Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 160 (2003) 65–73
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

The Xplor-NIH NMR molecular structure determination package

Charles D. Schwieters,[a,*] John J. Kuszewski,[a] Nico Tjandra,[b] and G. Marius Clore[c,*]

[a] *Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Building 12A, Bethesda, MD 20892-5624, USA*
[b] *Laboratory of Biophysical Chemistry National Heart, Lung, and Blood, Institute National Institutes of Health, Building 50, Bethesda, MD 20892-8013, USA*
[c] *Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, Bethesda, MD 20892-0510, USA*

Received 29 July 2002; revised 3 September 2002

Abstract

We announce the availability of the Xplor-NIH software package for NMR biomolecular structure determination. This package consists of the pre-existing XPLOR program, along with many NMR-specific extensions developed at the NIH. In addition to many features which have been developed over the last 20 years, the Xplor-NIH package contains an interface with a new programmatic framework written in C++. This interface currently supports the general purpose scripting languages Python and TCL, enabling rapid development of new tools, such as new potential energy terms and new optimization methods. Support for these scripting languages also facilitates interaction with existing external programs for structure analysis, structure manipulation, visualization, and spectral analysis.

1. Introduction

The XPLOR [1] and follow-on CNS [2] packages have proven highly popular for biomolecular structure determination using NMR spectroscopy and crystallography. XPLOR was initially derived from the generalized molecular dynamics and minimization program CHARMM [3], and in its initial applications to both NMR [4–9] and X-ray [10] structure determination, XPLOR was referred to as the CRAY-optimized version of CHARMM (Brünger, unpublished). Both the original XPLOR program and CNS are no longer in active development. Moreover, CNS lacks many NMR-specific features which have been recently developed at the NIH (i.e., post-1998 developments including the work in [11–19]), and incorporated into the internal NIH version of XPLOR which we call Xplor-NIH. Therefore, we have sought to distribute Xplor-NIH for research purposes, by reaching an agreement with the Accelrys corporation, which currently owns the XPLOR program. The agreement, now in place, allows us to distribute Xplor-NIH in its entirety (in both source and executable formats) freely to academic users, as well as to distribute freely to industrial users those portions of Xplor-NIH developed solely at the NIH.

Xplor-NIH is a generalized package for biomolecular structure determination from experimental NMR data combined with known geometric data. This is achieved by seeking the minimum of a target function comprising terms for the experimental NMR restraints, covalent geometry and non-bonded contacts using a variety of minimization procedures including molecular dynamics in Cartesian and torsion angle space, Monte Carlo methods and conventional gradient-based minimization.

Xplor-NIH was originally derived from XPLOR version 3.851 and contains all of the functionality therein. However, Xplor-NIH incorporates numerous completely new features designed to render its overall architecture highly flexible and to foster the rapid and easy development of new and improved functionality. These new features comprise the following:

- A number of additional NMR-specific features related to refinement against NMR observables [11,12,19] not included in XPLOR 3.851, as well as

* Corresponding authors.
E-mail addresses: Charles.Schwieters@nih.gov (C.D. Schwieters), mariusc@intra.niddk.nih.gov (G. Marius Clore).

1090-7807/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S1090-7807(02)00014-9

a variety of knowledge-based database potentials of mean force [13,15,18].
- A reduced variable dynamics module which permits completely generalized minimization and molecular dynamics in torsion angle and Cartesian coordinate space, as well as the effective treatment of rigid bodies [16].
- An entirely new overall framework to allow Xplor-NIH extensions to be written in C++.
- Interfaces between the C++ framework and the Python and TCL scripting languages which provide more flexible alternatives to the legacy XPLOR scripting language.
- A direct interface to the VMD-XPLOR visualization package [14].

We note that despite the numerous additions and modifications present in Xplor-NIH, great care has been taken to ensure that old XPLOR scripts still run as expected. CNS scripts can also be run with very minor modifications.

The structure of the paper is as follows. We first review the features of Xplor-NIH which are particularly useful in NMR structure determination, including the various energy terms associated with NMR experiments. Section 3 describes the C++ framework and its relationship to XPLOR's FORTRAN code, and Section 4 then discusses the advantages of modern general purpose scripting languages, and Xplor-NIH's current interfaces to the Python and TCL languages. Following this are short sections on regression testing of the package and on availability of the software.

2. Overview of XPLORs NMR functionality

Xplor-NIH contains many general purpose tools to generate and manipulate molecular structures. Protein and nucleic acid structures, including bond and angle definitions can be generated from sequence-only information, and the coordinate statement can be used to read and write (somewhat nonstandard) PDB structure coordinate files. Trajectory statements are used to read and write coordinate time series, generated, for example, by molecular dynamics runs. Once coordinates are generated it is straightforward to manipulate them with the aid of XPLOR's powerful atom selection language. Commonly performed tasks including fitting one structure to another based on a subset of atoms (such as α-carbons) and calculating average coordinates from an ensemble. These items are covered in depth in the XPLOR manual [1].

Optimizing the atomic coordinates such that they match the NMR observables can be achieved using a number of methods. In Cartesian coordinates Xplor-NIH provides Powell gradient minimization, and supports molecular dynamics simulated annealing optimization using the Verlet dynamics facility. In addition, it supports rigid-body minimization. However, more recently, we have found the internal variable module (IVM) [16] to be a more efficient and flexible facility for dynamics and optimization of atomic coordinates. The IVM allows one to perform dynamics and optimization in torsion angle coordinates and also allows for rigid-body optimization, or mixed cases in which there are some free atoms, some rigid regions, and other regions in which torsion angle are allowed to vary [16,17]. More exotic internal coordinates can also be defined in the IVM. For example, one can allow bond angles to vary in addition to torsion angles in a fashion which is appropriate for ring pucker (e.g., proline rings in proteins and sugar rings in nucleic acids and carbohydrates). The IVM allows a choice of molecular dynamics integration algorithm, including a sixth order predictor–corrector algorithm in addition to the usual velocity Verlet method. Cartesian gradient minimization remains useful in initial structure determination in order to achieve correct covalent geometries; one can then use a reduced set of internal coordinates for better efficiency. Xplor-NIH also supports structure determination through metric matrix distance geometry calculations [20], including substructure embedding without triangulation to provide starting structures for simulated annealing using the hybrid distance geometry-simulated annealing method [21].

Xplor-NIH structure calculations are generally performed in the absence of solvent molecules. While the inclusion of explicit water is possible (and sometimes performed [22]), this becomes quite expensive computationally, and it is our belief that the use of explicit solvent introduces an unnecessary, uncontrolled dependence on various force field parameters. We prefer to include solvent effects via knowledge-based potentials, described below.

The XPLOR language contains an interface to the VMD molecular structure viewer. This interface is used by the VMD-XPLOR package [14] to load structures, labels and trajectories from Xplor-NIH, and can be used to run structure calculations on one computer while displaying the results on another.

2.1. Potential energy terms

The following potential energy terms find frequent use in NMR structure determination.

2.1.1. Covalent geometry

Xplor-NIH has support for the usual idealized covalent energy terms (bonds, angles, and improper torsions). In NMR structure determination, it is often desirable to avoid artifacts introduced by molecular force fields, so only energy terms which encode geometric structure information are included. Covalent

terms ensure that fixed bond lengths, bond angles, planar atom groups, and chiral centers are correct.

2.1.2. Non-bonded contacts

Non-bonded contact potential terms include a repulsive quartic van der Waals term [21,23], which is used solely to prevent atoms from overlapping, and can be scaled down such that atoms may move through each other. This potential term has been augmented in an attractive–repulsive quartic van der Waals term [24] designed to combine the attractive well of the Lennard–Jones potential with the overall behavior, computational efficiency, and flexibility of the quartic repulsive potential. In addition to the quartic non-bonded terms are conventional empirical terms, such as Lennard–Jones, electrostatic, and dihedral angle potentials from the CHARMM (CHARMM19, CHARMM20) empirical energy function. Finally, a radius of gyration potential [11] can be used in combination with the repulsive vdw term (or the attractive–repulsive quartic term) to optimize internal packing. The radius of gyration is readily calculated for globular proteins from the number of residues, and for non-globular proteins, the protein can readily be subdivided into overlapping, approximately globular, segments [11].

2.1.3. NMR observables

A number of potential energy terms are included in Xplor-NIH which correspond to NMR observables. Of these terms, the most important is the NOE-derived interproton distance restraint term. Xplor-NIH supports multiple forms for the NOE potential energy, including harmonic, biharmonic [4], square-well [8], and a form containing an asymptotic cutoff for very large violations [23]. Corresponding to each NOE peak are two atom selections specifying the nuclei involved. An important feature of the NOE term is flexibility in the definition of the restraint distance r in the case of ambiguous or indistinguishable atoms. The center averaging method uses the distance between the centers of the two selections, and is appropriate for use with methyl groups and with non-stereospecifically assigned methylenes and aromatics when used with a pseudo-atom correction [25]. Also available are r^{-6} and r^{-3} averaging in which the distance is calculated as $\langle r^{-6}\rangle^{-1/6}$ and $\langle r^{-3}\rangle^{-1/3}$, respectively [7]. For ambiguous restraints it is best to use the sum averaging method, in which the distance is calculated as $[n\langle r^{-6}\rangle/n_{\mathrm{mono}}]^{-1/6}$, where n is the number of restraint distances specified by the selection and n_{mono} is the number of monomers [26].

Complete cross-validation for interproton distance restraints (as well as NOE intensity restraints) [27] and the automated iterative NOE assignment method ARIA [28,29] are also implemented in Xplor-NIH.

Complimentary to NOE distance restraint information are long-range orientational restraints which provide orientational information of a vector or atom group relative to an external axis system. These include residual dipolar couplings for fixed distance vectors [30,31] and vectors of variable length (e.g., ^{1}H–^{1}H vectors) [12], heteronuclear T_1/T_2 ratios for molecules that tumble anisotropically [32,33], and chemical shift anisotropy [19,34]. Currently, all the orientational restraints are treated similarly. The tensor orientation is represented by a tetratom pseudo-molecule, while the other parameters of the tensors need to be explicitly specified. In the case of the residual dipolar coupling potential simultaneous optimization of the magnitude of the alignment tensor is also available (C.D.S. and G.M.C., unpublished).

Additional potential terms for NMR observables include:

- Torsion angle restraints [8] derived from three-bond J couplings in combination with NOE/ROE data, available in square-well and harmonic forms.
- Three-bond J coupling constant restraints [35] which are related to torsion angles via Karplus relationships. Harmonic and square-well potentials are available, but it is best to use the former with the force constant chosen such that the rms difference between observed and calculated values matches experimental measurement error.
- $^1J_{\mathrm{C}_\alpha\text{–}\mathrm{H}_\alpha}$ coupling constant restraints [36]: these couplings are related to ϕ and ψ angles by an empirical relationship [37].
- Three-bond amide deuterium isotope effects on ^{13}C shifts: these are related to ψ angles by an empirical relationship that has the same functional form as a Karplus relationship [38], and can therefore be refined against in the same manner as ^3J coupling constants [39].
- $^{13}\mathrm{C}_\alpha/^{13}\mathrm{C}_\beta$ secondary shifts [40], which are empirically related to ϕ/ψ values [41].
- ^{1}H chemical shifts restraints [42,43], which include the capability of dealing with non-stereospecifically assigned methylene and methyl groups.
- Direct refinement against NOE intensities using complete relaxation matrix calculations [44,45]. This approach is compute intensive, and for this reason is used infrequently and then only in the very last stages of a refinement protocol.

2.1.4. Knowledge-based potentials of mean force

In addition to potential terms associated with NMR experiments, Xplor-NIH contains terms generated from high resolution structures present in the PDB. These potentials of mean force seek to bias conformational space sampling during simulated annealing to regions that are known to be physically realizable. These database potential terms are useful due to the limited nature of the structural restraints obtained from NMR, and serve as a substitute for realistic nonbonded

interactions. However, great care is taken during the construction of the database potentials such that they do not bias the structures away from the experimental restraints.

There are multidimensional torsion angle database potentials of mean force for proteins and nucleic acids, including raw potentials [24,46] as well as smoother potentials consisting of fitted Gaussians or quartic wells [13,18]. Examples of protein intraresidue data currently implemented include 2D (ϕ/ψ, χ_1/χ_2, χ_2/χ_3, χ_3/χ_4), 3D ($\phi/\psi/\chi_1$, $\chi_1/\chi_2/\chi_3$, $\chi_2/\chi_3/\chi_4$) and 4D ($\phi/\psi/\chi_1/\chi_2$, $\chi_1/\chi_2/\chi_3/\chi_4$) torsion angle correlations. In addition to intraresidue correlations, there are interresidue four-dimensional correlations of ϕ/ψ with ϕ/ψ of residue $i-1$, as well as residue $i+1$, which can be employed in cases of limited experimental data [46].

A separate set of potentials exists for groups that are close in space. These include the base-base positional potential for DNA [15,47]. These potentials are also available for RNA base-base interactions [48], protein sidechain-sidechain interactions, protein sidechain-DNA, and protein sidechain-RNA interactions [36]. For nucleic acids this leads to a significant increase in accuracy as judged by cross-validation and circumvents limitations associated with conventional representations of non-bonded contacts (e.g., the repel term tends to lead to expanded structures, and the Lennard-Jones term tends to lead to structural compression) [15,48].

2.1.5. Other potentials

There are also a set of generally useful potential terms. These include explicit planarity restraints, one application of which is to nucleic acid base pairs to allow propeller twisting without undue buckling. A non-crystallographic symmetry term is useful in treating homomultimeric structures (e.g., protein dimers, trimers, tetramers, and palindromic DNA). In addition, there is a potential term for direct refinement against crystallographic structure factors [10,49,50] and facilities are available to readily carry out joint NMR/Xray refinement calculations [51].

2.2. Parameter and topology information

Xplor-NIH is distributed with multiple parameter sets. These include parameter and topology sets for proteins and nucleic acids adapted for purely geometric refinements, which we recommend for NMR refinement. These parameters may be easily modified by end-users so as to ensure very small deviations from idealized covalent geometry, while satisfying experimental restraints and achieving good non-bonded contacts (i.e. no atomic overlaps). Full empirical energy functions (CHARMM19/20, OPLS [52]) are also available. Finally, it is straightforward to adapt parameter and topology files for any empirical energy function (e.g., AMBER [53], etc.) which employs the same analytic form for the electrostatics, hydrogen bonding and van der Waals Lennard-Jones potentials as CHARMM19/20. Note that obtaining parameters and topology information for other systems, including polysaccharides and small molecules, is relatively simple. Non-proton information can be generated from arbitrary molecular coordinates using the XPLO2D [54] program. Information relating to protons can then be easily added manually to these files. Similarly, parameter-only information can be generated using the learn facility of Xplor-NIH.

3. C++ framework

XPLOR [1] was written in FLECS, a dialect of FORTRAN77. We have translated this code to standard FORTRAN77 to remove a convoluted preprocessing stage, and to facilitate debugging. However, FORTRAN is not as expressive as more modern compiled languages, such as C++ or Java. FORTRAN suffers from a lack of language support for object-oriented techniques, poor character string support, and lack of standard dynamic memory management tools, making it quite cumbersome to develop new features, such as new potential energy terms. Most recent work on Xplor-NIH has employed the C++ as the compiled language. This language addresses the above deficiencies of FORTRAN without incurring a large performance penalty. To undertake the melding of code in a systematic fashion, we have developed a common framework for the C++ code analogous to that defined in XPLOR.

One important task of the C++ framework is to provide an efficient and convenient communication of structure information, such as atomic positions, to and from the FORTRAN framework when necessary. For instance, on the C++ side, if a value is modified, a flag is set signaling that copying is necessary. The FORTRAN side has no such mechanism, but if the C++ side detects that the size of an array (such as the number of atoms) has changed, it assumes that the XPLOR values have changed and will re-copy the whole array.

Within the C++ framework it is possible to define new potential energy terms, such as those associated with new NMR experiments. Such an energy term is defined as a C++ class. If one wishes this energy term to be evaluated, an instance of the class is created and added to a list of energy terms to be evaluated. So that these C++ energy terms are available to the XPLOR interface, a hook has been added to XPLOR's ENERGY function which calls all of the energy terms in the C++ energy list.

The C++ framework provides an atom selection type which stores selection strings and calls XPLOR's atom

selection routine. Thus, the C++ atom selection language maintains perfect fidelity with XPLOR's version.

Note that the C++ framework makes extensive use of a template library which borrows heavily from the standard C++ template library. In addition to classes for strings, regular expressions and container types such as lists, the library contains linear algebra components such as Matrix and Vector classes.

Some of the functionality in the C++ framework is directly accessible via an XPLOR language interface. Features accessible in this manner include the IVM [16] and the VMD interface [14]. But the preferred mechanism for interacting with the C++ framework is via the scripting interface described in the next section.

4. Scripting interface

In recent years, powerful interpreted general purpose scripting languages (GPSLs) such as Python, TCL and Perl have seen increasing popularity as a programming middle ground between a shell language, such as the Bourne or C shells, and compiled languages, such as FORTRAN, C++ and Java. These scripting languages feature a full set of tools for string and regular expression processing, container types (such as arrays), support for mathematical operations and access to system resources (like files and directories) in a rational, portable package. By contrast, shell languages are relatively slow, with their library consisting of external programs such as `sed`, `awk`, `cut`, and so on, providing unreliable interfaces. GPSLs have found use in applications ranging from web-page scripting to genome analysis. The popularity of these languages means that there are many packages implementing useful functionality, including the VMD biomolecular graphical viewer [55] and the NMRWish component of the NMRPipe spectrum analysis package [56]. These packages can be combined to further leverage each other's work. The fact that the GPSLs are standard, extensively documented, and widely known is an additional advantage.

The XPLOR language also ranks as a scripting language, but it lacks many features of a GPSL: in particular, simple string, regular expression, and mathematical handling. Also, it is not a full featured programming language in that it does not allow one to define functions, or provide object-oriented constructs. In the past, external programs such as `sed` and `awk` were used to manipulate XPLOR input files such as assignment tables (e.g., to convert NOE peak-pick tables from PIPP [57] to XPLOR restraint files; to separate out dipolar coupling restraint files into working and reference sets for cross-validation [58]). Since a GPSL is fully capable of performing these tasks, XPLOR outfitted with a GPSL interface would not require the external programs, and hence the structure determination process would be streamlined.

Being interpreted, the GPSLs lack the computational efficiency of a compiled language, and require a compiled language to perform compute-intensive calculations such as the evaluation of the full non-bonded potential energy term for a large system. Therefore an interface between the GPSL and the compiled language is necessary. For this purpose we employ the SWIG package [59], which semi-automatically generates the code needed to call C and C++ functions. Currently, SWIG is used to generate wrappers for the Python and TCL scripting languages. With the aid of SWIG, generating an interface for an additional GPSL can be achieved without much effort. In contrast, the scripting interface for the XPLOR scripting language is constructed manually in nested if-then loops which burden the FORTRAN code with considerable complexity.

4.1. Current features of the GPSL interface

The GPSL interface is currently under active development. Both Python and TCL interfaces have been implemented to maximize the number of external programs and tools which can be utilized in conjunction with Xplor-NIH. Developers can implement new features in whichever language they are more comfortable with, as the full basic functionality is available to both scripting languages. We expect to add new features shortly, but the basics of the existing interface will remain unchanged. Here we describe the current status.

Full inter-language scripting is provided to enable the widest possible reuse of existing scripts. For instance, one can execute arbitrary XPLOR or TCL commands from the Python interface, and likewise execute python commands from the XPLOR and TCL interfaces. Data can be interchanged between the scripting languages via various mechanisms. In interactive mode it is possible to jump from one interactive prompt to another. Thus, the full power of the XPLOR language is available from the GPSL interfaces.

Potential energy terms written in C++ are intended to be manipulated via the GPSLs. These terms are activated in the XPLOR interface by enabling the SCRIpting energy term within the FLAGs statement. This makes it possible to reuse existing XPLOR scripts with simple modifications to include new potential energy terms. Note that it is possible to code potential energy terms directly in the GPSL, although such terms would suffer a serious performance penalty relative to those coded in C++. A schematic diagram depicting the programmatic relationship among the various components of Xplor-NIH is shown in Fig. 1.

Currently, the following GPSL/C++ types include:

- The Simulation object which provides access to basic structure information. Through this object one can obtain atom coordinates, names, bonding information, and so on.

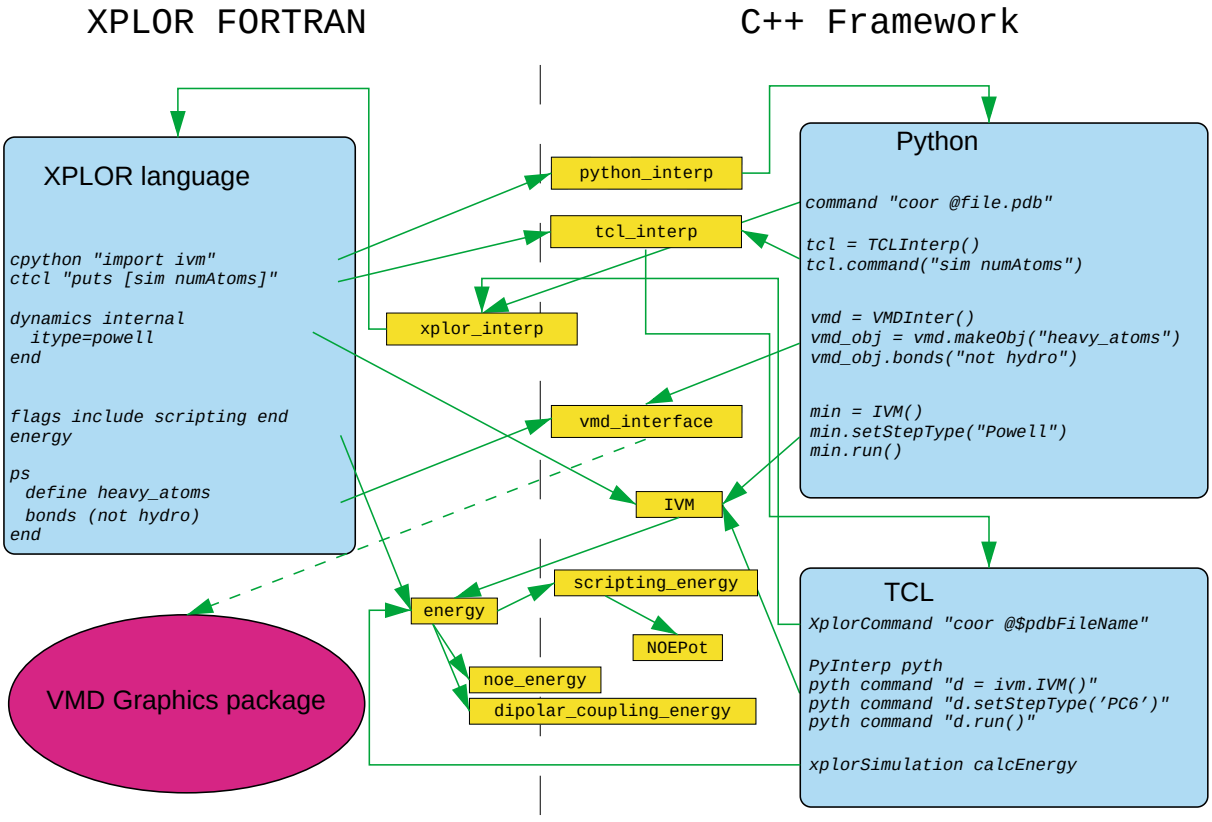

Fig. 1. Schematic of the programmatic organization of Xplor-NIH. The three blue boxes represent scripting language functionality, and contain representative script snippets. Yellow boxes represent functional computational units. Green arrowed lines denote the function call relationship between the script snippets and functional units. Items on the left of the dashed vertical line are implemented in FORTRAN77, while those on the right are primarily written in C++. (The Python and TCL languages are written in C.) The external VMD package [55] is shown as a magenta oval, and is connected via a dashed line, representing a socket communication interface.

- An IVM interface for molecular dynamics and minimizations. This is particularly useful for manipulations in rigid body and torsion angle coordinates.
- Access to the VMD interface. Graphical objects in VMD are accessed via instances of object entities within the GPSL interface.
- A new NOE potential term which is equivalent to XPLOR's AVERage = SUM potential [26], but which allows the following restraint distance average definition:

$$R = \left[\sum_{ij} r_{ij}^{-\alpha} / n_{\text{mono}} \right]^{-1/\alpha}, \quad (1)$$

where the sum is over all ambiguous nuclei, n_{mono} is the number of monomers, and α is a variable parameter which typically ranges from its definition in XPLOR of 6 down to 2. This definition enables tuning (reducing) the effective barrier between energy minima associated with each contribution to a set of ambiguous assignments.

- An AtomSel type which stores a selection string and associated atom indices. As covered in Section 3, these atom indices are generated by calling XPLOR's atom selection subroutine. Operations may be performed on atom selections using AtomSelAction functions to achieve the functionality of XPLOR's VECTor statement.
- An interface to generate reweighted atom density maps [60] without invoking VMD-XPLOR.

Global and module-specific interactive help functions are provided for all of the modules.

Figs. 2 and 3 contain scripting examples from Python and TCL, respectively. The corresponding commands in the XPLOR scripting language are included for comparison.

5. Regression testing

Xplor-NIH is distributed with a full suite of regression tests both in the source and binary-only packages. These tests are essential to obtain confidence that the package operates properly. For end users, possible problems can be caused by moving to a system with slightly different hardware or operating system configuration from those used for compilation. For developers, the slightest change in a base library might break

XPLOR	Python	description
	from ivm import IVM	load IVM module
dynamics internal	min = IVM ( )	create an IVM object
itype = powell	min.setStepType(powell)	choose minimization
depred = 0.0001	min.setDEpred(0.0001)	Prediction of initial energy drop
nstep = 1000	min.setNumSteps (1000)	Number of minimization steps
etol = 0.0000001	min.setETolerance(1e-7)	energy tolerance for exit condition
nprint = 50	min.setPrintInterval(50)	how often to print energy values
group = (resid 2:249)	min.setGroupList([AtomSel('resid 2:249'),	atom selections to group
group = (resid 301:385)	AtomSel ('resid 301:385')])	into rigid bodies
end	min.run( )	perform minimization

Fig. 2. Example scripts for performing rigid-body minimization in the XPLOR and Python scripting languages.

XPLOR	TCL	description
vector identify (store9) (name ca)	AtomSel aSel "name ca"	select the alpha carbons
foreach i in id (store9) loop main	foreach i [aSel indicies] {	loop through each selected atom
vector show elem (name) (id $i)	set curName [xplorSim getAtomName $i]	grab the current atom's name
eval ($curName = $result)		
vector show elem (resid) (id $i)	set curNum [xplorSim getResidueNum $i]	grab the current residue number
eval ($curNum = $result)		
display "name" $curName "res num" $curNum	puts [format "name %s res num %d" $curName $curNum]	print them out
end loop main	}	end of loop

Fig. 3. Example scripts for printing the atom name and residue number of a subset of atoms in the XPLOR and TCL scripting languages.

some feature. These tests provide some assurance that the package does indeed behave as it is expected to, and allows quick identification as to where an error is located. In the source package, we test components hierarchically: the C++ template library contains a test suite, as does the IVM and each potential term. The XPLOR and GPSL interfaces contain a large collection of test scripts along with the expected output. An automated procedure runs each script and reports discrepancies. These tests are essential to validate a new installation of Xplor-NIH.

6. Availability

Source and binary versions of Xplor-NIH are available from the URL http://nmr.cit.nih.gov/xplor-nih/ for noncommercial use. Commercial use of Xplor-NIH should be arranged with the Accelrys Corporation, but code developed outside the old XPLOR framework (the C++ interface including the IVM, etc.) are available to commercial and noncommercial entities. Precompiled binary executables are provided for multiple hardware/operating system platforms including Intel/Linux, Alpha/Linux, Alpha/OSF, Mips/Irix, Sparc/Solaris, PowerPC/Darwin (MacOS X). Where possible, these binaries have been compiled with optimizing compilers. Also provided with these distributions are a set of example scripts, supporting databases, and test suites.

Acknowledgments

It should be emphasized that many of the components of this package are the product of a large number of workers over almost two decades. In particular, the Xplor-NIH software package owes a great debt to Axel Brünger as the original creator of XPLOR. In addition, major contributions have been made by Michael Nilges and the original developers of CHARMM [3] and CNS [2]. We thank Dave Edwards, Scott Kahn and James Holden at Accelrys, and Mike Edwards and Pat Lake at NIH for their work arranging the Accelrys agreement.

References

[1] A.T. Brünger, XPLOR Manual Version 3.1, Yale University Press, New Haven, 1993. Available from <http://xplor.csb.yale.edu/xplor-info/>.

[2] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G. Warren, Crystallography and NMR System (CNS): a new software suite for macromolecular structure determination, Acta Cryst. Ser. D 54 (1998) 905–921.

[3] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM—a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[4] G.M. Clore, A.M. Gronenborn, A.T. Brünger, M. Karplus, The solution conformation of a heptadecapeptide comprising the DNA binding helix F of the cyclic AMP receptor protein of

Escherichia coli: combined use of ^{1}H-nuclear magnetic resonance and restrained molecular dynamics, J. Mol. Biol. 186 (1985) 435–455.

[5] L. Nilsson, G.M. Clore, A.M. Gronenborn, A.T. Brünger, M. Karplus, Structure refinement of oligonucleotides by molecular dynamics with NOE interproton distance restraints: application to 5′d$(CGTACG)_2$, J. Mol. Biol. 188 (1986) 455–475.

[6] A.T. Brünger, G.M. Clore, A.M. Gronenborn, M. Karplus, Three-dimensional structures of proteins determined by molecular dynamics with interproton distance restraints: application to crambin, Proc. Natl. Acad. Sci. USA 83 (1986) 3801–3805.

[7] G.M. Clore, A.T. Brünger, M. Karplus, A.M. Gronenborn, Application of molecular dynamics with interproton distance restraints to three-dimensional protein structure determination: a model study of crambin, J. Mol. Biol. 191 (1986) 523–551.

[8] G.M. Clore, M. Nilges, D.K. Sukumaran, A.T. Brünger, M. Karplus, A.M. Gronenborn, The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics, EMBO J. 5 (1986) 2729–2735.

[9] G.M. Clore, D.K. Sukumaran, M. Nilges, A.M. Gronenborn, The three-dimensional structure of phoratoxin in solution: combined use of nuclear magnetic resonance, distance geometry, and restrained molecular dynamics, Biochemistry 26 (1987) 1732–1745.

[10] A.T. Brünger, J. Kuriyan, M. Karplus, Crystallographic *R*-factor refinement by molecular-dynamics, Science 235 (1987) 458–460.

[11] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration, J. Am. Chem. Soc. 121 (1999) 2337–2338.

[12] N. Tjandra, J. Marquardt, G.M. Clore, Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules, J. Magn. Reson. 142 (2000) 393–396.

[13] J. Kuszewski, G.M. Clore, Source of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force, J. Magn. Reson. 146 (2000) 249–254.

[14] C.D. Schwieters, G.M. Clore, The VMD-XPLOR visualization package for NMR structure refinement, J. Magn. Reson. 149 (2001) 239–244.

[15] J. Kuszewski, C.D. Schwieters, G.M. Clore, Improving the accuracy of NMR structures of DNA by means of a database potential of mean force describing base–base positional interactions, J. Am. Chem. Soc. 123 (2001) 3903–3918.

[16] C.D. Schwieters, G.M. Clore, Internal coordinates for molecular dynamics and minimization in structure determination and refinement, J. Magn. Reson. 152 (2001) 288–302.

[17] G.M. Clore, C.A. Bewley, Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings, J. Magn. Reson. 154 (2002) 329–335.

[18] G.M. Clore, J. Kuszewski, χ_1 rotamer populations and angles of mobile surface sidechains are accurately predicted by a torsion angle database potential of mean force, J. Am. Chem. Soc. 124 (2002) 2866–2867.

[19] Z.R. Wu, N. Tjandra, A. Bax, ^{31}P chemical shift anisotropy as an aid in determining nucleic acid structure in liquid crystals, J. Am. Chem. Soc. 123 (2001) 3617–3618.

[20] J. Kuszewski, M. Nilges, A.T. Brünger, Sampling and efficiency of metric matrix distance geometry—a novel partial metrization algorithm, J. Biomol. NMR 2 (1992) 33–56.

[21] M. Nilges, G.M. Clore, A.M. Gronenborn, Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations, FEBS Lett. 229 (1988) 317–324.

[22] J.P. Linge, M. Nilges, Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation, J. Biomol. NMR 13 (1999) 51–59.

[23] M. Nilges, A.M. Gronenborn, A.T. Brünger, G.M. Clore, Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor, and barley serine proteinase inhibitor 2, Protein Eng. 2 (1988) 27–38.

[24] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases, Protein Sci. 5 (1996) 1067–1080.

[25] K. Wüthrich, M. Billeter, W. Braun, Pseudo-structures for the 20 common amino-acids for use in studies of protein conformations by measurements of intramolecular proton proton distance constraints with nuclear magnetic-resonance, J. Mol. Biol. 169 (1983) 949–961.

[26] M. Nilges, A calculational strategy for the solution structure determination of symmetric dimers by ^{1}H-NMR, Proteins 17 (1993) 297–309.

[27] A.T. Brünger, G.M. Clore, A.M. Gronenborn, R. Saffrich, M. Nilges, Assessing the quality of solution nuclear magnetic resonance structures by complete cross-validation, Science 261 (1993) 328–331.

[28] M. Nilges, M.J. Macias, S.I. O'Donoghue, S. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin, J. Mol. Biol. 269 (1997) 408–422.

[29] M. Nilges, S.I. O'Donoghue, Ambiguous NOEs and automated NOE assignment, Prog. NMR Spectrosc. 32 (1998) 107–139.

[30] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, Use of dipolar ^{15}N–^{1}H and ^{13}C–^{1}H couplings in the structure determination of magnetically oriented macromolecules in solution, Nat. Struct. Biol. 4 (1997) 732–738.

[31] G.M. Clore, A.M. Gronenborn, N. Tjandra, Direct refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude, J. Magn. Reson. 131 (1998) 159–162.

[32] N. Tjandra, D.S. Garrett, A.M. Gronenborn, A. Bax, G.M. Clore, Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy, Nat. Struct. Biol. 4 (1997) 443–449.

[33] G.M. Clore, A.M. Gronenborn, A. Szabo, N. Tjandra, Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information, J. Am. Chem. Soc. 120 (1998) 4889–4890.

[34] R.S. Lipsitz, N. Tjandra, Carbonyl CSA restraints from solution NMR for protein refinement, J. Am. Chem. Soc. 123 (2001) 11065–11066.

[35] D.S. Garrett, J. Kuszewski, T.J. Hancock, P.J. Lodi, G.W. Vuister, A.M. Gronenborn, G.M. Clore, The impact of direct refinement against three-bond HN–CαH coupling constants on protein structure determination by NMR, J. Magn. Reson. Ser. B 104 (1994) 99–103.

[36] J. Kuszewski, G.M. Clore, unpublished.

[37] G.W. Vuister, F. Delaglio, A. Bax, The use of $^{1}JC_{\alpha}H_{\alpha}$ coupling-constants as a probe for protein backbone conformation, J. Biomol. NMR 3 (1993) 67–80.

[38] M. Ottiger, A. Bax, An empirical correlation between amide deuterium isotope effects on $^{13}C(\alpha)$ chemical shifts and protein backbone conformation, J. Am. Chem. Soc. 119 (1997) 8070–8075.

[39] M. Caffrey, M. Cai, J. Kaufman, S.J. Stahl, P.T. Wingfield, D.G. Covell, A.M. Gronenborn, G.M. Clore, Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41, EMBO J. 17 (1998) 4572–4584.

[40] J. Kuszewski, J. Qin, A.M. Gronenborn, G.M. Clore, The impact of direct refinement against $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts on protein structure determination by NMR, J. Magn. Reson. Ser. B 106 (1995) 92–96.

[41] S. Spera, A. Bax, Empirical correlation between protein backbone conformation and C_α and C_β ^{13}C nuclear-magnetic-resonance chemical shifts, J. Am. Chem. Soc. 113 (1991) 5490–5492.
[42] J. Kuszewski, A.M. Gronenborn, G.M. Clore, The impact of direct refinement against proton chemical shifts in protein structure determination by NMR, J. Magn. Reson. Ser. B 107 (1995) 293–297.
[43] J. Kuszewski, A.M. Gronenborn, G.M. Clore, A potential involving multiple proton chemical shift restraints for non-stereospecifically assigned methyl and methylene protons, J. Magn. Reson. Ser. B 112 (1996) 79–81.
[44] P. Yip, D.A. Case, A new method for refinement of macromolecular structures based on nuclear overhauser effect spectra, J. Magn. Reson. 83 (1989) 643–648.
[45] M. Nilges, J. Habazettl, A.T. Brünger, T.A. Holak, Relaxation matrix refinement of the solution structure of squash trypsin-inhibitor, J. Mol. Biol. 219 (1991) 499–510.
[46] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids, J. Magn. Reson. 125 (1997) 171–177.
[47] E.C. Murphy, V.B. Zhurkin, J.M. Louis, G. Cornilescu, G.M. Clore, Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occuring point mutation, J. Mol. Biol. 312 (2001) 481–499.
[48] G.M. Clore, J. Kuszewski, Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation, J. Am. Chem. Soc. (in press).
[49] A.T. Brünger, Crystallographic refinement by simulated annealing, in: N.W. Isaacs, M.R. Taylor (Eds.), Crystallographic Computing 4: Techniques and New Technologies, Clarendon Press, Oxford, 1988, pp. 126–140.
[50] A.T. Brünger, Free *R*-value—a novel statistical quantity for assessing the accuracy of crystal-structures, Nature 355 (1992) 472–475.
[51] B. Shaanan, A.M. Gronenborn, G.H. Cohen, G.L. Gilliland, B. Veerapandian, D.R. Davies, G.M. Clore, Combining experimental information from crystal and solution studies: joint X-ray and NMR refinement, Science 257 (1992) 961–964.
[52] W.L. Jorgensen, J. Tirado-Rives, The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin, J. Am. Chem. Soc. 110 (1988) 1666–1671.
[53] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force-field for molecular mechanical simulation of nucleic-acids and proteins, J. Am. Chem. Soc. 106 (1984) 765–784.
[54] G.J. Kleywegt, J.Y. Zou, M. Kjeldgaard, T.A. Jones, Around O, in: M.G. Rossmann, E. Arnold (Eds.), International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules, Kluwer Academic Publishers, Dordrecht, 2001, 366–367. Available from <http://xray.bmc.uu.se/~gerard/manuals/xplo2d_man.html>, pp. 353–356.
[55] W. Humphrey, A. Dalke, K. Schulten, VMD-visual molecular dynamics, J. Mol. Graphics 14 (1996) 33–38. Available from <http://www.ks.uiuc.edu/Research/vmd/>.
[56] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe—a multidimensional spectral processing system based on Unix pipes, J. Biomol. NMR 6 (1995) 277–293.
[57] D.S. Garrett, R. Powers, A.M. Gronenborn, G.M. Clore, A common sense approach to peak picking two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams, J. Magn. Reson. 95 (1991) 214–220.
[58] G.M. Clore, D.S. Garrett, *R*-factor, Free *R* and complete cross-validation for dipolar coupling refinement of NMR structures, J. Am. Chem. Soc. 121 (1999) 9008–9012.
[59] . Available from <http://www.swig.org/>.
[60] C.D. Schwieters, G.M. Clore, Reweighted atomic densities to represent ensembles of NMR structures, J. Biomol. NMR 23 (2002) 221–225.