å and the other at 50 Å with equal weights. Therefore, we conclude that in solution mPNK is populating a wide range of conformations. These can be classified into an open state with the extended linker and the closed state with the FHA domain in close proximity to the catalytic segment.

FoXSDock

The input was the catalytic segment of mPNK, the DNA hairpin H1, the SAXS profile of their complex and a distance constraint between ASP396 (OD1 atom) and the 5′-hydroxyl group of the DNA substrate, as described previously (37). FoXSDock produced a list of complex models (Supplementary Figure S9A). In this case, only the SAXS score is used for ranking the models, since our statistical potential does not extend to DNA atoms. The top scoring model had a χ value of 2.5 and was consistent with the constrained distance and additional biochemical information regarding mPNK mutants (37).

CONCLUSION

The three servers facilitate the use of SAXS data in a variety of molecular modeling applications, such as comparing solution and crystal structures, structural characterization of flexible proteins, assembly of multi-protein complexes and modeling of missing regions in the high-resolution structure. Atomic resolution representation of the modeled system provides strong constraints on possible solutions consistent with SAXS data, thus making SAXS-based modeling helpful for characterizing biomolecular systems. To maximize the accuracy of the predictions, the servers rely on: (i) scoring functions for fitting multi-state models with single set of fitting parameters to reduce data overfitting, (ii) efficient deterministic approach for enumeration of multiple states and (iii) advanced methods for exhaustive sampling of conformations and complexes. Moreover, the three servers provide user-friendly interface and visualization of the modeling results. An illustration was provided by the SAXS-based modeling of the ST2-IL33 complex and its components. The accuracy and precision of SAXS-based modeling will improve as the remaining challenges, such as incomplete representations of system components at atomic resolution, inaccurate hydration layer modeling, insufficient conformational sampling, data overfitting and sample heterogeneity, are addressed.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Hura,G.L., Menon,A.L., Hammel,M., Rambo,R.P., Poole,F.L. 2nd, Tsutakawa,S.E., Jenney,F.E. Jr, Classen,S., Frankel,K.A., Hopkins,R.C. *et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods*, **6**, 606–612.

2. Dyer,K.N., Hammel,M., Rambo,R.P., Tsutakawa,S.E., Rodic,I., Classen,S., Tainer,J.A. and Hura,G.L. (2014) High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.*, **1091**, 245–258.

3. Hura,G.L., Budworth,H., Dyer,K.N., Rambo,R.P., Hammel,M., McMurray,C.T. and Tainer,J.A. (2013) Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods*, **10**, 453–454.

4. Putnam,C.D., Hammel,M., Hura,G.L. and Tainer,J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.*, **40**, 191–285.

5. Rambo,R.P. and Tainer,J.A. (2013) Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.*, **42**, 415–441.

6. Chacon,P., Moran,F., Diaz,J.F., Pantos,E. and Andreu,J.M. (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.*, **74**, 2760–2775.

7. Svergun,D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, **76**, 2879–2886.

8. Svergun,D.I., Petoukhov,M.V. and Koch,M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946–2953.

9. Petoukhov,M.V. and Svergun,D.I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, **89**, 1237–1250.

10. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

11. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

12. Forster,F., Webb,B., Krukenberg,K.A., Tsuruta,H., Agard,D.A. and Sali,A. (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. *J. Mol. Biol.*, **382**, 1089–1106.

13. Schneidman-Duhovny,D., Hammel,M. and Sali,A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.*, **38**, W540–W544.

14. Schneidman-Duhovny,D., Hammel,M. and Sali,A. (2011) Macromolecular docking restrained by a small angle X-ray scattering profile. *J. Struct. Biol.*, **173**, 461–471.

15. Schneidman-Duhovny,D., Rossi,A., Avila-Sakar,A., Kim,S.J., Velazquez-Muriel,J., Strop,P., Liang,H., Krukenberg,K.A., Liao,M., Kim,H.M. *et al.* (2012) A method for integrative structure determination of protein-protein complexes. *Bioinformatics*, **28**, 3282–3289.

16. Hammel,M. (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur. Biophys. J.*, **41**, 789–799.

17. Schneidman-Duhovny,D., Kim,S.J. and Sali,A. (2012) Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.*, **12**, 17.

18. Rambo,R.P. and Tainer,J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559–571.

19. Petoukhov,M.V., Franke,D., Shkumatov,A.V., Tria,G., Kikhney,A.G., Gajda,M., Gorba,C., Mertens,H.D., Konarev,P.V. and Svergun,D.I. (2012) New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.

20. Pons,C., D'Abramo,M., Svergun,D.I., Orozco,M., Bernado,P. and Fernandez-Recio,J. (2010) Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. *J. Mol. Biol.*, **403**, 217–230.

21. Jimenez-Garcia,B., Pons,C., Svergun,D.I., Bernado,P. and Fernandez-Recio,J. (2015) pyDockSAXS: protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Res.*, **43**, W356–W361.

22. Xia,B., Mamonov,A., Leysen,S., Allen,K.N., Strelkov,S.V., Paschalidis,I., Vajda,S. and Kozakov,D. (2015) Accounting for observed small angle X-ray scattering profile in the protein-protein docking server ClusPro. *J. Comput. Chem.*, **36**, 1568–1572.

23. Schneidman-Duhovny,D., Hammel,M., Tainer,J.A. and Sali,A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.*, **105**, 962–974.

24. Debye,P. (1915) Zerstreuung von Röntgenstrahlen. *Ann. Phys.*, **351**, 809–823.

25. Herraez,A. (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ.*, **34**, 255–261.

26. Lawler,E.L. and Wood,D.E. (1966) Branch-and-bound methods: A survey. *Oper. Res.*, **14**, 699–719.

27. Duhovny,D., Nussinov,R. and Wolfson,H.J. (2002) In: Guigó,R. and Gusfield,D. (eds.), *Second International Workshop, WABI 2002*. Springer Berlin/Heidelberg, Rome, Italy, Vol. **2452**/**2002**, pp. 185–200.

28. Schneidman-Duhovny,D., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.

29. Dong,G.Q., Fan,H., Schneidman-Duhovny,D., Webb,B. and Sali,A. (2013) Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics*, **29**, 3158–3166.

30. LaValle,S.M. and Kuffner,J.J. (2001) Rapidly-exploring random trees: progress and prospects. *Algorithmic Comput. Robotics*, 293–308.

31. Amato,N.M. and Song,G. (2002) Using motion planning to study protein folding pathways. *J. Comput. Biol.*, **9**, 149–168.

32. Cortes,J., Simeon,T., Ruiz de Angulo,V., Guieysse,D., Remaud-Simeon,M. and Tran,V. (2005) A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics*, **21 Suppl 1**, i116–i125.

33. Raveh,B., London,N. and Schueler-Furman,O. (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins*, **78**, 2029–2040.

34. Liu,X., Hammel,M., He,Y., Tainer,J.A., Jeng,U.S., Zhang,L., Wang,S. and Wang,X. (2013) Structural insights into the interaction of IL-33 with its receptors. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14918–14923.

35. Rambo,R.P. and Tainer,J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477–481.

36. Bernstein,N.K., Williams,R.S., Rakovszky,M.L., Cui,D., Green,R., Karimi-Busheri,F., Mani,R.S., Galicia,S., Koch,C.A., Cass,C.E. *et al.* (2005) The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol. Cell*, **17**, 657–670.

37. Bernstein,N.K., Hammel,M., Mani,R.S., Weinfeld,M., Pelikan,M., Tainer,J.A. and Glover,J.N. (2009) Mechanism of DNA substrate recognition by the mammalian DNA repair enzyme, Polynucleotide Kinase. *Nucleic Acids Res.*, **37**, 6161–6173.