Native state protein dynamics
Groot, Berend Lammert de

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1999

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CONCLUDING REMARKS

Current state of the art

As summarised in the introduction of this thesis, recent developments in computer simulations of biological macromolecules have enhanced the range of applicability of these techniques for the study of conformational properties of proteins. The methods described in this thesis form another contribution to this field, and applications to several proteins have yielded interesting results. The Essential Dynamics technique has proven a powerful analysis tool not only for the interpretation of MD simulations of proteins, but also of experimental and CONCOORD generated protein conformations. Enhanced sampling protocols based on the Essential Dynamics technique have shown to reach rates of conformational sampling that are 5-10 times higher than those reached by conventional MD. However, success of the ED sampling technique depends on the accuracy of the initial definition of the collective coordinates along which the positions are constrained in the sampling technique. Comparisons of subspaces spanned by these coordinates from different simulations have shown that the subspace overlap is usually not larger than 60 % (see chapters 2, 5, 6). Thus, although an approximate convergence of the definition of the principal collective coordinates can be obtained from MD simulations in the nanosecond time range, there is still a significant level of noise in this definition. It has recently been shown that several shorter simulations sample the conformational space of proteins more efficiently than one single, longer simulation\textsuperscript{19}. It would be interesting to investigate whether the use of multiple short simulations also yields a faster convergence of the essential subspace.

The CONCOORD method has been described and applied in chapters 6 and 7. It has proven a simple yet powerful technique to study protein conformational freedom. Because it is based on very different principles than for instance Molecular Dynamics, the specific strengths and weaknesses of the CONCOORD method differ from those of the MD technique. Therefore, the two methods partially complement each other, enabling a deeper insight in conformational properties of proteins than can be obtained from either technique individually. Moreover, the ability of the CONCOORD method to yield modes of collective protein fluctuation that are similar to those obtained from MD and experiment proves that protein dynamics is largely governed by restrictions imposed by interactions in the native structure.

Normal Mode analyses form another computational tool to study fluctuations in proteins. Although limited to the fluctuations in a single harmonic well, the method has been shown to sample biologically relevant motions (e.g. ref. 220), and can be applied to relatively large proteins\textsuperscript{212,221}. In a recent study, we have shown that combined Normal Modes from multiple (local) minima are more similar to collective modes of fluctuation derived from MD simulations, than are Normal Modes extracted from a single minimum\textsuperscript{222}. 
An efficient, automated procedure to combine Normal Mode results from multiple local minima could form an alternative method to study protein dynamics. Further progress may be obtained by combinations of Molecular Dynamics, CONCOORD and Normal Modes. A method that exploits the specific strengths of each of the three techniques may prove a valuable tool for the study of conformational fluctuations in proteins. It should be noted, however, that both NM and CONCOORD critically depend on the presence of a high-resolution starting structure, whereas in MD, structures can be allowed to equilibrate from a low resolution model or from a structure determined under different conditions (e.g. different solvent).

Limitations

A good illustration of the limitations of the methods described in this thesis is formed by a project in which we studied the coupled tertiary and quaternary structural changes in haemoglobin. After the concept of Essential Dynamics was first conceived, the idea arose that the study of allosteric proteins would be an ideal application of the technique. Allosteric proteins are multi-subunit proteins that are characterised by a cooperative substrate binding. Communication between subunits is responsible for the dependence of the substrate affinity of one subunit on the binding state of the others. The binding affinity can often be further regulated by binding of other molecules at sites distinct from the substrate binding site. Most allosteric proteins exist in two or more conformations that differ in the packing of their subunits (quaternary conformation). One of these states is the preferred conformation in the absence of substrate, and another quaternary conformation is associated with the fully liganded state. In the traditional view\(^2\), the binding of substrate to one of the subunits (slightly) changes the conformation of that subunit, triggering a quaternary conformational change that changes the substrate affinity of the other subunits. Molecules that regulate the activity of such proteins specifically stabilise one of the quaternary conformations. The correlation between the (usually small) tertiary structural changes and the larger global changes, if sampled realistically, would be detected by a covariance analysis like Essential Dynamics, and therefore such an analysis technique could improve our understanding of the mechanisms involved in such conformational changes.

Haemoglobin, probably the best studied allosteric protein, was studied with Molecular Dynamics techniques with the hope to learn about the coupling between the changes that take place in the subunits upon oxygen binding, and overall structural changes. Simulations of 1 nanosecond did not significantly sample the experimentally known quaternary conformational change. This structural change is known to take place in a time scale of microseconds after binding (or removal) of oxygen\(^2\). The three orders of magnitude time difference between the simulations and the experiment are
probably the explanation for the absence of the conformational change in the simulations. However, applications to T4 lysozyme had shown that domain motions were sampled to an appreciable amplitude in simulations of the same length. The critical difference between the two proteins is the presence of a (free) energy barrier between the different conformational states of haemoglobin. In T4 lysozyme the domain motion(s) are not restricted by such an internal barrier and fluctuate diffusively during simulation. CONCOORD simulations of haemoglobin did sample the conformational changes between the different quaternary structures. The CONCOORD method is less sensitive to internal barriers since there is no path dependence between successively generated structures. The conformational changes sampled by CONCOORD, however, did not show a unique mechanism of coupling between tertiary and quaternary structural changes. The specific interaction of oxygen with the haem prosthetic group and the local structural changes are not modeled accurate enough to allow identification of such a coupling mechanism. The CONCOORD results, however, did indicate a direct role of the C-terminal residues of each subunit in the allosteric mechanism.

**Outlook**

Computational techniques are widely used for the study of conformational properties of biological macromolecules, and their range of application will only grow in the future. From the refinement of experimental structures to the *ab initio* folding of proteins, computer simulation techniques have proven to be valuable tools that can complement insights obtained from experiment. The predictive power of computer simulation techniques applied to proteins is still limited because of the large number of degrees of freedom that need to be treated explicitly. In the introduction of this thesis an overview was given of methods that are currently used to enhance the efficiency of computer simulation techniques to study protein dynamics. Only time can tell which (combination of) techniques will prove most useful in the future.

Based on the results presented in this thesis, it follows that a large portion of the configurational freedom is defined by restrictions that are imposed directly by the structure (chapters 6 and 7). Methods that do not use a molecular description on the atomic level will lack features that are directly related to the specific atomic interactions or packing. Another source of artifacts in computer simulation techniques is the representation of atomic interactions. Whereas many aspects of collective protein fluctuations may be correctly described by methods that lack a sophisticated treatment of interactions (e.g. quaternary structural changes in haemoglobin by CONCOORD), other, more subtle, mechanisms may not be correctly represented by such techniques (e.g. the coupling between tertiary and quaternary structural changes in haemoglobin). Therefore, the kind of application defines the method of choice.
Some applications do not require sophisticated all-atom treatment (e.g. determining the hinge-bending mode in lysozyme). Other processes however, depend on subtle interactions and/or take place at time scales that can currently not be simulated realistically (e.g. substrate entry or exit in enzymes or the protein folding process).

In chapters 6 and 7 it was shown that many conformational properties of proteins can be obtained by a much more simple technique (CONCOORD) than Molecular Dynamics. Although the applicability of the CONCOORD method is limited because of the absence of a realistic atomic description, it has the advantage that it does not suffer from sampling problems, at least within predefined limits. Progress with respect to the current implementation can be obtained by releasing some of the constraints imposed by a single conformation. This would allow the generation of conformations more distinct from the starting structure, and a sampling of the paths between those conformations. The difficulty in the design of a method to accomplish this is the prediction of which interactions are to be maintained for each generated conformer, and for which there are alternatives available. One straightforward approach would be to use the method as it is, in a recursive manner. The first step would be the generation of structures based on a single conformation. In next steps, structures generated in the previous steps could be used to define new sets of distance limitations, on the basis of which new structures could be generated. Preferable would be to have multiple experimental structures or reliable MD structures that could be used in the same fashion. Future studies will have to resolve whether meaningful results can be obtained in this way.

Summarising, although biomolecular computer simulations have come of age, many interesting processes involving dynamics of biological macromolecules are still beyond the scope of current computational techniques. Simulations employing sophisticated atomic models are limited to short time scales, and more coarse-grained methods lack the atomic detail that is often essential for a full understanding of a dynamical process. However, constant improvement of methodologies, together with a steady increase of (affordable) computer power will allow the study of more complex systems on longer timescales. Backed up by constant thorough experimental validation, methods will be developed in the next decade(s) that will allow detailed simulation of functional dynamics of proteins, the interactions of proteins with other molecules (docking) and the protein folding process.