Optimal isotope labelling for NMR protein structure determinations

Masatsune Kainosho¹, Takuya Torizawa¹, Yuki Iwashita¹, Tsutomu Terauchi¹, Akira Mei Ono¹ & Peter Güntert²

Nuclear-magnetic-resonance spectroscopy can determine the three-dimensional structure of proteins in solution. However, its potential has been limited by the difficulty of interpreting NMR spectra in the presence of broadened and overlapping resonance lines and low signal-to-noise ratios. Here we present stereo-array isotope labelling (SAIL), a technique that can overcome many of these problems by applying a complete stereospecific and regiospecific pattern of stable isotopes that is optimal with regard to the quality and information content of the resulting NMR spectra. SAIL uses exclusively chemically and enzymatically synthesized amino acids for cell-free protein expression. We demonstrate for the 17-kDa protein calmodulin and the 41-kDa maltodextrin-binding protein that SAIL offers sharpened lines, spectral simplification without loss of information, and the ability to rapidly collect the structural restraints required to solve a high-quality solution structure for proteins twice as large as commonly solved by NMR. It thus makes a large class of proteins newly accessible to detailed solution structure determination.

In the past two decades, nuclear-magnetic-resonance spectroscopy has become one of the two accepted methods (along with X-ray crystallography) for determining three-dimensional structures of proteins. NMR spectroscopy provides information about the structure and dynamic properties of proteins in solution, and offers an approach for determining the three-dimensional protein structures of systems that fail to crystallize. NMR currently provides about 15% of the protein structures in the Protein Data Bank and has a role in structural genomics⁵, but complete automation of the structure determination process or the structural analysis of proteins with a molecular mass greater than 25 kDa have not yet become routine with NMR. Conventional uniform labelling of proteins with¹³C and¹⁵N, coupled with double and triple resonance, two-dimensional to four-dimensional NMR data collection, supports the determination of NMR solution structures of proteins, in favourable cases as large as 25 kDa (refs 2, 3). Cryogenic probes are used to improve the signal-to-noise ratio, and higher-field magnets provide increased resolution and a further gain in sensitivity. However, as molecular masses increase, NMR spectra become increasingly difficult to interpret because of spectral crowding and line broadening due to fast transverse relaxation. Spin-diffusion effects decrease the ability to determine inter-proton distances from nuclear Overhauser enhancement (NOE) data. A common approach for addressing these problems is to label proteins with deuterium to simplify the spectra and to minimize spin-diffusion effects⁶,⁷. Data collection by transverse relaxation optimized spectroscopy (TROSY)⁸ provides sharper lines for amide and aromatic groups. With these methods the global folds of a limited number of proteins larger than 30 kDa could be derived on the basis of conformational restraints for amide and selected methyl groups⁷. However, conformational data from other aliphatic or aromatic groups and thus for most of the side chains remain difficult to collect for proteins larger than 25 kDa. Consequently, by June 2005 only 1% of all NMR protein structures deposited in the Protein Data Bank were for proteins with a molecular mass of more than 25 kDa, and backbone and side-chain chemical shift assignments more than 70% complete had been recorded in the BioMagResBank (http://www.bmrb.wisc.edu) for only eight proteins larger than 25 kDa.

It has long been recognized that deuteration can be used to simplify NMR spectra⁶, to obviate the need for chiral assignments⁹, to facilitate the measurement of spin–spin⁹ and dipolar couplings⁷, and to increase resolution⁶. However, labelling patterns and the approaches for achieving them have so far been suboptimal. Random fractional deuteration methods⁷ suffer from the production of
numerous isotopomer proteins with chemical shift heterogeneity and decreased signal intensities. Perdeuteration methods remove all carbon-bound protons such that much potential NOE information is lost. Methods for introducing methyl and/or aromatic protons into a perdeuterated background improve this situation, but the additional structural information is localized and unevenly distributed. The same holds true for residue-selective and segmental labeling. As a consequence, the quality of three-dimensional protein structures based on these methods remains limited. As an alternative presented here, optimal labeling patterns for protein NMR can be realized by the chemical or enzymatic synthesis of amino acids followed by in vitro (cell-free) protein expression to build a protein exclusively from such amino acids. Cell-free methods are crucial to making efficient use of the labeled amino acids and to the prevention of scrambling of the label, which would occur through metabolic pathways present in cell-based protein expression systems.

**Stereo-array isotope labeling (SAIL)**

The basic strategy of the SAIL approach is to prepare amino acids with the following features (Fig. 1): first, stereo-selective replacement of one \(^1\)H in methylene groups by \(^2\)H; second, replacement of two \(^1\)H in each methyl group by \(^3\)H; third, stereo-selective modification of the prochiral methyl groups of Leu and Val such that one methyl is \(-^{13}\)C(\(^2\)H)\(^3\) and the other is \(-^{12}\)C(\(^1\)H)(\(^2\)H)\(^2\); and last, labeling of six-membered aromatic rings by alternating \(^1\)C–\(^2\)H and \(^1\)C–\(^1\)H couplings, and decreased signal intensities. Perdeuteration methods remove all carbon-bound protons such that much potential NOE information is lost. Methods for introducing methyl and/or aromatic protons into a perdeuterated background improve this situation, but the additional structural information is localized and unevenly distributed. The same holds true for residue-selective and segmental labeling. As a consequence, the quality of three-dimensional protein structures based on these methods remains limited. As an alternative presented here, optimal labeling patterns for protein NMR can be realized by the chemical or enzymatic synthesis of amino acids followed by in vitro (cell-free) protein expression to build a protein exclusively from such amino acids. Cell-free methods are crucial to making efficient use of the labeled amino acids and to the prevention of scrambling of the label, which would occur through metabolic pathways present in cell-based protein expression systems.

Table 1 | Expected and observed features of CaM and MBP samples prepared by conventional uniform labeling and SAIL methods

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Uniform labelling</th>
<th>CaM</th>
<th>SAIL</th>
<th>Uniform labelling</th>
<th>MBP</th>
<th>SAIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^1)H atoms per molecule in (^2)H(^2)O</td>
<td>1,095</td>
<td>697 (64%)</td>
<td>2,860</td>
<td>1,802 (63%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^1)H atoms per molecule in (^3)H(^4)O*</td>
<td>851</td>
<td>453 (53%)</td>
<td>2,249</td>
<td>1,911 (53%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-exchangeable side-chain (^1)H atoms</td>
<td>692</td>
<td>305 (44%)</td>
<td>1,850</td>
<td>821 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE cross-peaks expected†</td>
<td>9,812</td>
<td>5,642 (58%)</td>
<td>17,076</td>
<td>9,382 (55%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE cross-peaks used</td>
<td>–</td>
<td>4,576 (47%)</td>
<td>–</td>
<td>7,485 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE distance restraints expected‡</td>
<td>2,883</td>
<td>2,720 (94%)</td>
<td>4,293</td>
<td>4,347 (101%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE distance restraints used</td>
<td>–</td>
<td>2,422 (84%)</td>
<td>–</td>
<td>3,818 (89%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†NOE cross-peaks are assumed to be observable between all aliphatic, aromatic, backbone amide and Asn/Gln side-chain amide \(^1\)H nuclei that are less than 4.75 A\(^\circ\) apart in the 20 conformers of the SAIL-CaM solution structure, or less than 4.2 A\(^\circ\) apart in the 20 conformers of the SAIL-MBP solution structure.

‡Excluding duplicate restraints from symmetry-related peak pairs and NOEs for fixed distances. For uniform labeling, pseudo-atoms were assumed.

**SAIL versus uniform labeling with \(^13\)C and \(^15\)N**

On the basis of calculations (Table 1), we expected that the NMR spectra of SAIL proteins would be simpler than those of the corresponding uniformly labeled proteins. The numbers of observable protons are reduced by SAIL relative to uniform labeling. The SAIL approach reduces the number of non-exchangeable side-chain protons, which are prone to overlap but essential for defining the side-chain conformations, to less than half, and decreases the number of expected NOESY cross-peaks by 40–45%. Most of the additional NOEs from uniformly labeled proteins either involve fixed (geminal) distances or become redundant in the absence of stereo-specific assignments and thus contribute to spectral overlap without furnishing independent information. In principle, stereo-specific assignments could allow slightly more precise structures in the case of uniform labeling. However, in practice only a limited number of stereospecific assignments can be made. The BioMagResBank reports stereospecific assignments for 7% of the methylene protons and Val/Leu methyl groups, and for only 3.4% of the prochiral groups in proteins larger than 20 kDa. Therefore, in general the expected number of non-redundant, structurally relevant NOE restraints is retained with SAIL (Table 1), even if the effect of the better signal-to-noise ratio with SAIL is neglected. These numbers are corroborated by the experimental findings (Supplementary Table 1). More detailed theoretical considerations (see Methods), which take into account the enhanced signal strength and sharper lines in SAIL spectra and the fact that overlap can render peaks unidentifiable, show that in practice SAIL is expected to increase rather than decrease the number of identifiable NOE cross-peaks. The expected increase is moderate in regions without overlap but is significant in regions with strong overlap and therefore for larger proteins, for which SAIL is expected to yield two or more times the number of relevant conformational restraints than uniform labeling.

Precautions to be taken when collecting and analyzing data for a SAIL protein are that deuterium decoupling should be applied during \(^13\)C evolution times and that what normally are methyl and methylene groups should be treated as methine groups. \(^1\)H–\(^13\)C CT-HSQC (constant-time heteronuclear single-quantum coherence) NMR data sets collected from SAIL and uniformly labeled proteins demonstrate the superiority of the SAIL method. Each \(^1\)H–\(^13\)C pair in the protein is associated with a peak in the spectrum. Severe signal overlap in uniform labeling is alleviated in the spectrum from the SAIL proteins. The improvements, which are particularly apparent for larger proteins, result from the decrease in the number of \(^1\)H signals as well as from sharpening of the remaining signals.

The SAIL method also improves sensitivity. Part of the gain arises from longer \(^1\)H and \(^13\)C transverse relaxation times resulting from replacements of \(^1\)H by \(^2\)H. Reduced relaxation during magnetization transfer steps in experiments such as \(^1\)H–\(^13\)C CT-HSQC leads to an increased signal-to-noise ratio. Reduced long-range couplings result in a further sharpening of signals. The signal intensities for methylene...
groups are threefold to sevenfold higher with SAIL than with uniform labelling under the same conditions. Improvements are more pronounced for the 41-kDa SAIL-MBP protein (Fig. 3e) than for the 17-kDa SAIL-CaM (Fig. 2i). Although each observed SAIL methyl group contained only one $^1$H in comparison with three equivalent $^1$H with uniform labelling, equivalent signal intensities were observed (data not shown) as a result of the longer $^1$H and $^{13}$C transverse relaxation times.

**SAIL protein structure determinations**

All peaks benefited from improved sensitivity and resolution and allowed the signals of SAIL-CaM and SAIL-MBP to be assigned readily by established methods. Side-chain assignments were determined completely from the analysis of two data sets: HCCH-TOCSY (total correlation spectroscopy) data provided connectivities among all side-chain signals, and HCCH-COSY (correlation spectroscopy) data were used to identify the spin systems. The detection and assignment of signals from aromatic rings containing alternating $^{12}$C and $^{13}$C nuclei (Fig. 1) is straightforward by an unconventional approach that we describe separately. The expected $^1$H, $^{13}$C and $^{15}$N chemical shifts could be assigned without exception for SAIL-CaM, and to 94% for SAIL-MBP, including more than 90% of the aliphatic and aromatic side-chain protons (Supplementary Table 1). Many of the shifts that could not be assigned are in the region of residues 229–241, which have been shown to interact with the bound cyclodextrin. Because of conformational heterogeneity, most of the expected resonances of residues 229–241 could also not be assigned in earlier studies.

We obtained distance restraints for the structure calculation from three-dimensional $^{15}$N- and $^{13}$C-edited NOESY spectra and from a two-dimensional NOESY spectrum for the aromatic region. These spectra were simplified by a decreased number of signals, as expected (Table 1). Because of lower spin diffusion, maximum NOE intensities for SAIL proteins were typically reached at mixing times 1.5-fold to 3.0-fold those for uniform labelling (data not shown). On the basis of the simplified and more quantitative NOESY spectra, we were able to obtain structures of SAIL-CaM and SAIL-MBP by means of the combined automated NOE assignment and structure calculation protocol in the program CYANA. (Supplementary Table 1). For instance, a dense network of NOEs including 949 non-redundant, long-range distance restraints was established for SAIL-MBP. Of the 3,818 non-redundant NOE distance restraints, 1,879 involve side-chain atoms beyond Hb. The SAIL-CaM and SAIL-MBP solution structures show good quality in terms of the agreement with the experimental data and other validation parameters.

Structures of the calcium-bound form of calmodulin have been determined previously by crystallography and NMR. The NMR structure was based on residual dipolar couplings (RDCs),

---

**Figure 2** $^1$H-$^{13}$C CT-HSQC spectra of CaM. a, SAIL-CaM, aliphatic region. b, SAIL-CaM, methyl region. c, SAIL-CaM, methylene region. d, UL-CaM, aliphatic region. e, UL-CaM, methyl region. f, UL-CaM, methylene region. g, SAIL-CaM, Arg δ region. h, UL-CaM, Arg δ region. i, Cross-sections from g and h. The spectra for SAIL-CaM and UL-CaM were recorded under identical conditions and scaled for equal noise levels. Assignments are indicated by one-letter amino-acid code, residue number and atom identifier. Assignments of UL-CaM are as reported previously.
measured mainly for the polypeptide backbone, and on the crystal structure. Our NMR structure, the first one for calcium-bound calmodulin based exclusively on NOEs, provides detailed information on the side-chain conformations. It is in close agreement with the crystal structure and the RDC-based backbone structure (Fig. 4a).

In addition, the solution structure of the 41-kDa SAIL-MBP coincides closely with the crystal structure27 determined previously under slightly different conditions (Fig. 4b, c, and Supplementary Table 1). The 41-kDa SAIL-MBP solution structure is of similar precision and accuracy to those of smaller proteins, and the structural statistics are comparable to those commonly found in NMR structure determinations of smaller proteins. Previously7, a global fold of MBP was determined by NMR on the basis of NOEs between amide and methyl protons, residual dipolar couplings for the polypeptide backbone, and hydrogen-bond restraints. That NMR study could provide a good determination of the global fold of the polypeptide backbone and the conformations of the methyl-containing side chains of valine, leucine and isoleucine, but the approach used cannot provide direct structural information on the other side chains. Their conformations therefore remained largely undetermined, resulting in root-mean-square deviations (RMSDs) of more than 3.8 Å for all side-chain heavy atoms of the amino-terminal and carboxy-terminal domains of MBP. The corresponding

Figure 3 | 1H–13C CT-HSQC spectra of MBP. a, Aliphatic region of CDH groups in SAIL-MBP. b, Enlargement of the rectangular region marked in a. Assignments are indicated with one-letter amino-acid code, residue number and atom identifiers. c, d, Corresponding regions for UL-MBP. e, Cross-sections taken at the positions indicated in b and d. The spectra for SAIL-MBP and UL-MBP were recorded under identical conditions and scaled for equal noise levels.

Figure 4 | CaM and MBP solution and crystal structures. a, SAIL-CaM (backbone in cyan, Ca2+ in white), CaM X-ray structure31 (red), and three solution conformers of UL-CaM determined from residual dipolar coupling data32 (blue). b, MBP solution and crystal structures32. Backbone of the N-terminal domain (SAIL-MBP in green, X-ray in red) and the C-terminal domain (SAIL-MBP in blue, X-ray in red). c, Aromatic side chains (SAIL-MBP in green and blue, X-ray in red), and backbone ribbon representation of the X-ray structure (gold). Superpositions of CaM and MBP solution conformers on the X-ray structures were performed separately for the two flexibly connected domains.
side-chain RMSDs for the SAIL-MBP structure are 2.3 Å for both domains.

Discussion

The SAIL strategy described here is expected to support high-throughput protein structure determination without loss of structural quality. The much smaller numbers of $^1$H shifts that need to be assigned make SAIL proteins particularly amenable to automated resonance assignment, which also benefits from the complete absence of uncertainties associated with the lack of stereospecific assignments and frequent accidental chemical shift degeneracies for diasteroretic pairs in uniformly labelled proteins. SAIL improves the quality of virtually all commonly used multidimensional NMR spectra and can be used in conjunction with other techniques to improve the sensitivity of NMR experiments, such as TROSY, cryogenic probes and high-field magnets. Our structure determination of MBP shows that high-quality structure solutions for proteins up to at least 40 kDa can now be solved by NMR.

METHODS

Expected numbers of identifiable NOE signals with peak area and integral

For MBP with a SAIL-to-UL peak height ratio of $h / H$ = 4–7 (Fig. 3c) and a ratio $N_0 / N = 0.55$ (Table 1) for the number of NOEs below 4.2 Å, SAIL is expected to increase the number of identifiable NOEAs by at least 10% for regions without overlap ($N_{obs}/1 = 0$) and by more than 50% for a crowded region with $N_{obs}/1 = 0.25$.

NMR spectroscopy

The SAIL-MaM and UL-MaM samples each contained 0.33 mM protein, 3.3 mM HEPES, pH 7.2. NMR experiments were performed at 300 MHz and 800 MHz (Varian, Santa Clara, CA, and Bruker, Billerica, MA, respectively) using 2D NOESY-HSQC, 2D NOESY-ROESY, 3D NOESY-HSQC, 3D NOESY-ROESY, 3D ROESY-HSQC and 3D ROESY-ROESY experiments.

Structure calculation

The SAIL-MaM and SAIL-MBP structures were obtained with the program CYANA[22] with the use of automated NOE assignment and torsion angle dynamics for the structure calculation, which was started from 100 (CaM) or 200 (MBP) conformers with random torsion angle values. The standard CYANA-simulated annealing schedule was applied with 10,000 (CaM) or 20,000 (MBP) torsion angle dynamics steps each. Backbone torsion angle restraints obtained from chemical shifts with the program TALOS[24] were added to the input for CYANA. Hydrogen-bond restraints were not used. The 20 conformers with the lowest final CYANA target function values were embedded in a water shell of 8 Å thickness and energy-minimized against the Amber force field[25] with the program OPLS[26] in the presence of the NOE distance restraints as the only experimental data.

Received 8 September; accepted 14 December 2005.


Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Ikura and T. Yamazaki for providing the calmodulin and maltodextrin-binding protein genes, respectively, and M. Takeda for help with the preparation of figures. This work was supported by CREST/JST.

Author Information Atomic coordinates of the SAIL-CaM and SAIL-MBP structures have been deposited in the Protein Data Bank with accession codes 1X02 and 2D21, respectively. Chemical shifts have been deposited in the BioMagResBank with accession numbers 6541 and 6807. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (kainosho@nmr.chem.metro-u.ac.jp).
New Tools Provide New Insights in NMR Studies of Protein Dynamics

Anthony Mittermaier1 and Lewis E. Kay2

There is growing evidence that structural flexibility plays a central role in the function of protein molecules. Many of the experimental data come from nuclear magnetic resonance (NMR) spectroscopy, a technique that allows internal motions to be probed with exquisite time and spatial resolution. Recent methodological advancements in NMR have extended our ability to characterize protein dynamics and promise to shed new light on the mechanisms by which these molecules function. Here, we present a brief overview of some of the new methods, together with applications that illustrate the level of detail at which protein motions can now be observed.

NMR spectroscopy is an experimental tool developed over half a century ago by physicists who were interested in exploring fundamental properties of matter. They could have hardly imagined the wide utility of their tool developed over half a century ago. Because the frequency of the magnetic energy absorbed by each nucleus depends on its chemical environment, a given probe will likely have distinct chemical shifts in each conformation, separated by Δν (Fig. 1A). If the exchange rate, \( k_{ex} \), is not much smaller than \( 2\pi\Delta\nu \), peaks derived from the weakly populated conformer (excited state) are most often not observed, because the transient nature of this state leads to substantial peak broadening. As a result, a spectrum is obtained (Fig. 1B) where for each probe a peak is observed only from the more populated state, slightly shifted from its position in the absence of a slow exchange limit. How, then, does one obtain information about the excited state when it is essentially invisible in NMR spectra?

One way is to use an experimental approach, based on an idea from Erwin Hahn in the 1950s, called a spin echo. The basic phenomenon can be explained as follows. Imagine that a group of runners, composed of both slow and fast individu-
uals, start a race at the same time. If at the halfway point of the race all runners are made to stop, turn around, and run back to the starting position, it is clear that both slow and fast individuals will cross the starting line at the same time, giving rise to a spontaneous ordering referred to as an echo. Now suppose that the runners can interconvert stochastically during the course of the race, in the sense that a slow runner can become fast and vice versa (corresponding to molecules interconverting between a pair of states). Although the positions of the runners at the end will depend on the details of the exchange process, in general they will not all finish the race simultaneously, and a plot of the distribution of runners crossing the starting line versus time gives rise to a peak that is broader than that observed in the absence of exchange. The breadth of the peak provides information about the relative numbers of fast and slow runners (thermodynamics), their rate of interconversion (kinetics), and the difference in running speeds between the two groups (analogous to chemical shift differences between exchanging conformations, which relate to structure). The basic NMR experiment [so-called Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (3)] builds on this concept by applying a variable number of refocusing pulses during a fixed time interval, where each pulse corresponds to runners stopping and turning around in the analogy above. If the rate of application of pulses is such that the runners are allowed to get out of phase with each other, then the exchange process leads to broad lines as described above (Fig. 1C, top). In contrast, when the pulses are applied at a fast rate the effects of the interconversion are reduced and narrower peaks are obtained (Fig. 1C, bottom). The dependence of line width on pulsing rate (the relaxation dispersion profile) is subsequently fit to extract the exchange parameters (3).

Fig. 2. (A) Experimental relaxation dispersion profiles for amide $^{1}$H-15N single-quantum (SQ), double-quantum (DQ), and multiple-quantum (MQ) coherences of Gly142 in the G48M mutant of the Fyn SH3 domain (8). Fits to a global three-site model of exchange are indicated by lines. (Inset) An energy level diagram for a $^{1}$H-15N spin pair with SQ, DQ, MQ coherences indicated by black, blue, and sum of red lines, respectively. (B) The 15N chemical shifts of the G48M U state calculated from relaxation dispersion data plotted as a function of corresponding random coil chemical shifts. (C) Deviations from unfolded (U) state backbone chemical shifts, $\frac{\Delta \sigma_{\text{U}(15N)}}{\Delta \sigma_{\text{random coil}(15N)}}$ + $\frac{\Delta \sigma_{\text{U}(15N)}}{\Delta \sigma_{\text{random coil}(15N)}}$, calculated in absolute frequency units (800-MHz spectrometer) for intermediate (I, solid blue) and folded (F, dashed red) states, plotted as a function of residue number (8). (D) Structure of the wild-type Fyn SH3 domain (1SHF) with the location of Gly48 indicated in red.

As mentioned above, the basic CPMG scheme dates back to the formative days of NMR spectroscopy. However, the routine application of the method to proteins had to await a number of advances. One such advance was the development of multidimensional spectroscopy, where the great majority of sites in the protein can, in principle, be visualized in the form of cross peaks in $^{1}$H-15N or $^{1}$H-13C correlation spectra. Such spectra provide the site-specific information that distinguishes modern NMR from so many other physical techniques. The second breakthrough was due to Loria and Palmer, who developed a clever scheme for allowing slow dynamical processes to be characterized without interference from interactions between the nucleus of interest and directly bonded NMR-active nuclei (3). CPMG relaxation dispersion experiments most often make use of backbone amide $^{1}$H or $^{15}$N spin probes (4–6), focusing on proton or nitrogen magnetization (single-quantum transitions), respectively. However, in the context of a $^{1}$H-15N spin pair, there are additional transitions that can be explored to gain further insight into the dynamics. A comparison of experimental dispersion profiles for the same site but derived from experiments that monitor different transitions (Fig. 2A) emphasizes that the same exchange process can lead to very different dispersion shapes (7, 8). Here $R_{2,\text{eff}}$ (peak line width) is plotted along y, versus $v_{\text{CPMG}}$ (proportional to the number of refocusing pulses applied) along x.

CPMG relaxation dispersion experiments in which exchange effects are quenched through the application of pulses are often the method of choice for the study of millisecond dynamics in proteins (3). A similar quenching effect can be achieved by applying a continuous radio-frequency field that is allowed to vary both in magnitude and in frequency (3), and this approach is used to study faster processes ($k_{\text{ex}}$ up to $\approx 100,000\ s^{-1}$).

In order to illustrate the utility of the dispersion methodology, we focus on a single application in the area of protein folding that makes use of the CPMG method (9); additional studies can be found in the literature, with the work of Kern and colleagues relating protein dynamics to enzyme catalysis being particularly noteworthy (10–14). The present example concerns several Gly48 mutants of the Fyn SH3 domain that had been shown with use of stopped-flow fluorescence denaturation and renaturation measurements to fold with a two-state mechanism, $F \frac{k_{\text{ex}}}{k_{\text{U}}} U$. The exchange parameters obtained by using fluorimetry are well within the ranges of $k_{\text{ex}}$ values (several hundred to a few thousand s$^{-1}$) and excited state populations ($p_{\text{U}} > \sim 0.5\%$) that are necessary for the measurement of NMR relaxation dispersions. Yet 15N CPMG studies of a pair of these mutants, Gly48→Met48 (G48M) and Gly48→Val48 (G48V), performed in the absence of denaturant, are inconsistent with a two-state folding mechanism (9). Folding and unfolding rates obtained on a per-residue basis from fits of the dispersion profiles to a two-state folding model do not give the same values at each site, with differences as large as a factor of 10 in some cases; for cooperative folding, $k_{\text{U}}$ and $k_{\text{F}}$ values
are expected to be uniform among all sites in the protein. However, the data for all of the sites in both mutants can be well fit globally to a three-site model of conformational exchange, 

\[
\frac{k_{22}}{k_{21}} I = \frac{k_{32}}{k_{31}} U
\]

(Fig. 2A). The temperature dependence of the four rates allows values of free energy, enthalpy, and entropy differences to be extracted between pairs of states as well as activation barriers, assuming that the temperature dependence of the rates obeys transition state theory (9).

Further information is obtained in the form of chemical shift differences between the \( F \), \( I \), and \( U \) states. In fits of the dispersion data, no assumption is made a priori as to the positions of the \( I \) and \( U \) states along the reaction scheme; the chemical shifts obtained for an end point state are found to correspond to those of an unfolded protein (i.e., the \( U \) state) (Fig. 2B). It is well known that chemical shifts are sensitive indicators of molecular structure, and a long-standing goal is to exploit shifts of NMR “invisible” excited states so that structural information can be obtained. The deviations in chemical shifts of the \( I \) and \( U \) states from those of \( U \) obtained for the G48M Fyn SH3 domain (Fig. 2C), show that the central \( \beta \) sheet is largely formed in the \( I \) state (\( \beta2 \), \( \beta3 \), and \( \beta4 \)) but that there is considerably less structure at the termini of the domain. For reference, the \( F \) state structure of the domain is also illustrated (Fig. 2D).

**Measuring Dynamics on the Pico- to Nanosecond Time Scale**

Consider a protein that is tumbling in solution such that every orientation is equally probable. For the moment, let us focus on a single bond vector in the molecule that connects a pair of NMR active spins, such as a \( ^1\text{H}-^{15}\text{N} \) spin pair. When internal motions and molecular tumbling cause reorientation of the \( ^1\text{H}-^{15}\text{N} \) bond vector with respect to the external magnetic field, the local magnetic field at the site of the \( ^1\text{H} \) spin that derives from the directly attached \( ^1\text{H} \) magnetic dipole fluctuates (Fig. 3A). It can be shown that, although the local dipolar interaction between \( ^1\text{H} \) and \( ^{15}\text{N} \) spins averages to zero because of the molecular tumbling, the time-dependent variations in the field lead a spin system that has been perturbed by radio-frequency pulses to return, or relax, to thermal equilibrium. Because the fluctuations of the local magnetic fields are sensitive to internal motions, measurement of NMR relaxation rates provides a direct avenue to extracting dynamics parameters.

A pair of basic nitrogen spin relaxation experiments are used to probe backbone dynamics

\[
\begin{align*}
\text{equilibrium value (} T_2 &\text{). The } ^1\text{H} \text{ spin has a magnetogyric ratio (} \gamma \text{) that is 10 times larger than that of } ^{15}\text{N, and the inherent sensitivity of the NMR experiment scales as } \gamma^2; \text{ therefore experimental sensitivity is optimized by shuffling magnetization from an amide } ^1\text{H} \text{ to its directly coupled } ^{15}\text{N} \text{ and then back again to } ^1\text{H} \text{ for detection as a 2D data set. Here, one peak is obtained for each } (^1\text{H} , ^{15}\text{N}) \text{ pair in the protein, with an intensity proportional to } \exp(-1/2T_i^2) \text{ where } T_i \text{ is the } T_i \text{ value of the particular } ^{15}\text{N} \text{ nucleus and } t \text{ is a variable relaxation delay; relaxation times are measured by recording a series of spectra and fitting the peak intensities as a function of } t. \text{ The values of } T_i \text{ so obtained are usually interpreted in terms of generalized order parameters that describe the amplitude of bond vector motions and time constants that indicate the time scale of the internal motions (15), but specific models can be used as well (16). In addition, many experiments that provide complementary information to the } ^{15}\text{N} \text{ studies described above have been developed, such as those that measure correlated motions of successive residues in proteins (17, 18) or the dynamics of backbone carbonyl-C\( ^\alpha \) bonds (19).}
\end{align*}
\]

Side chain motions in proteins are also amenable to study with the use of spin relaxation techniques (20, 21). In particular, the use of the \( ^2\text{H} \) nucleus as a probe of side chain methyl group dynamics has seen substantial advances in recent years. It has long been known that the deuterium spin is an excellent probe of molecular motion, but applications in the solution state were limited, primarily because the poor resolution of \( ^2\text{H} \) spectra make site-specific studies extremely difficult. This is illustrated by protein spectra in Fig. 3C, where the broad, featureless \( ^2\text{H} \) spectrum may be compared with the high resolution \( ^1\text{H} \) spectrum. The deuteron can be exploited as a probe of molecular dynamics, however, by preparing uniformly \( ^1\text{C} \) labeled, fractionally deuterated protein and using the scheme illustrated in Fig. 3D that can be “tuned” to select, in this case, for \( ^1\text{CH}_2\text{D} \) methyl groups. A series of high resolution \( ^1\text{C} \)-\( ^2\text{H} \) maps are produced, with \( ^2\text{H} \) relaxation rates encoded by the intensities of the cross peaks.

One of the most exciting applications of the methodology described above is in the use of
Residual Dipolar Couplings

Probing Dynamics Spanning a Broad Time Scale Range by Using Residual Dipolar Couplings

Let us return to the example of the $^1$H-$^{15}$N spin pair in Fig. 3A, but now suppose that the protein is not dissolved in isotropic solution but rather in a media that leads to fractional alignment (typically about 0.1%) (23) (Fig. 4A). In this case, the dipolar interaction considered above does not average to zero. Instead, for a fixed orientation of the $^1$H-$^{15}$N spin vector, the effective magnetic field at the $^{15}$N spin is either increased or decreased depending on the $^1$H spin state, leading to dipolar splittings ($^{15}$N peak doublets) in spectra. These splittings are a rich source of structural information, because they report on the orientation of bond vectors with respect to an external coordinate frame (the magnetic field) (23). However, there is also potential for studies of biomolecular dynamics because motions that modulate local fields over a broad time regime (picosecond to millisecond) can affect dipolar splittings (24–27).

In this study, the authors used an extensive set of dipolar couplings recorded on the immunoglobulin-binding B1 domain of streptococcal protein G (protein G). Data from experiments performed in a number of different alignment media were pooled to obtain as many as 27 measurements per residue, with six different types of dipolar couplings measured per peptide plane (Fig. 4B). The dipolar coupling data were interpreted by using the x-ray structure of the protein along with a 3D Gaussian axial fluctuation model (GAF) of the motion (Fig. 4C) (16). A remarkable distribution of motion about the $\gamma$ axis was found for residues throughout the $\beta$ sheet of the protein (Fig. 4D), with hydrogen-bonded residues on adjacent strands experiencing similar levels of dynamics. In order to address whether the motions of interacting peptide planes are correlated, the authors recorded three-bond ($^{15}$N-$^{13}$C) scalar couplings ($^{15}$N-$^1$H-O-$^{13}$C), which depend on lengths and angles of H-O hydrogen bonds. Experimental couplings were compared with values calculated by using the GAF fluctuation amplitudes from the dipolar coupling measurements, assuming either correlated, anticorrelated, or uncorrelated motion. Significantly better agreement between computed and experimental couplings was obtained for the correlated motional model compared with the other two models. Although dipolar coupling data per se do not report on the time scale of motion, information can be obtained by comparing the per-residue generalized order parameters from the 3D GAF model of dynamics with order parameters extracted from backbone $^{15}$N spin relaxation data (see above) that report on picosecond to nanosecond time scale motions (Fig. 4E).

For the sites involved in the correlated dynamics described above, dipolar coupling–derived generalized order parameters are lower than those obtained from $^{15}$N relaxation experiments, implying that the correlated dynamics involve motional time scales that are slower than the picosecond to nanosecond range.

Lastly, we note that the residues with the highest level of dynamics are those that interact with the protein G binding partner and that the direction of the motion (about the $\gamma$ axis) coincides with the conformational adjustment needed for molecular recognition and for the formation of a hydrogen bonded complex.

Concluding Remarks

Over the past several years, new NMR experiments have been developed to provide site-specific information about protein motions spanning a range of time scales. Some of the most exciting new applications involve large molecular complexes, where motion is likely to be critical for function. NMR methods exploiting the TROSY (transverse relaxation optimized spectroscopy) principle (28) have emerged for both backbone positions and side chain methyl groups, allowing site-specific studies of dynamics to be performed on large protein complexes such as the GroEL-GroES chaperone (30) and the ClpP protease (31). It is clear from these studies, and from the applications to the smaller proteins described above, that the insights obtained from NMR dynamics studies will have important implications for our understanding of biological function.

References

TOOLS FOR BIOCHEMISTRY


PERSPECTIVE
Living Cells as Test Tubes
X. Sunney Xie,* Ji Yu, Wei Yuan Yang

The combination of specific probes and advanced optical microscopy now allows quantitative probing of biochemical reactions in living cells. On selected systems, one can detect and track a particular protein with single-molecule sensitivity, nanometer spatial precision, and millisecond time resolution. Metabolites, usually difficult to detect, can be imaged and monitored in living cells with coherent anti-Stokes Raman scattering microscopy. Here, we describe the application of these techniques in studying gene expression, active transport, and lipid metabolism.

Much of our quantitative understanding of molecular reactions in cells has come from traditional biochemistry—experiments done in test tubes with purified biomolecules. Although this approach is extremely productive, we understand that the milieu of the cell is fundamentally different from an in vitro solution in several ways: (i) DNA, many mRNA molecules, and some enzymes exist in low copy numbers and participate in stochastic reaction events in the cell that are hidden in test tubes with large numbers of molecules. (ii) Reactions are often at nonequilibrium steady state in the cell, with a constant supply of free energy and reactants. (iii) Many reactions are coupled in the cell, resulting in networks of complex interactions. Consequently, a biochemical reaction in a single cell could have different thermodynamic and kinetic properties from the same biochemical reaction in a test tube. The challenge now is to observe the biochemical reactions in living cells, and techniques are in place to do this in selected systems. Central to these techniques is optical imaging, which offers millisecond time resolution and nanometer spatial precision, single-molecule sensitivity, and most importantly, biochemical specificity. Here, we highlight advances that allow investigation of gene expression, active transport, and metabolism in living cells.

In an individual cell, gene expression is a single-molecule problem. On genomic DNA, a particular gene only exists in one (or a few) copy, switching on and off stochastically to regulate biological functions [for a review, see (1)]. Gene expression has been studied by biochemical assays, such as Northern and Western blotting, polymerase chain reaction, and more recently, mRNA arrays and mass spectrometry. However, these techniques are not sensitive enough to allow single-cell analysis of genes that are expressed at low levels. Furthermore, these ensemble-averaged methods often mask stochastic gene expression events. Single-molecule experiments in vitro have provided valuable insight into the mechanisms of gene expression machines (2–4). The next frontier will carry out single-molecule studies in individual living cells.

Imaging of gene expression at a single-molecule level in living cells has been made possible by two developments. At the transcriptional level, single mRNA molecules were detected and tracked in a living cell using multiple copies of a fluorescent mRNA binding protein (5, 6). At the translational level, we have tracked expression of single-protein molecules using a fast-maturing and membrane-targeting yellow fluorescent protein (YFP) (Venus) as a reporter (7).

Immobilizing the fluorescence protein reporter on the cell membrane (7, 8) overcame the difficulty in detecting single-protein molecules inside the cytoplasm; the fluorescence distributes throughout the cell because of fast protein diffusion during the image acquisition time and drops below the strong cellular autofluorescence. We monitored repressed expression from the lac promoter in Escherichia coli (Fig. 1A) and showed that protein expression occurs in small bursts (Fig. 1B), each originating from multiple ribosomes on an mRNA molecule. The protein copy numbers within a burst adhered to a geometric distribution (7), which was verified with the use of a different reporter (9). These assays provided quantitative details about the stochastic fluctuations in gene expression.

Is there a way to detect a single cytoplasmic protein molecule? The answer is yes, by extending the idea of detection by immobilization. We borrowed a method from strobe photography, which makes it possible to take a sharp picture of a bullet going through an apple (Fig. 1C). The sharpness is achieved because the light flash is so short that the bullet does not move far during the flash. Likewise, we applied an intense laser exposure for a very short duration (~300 μs), during which a protein reporter does not diffuse beyond the diffraction-limited spot. Figure 1D shows detection of single red fluorescent proteins [tdTomato (10)] in E. coli cytoplasm with a high signal-to-background ratio. The method could be used, for example, to determine the cellular concentration of a weakly expressed protein without calibration. To further develop this method, we need reporters with high photostability and better-controlled photochemistry.

The next step is to probe the expression of multiple genes simultaneously with the use of different colors of reporters (10) in order to study their interactions. In addition to transcription and translation, similar live-cell single-molecule assays offer the prospect of studying cellular processes, such as cell signaling (11), protein folding, DNA replication, and RNA trafficking (5, 12).

No less important than gene expression is energy transduction in living cells. Motor proteins convert chemical energy in the form of adenosine 5’-triphosphate (ATP) into mechanical work. Kinesin and dynein motors transport organelles along microtubules in opposite directions. Much has been learned about these motors at the single-
center of its circular swimming path. These particles move randomly with an apparent diffusion coefficient of \( D = 9.0 \pm 2.0 \, \mu m^2/s \), measured for isolated spermatozoa. A short-range pairwise attraction, arising from the hydrodynamic forces leading to the observed synchronization (20), and a longer range repulsion, which could be of steric or hydrodynamic origin (21), are assumed (Fig. 4D). Although one cannot describe circular flow by a potential (22), the important features of the observed pattern are captured by our model.

Stochastic simulations of this model (SOM text) also revealed two regimes: a random distribution of particles at low densities with a transition toward a hexagonal array of clusters at a critical particle density (Fig. 4E). Assigning to each particle a spermatozoon circling around that position, we generated simulated movies (3) mimicking the experimental observation (Fig. 4F versus Fig. 1B). Moreover, the order parameter \( \gamma \) computed for different simulated sperm densities agreed with the experimentally observed dependency (Fig. 4C). Our numerical results were further supported by a 1D mean-field analysis (SOM text), which indicated the existence of a supercritical pitchfork bifurcation at a critical sperm density (23). This critical density was proportional to the inverse sixth power of the distance \( D \) of two spins: \( I_s = \mu d_s^{-6} \). This isolated spin pair approximation (ISPA) involves an unknown scaling factor \( \gamma \). It seems straightforward to obtain the structure in the example: simply use the observed intensities to calculate sufficient distances to define the structure.

In realistic applications, this approach runs into difficulties. One problem is that the forward model is usually inherently degenerate, meaning that different conformations can lead to the same observations and therefore cannot be distinguished experimentally, and even a formally invertible forward model is practically degenerate if the data are incomplete. A further complication is that there are uncertainties in both the data and the forward model: Data are subject to experimental errors, and theories rest on approximations. Moreover, the forward model typically involves parameters that are not measurable. Algorithms for structure calculation from x-ray reflections, NMR spectra, or homology-derived restraints should account for these fundamental difficulties in some way.

Structure determination in general is an ill-posed inverse problem, meaning that going from the data to a unique structure is impossible. However, the current paradigm in structure calculation is to attempt an inversion of the forward model. Most algorithms minimize a hybrid energy \( E_{\text{hybrid}} = E_{\text{phys}} + E_{\text{ISPA}} \)

---

**Inferential Structure Determination**

*Wolfgang Rieping,* Michael Habeck, *Michael Nilges*

Macromolecular structures calculated from nuclear magnetic resonance data are not fully determined by experimental data but depend on subjective choices in data treatment and parameter settings. This makes it difficult to objectively judge the precision of the structures. We used Bayesian inference to derive a probability distribution that represents the unknown structure and its precision. This probability distribution also determines additional unknowns, such as theory parameters, that previously had to be chosen empirically. We implemented this approach by using Markov chain Monte Carlo techniques. Our method provides an objective figure of merit and improves structural quality.

A major difficulty in the determination of three-dimensional macromolecular structures is that experimental data are indirect. We observe physical effects that depend on the atomic geometry and use a forward model to relate the observed data to the atomic coordinates. For example in nuclear magnetic resonance (NMR), the intensity \( I_s \) of peaks in nuclear Overhauser effect spectroscopy (NOESY) data is proportional to the inverse sixth power of the distance \( d \) of two spins: \( I_s = \mu d_s^{-6} \). This isolated spin pair approximation (ISPA) involves an unknown scaling factor \( \gamma \). It seems
w_{\text{data}}E_{\text{data}}(2), where a nonphysical energy $E_{\text{data}}$ uses the forward model and a restraining function to assess the agreement between data and structure. A force field $E_{\text{phys}}$ describes the physical properties of the macromolecule, such as bonded and nonbonded interactions between the atoms, and partially removes the degeneracy of the problem. The rationale is that minimization of the hybrid energy effectively inverts the forward model, yielding the “true” structure.

This strategy works in the case of many data of good quality. In less favorable situations, the ill-posed nature of the inverse problem becomes apparent. Specifically, it remains unclear how to choose auxiliary parameters like the weight $w_{\text{data}}$ or theory parameters such as the scaling factor $\gamma$ in the ISPA. Because the hybrid energy minimization paradigm offers no principle to settle these issues, such parameters need to be determined heuristically.

The principal difficulty in structure determination by NMR is the lack of information that is indispensible to reconstruct the structure unambiguously. By formulating an optimization problem ("search for the minimum of $f_{\text{hybrid}}"), one however implicitly assumes that there is a unique answer. Repeating the optimization procedure multiple times to obtain several “unique” solutions hides but does not solve the ambiguity and makes it difficult to judge the validity and precision of NMR structures in an objective way.

We suggest that it is a misconception to use structure calculation methods that are only appropriate if the objective is to obtain a unique structure. Instead, we view structure determination as an inference problem, requiring reasoning from incomplete and uncertain information. We consider the entire conformational space and use the data only to rank the molecule’s possible conformations. We assign a number $P_i$ to every conformation $X_i$. If $P_i > P_j$, conformation $X_i$ is more supported by the data than $X_j$. Cox (3) proved that such rankings are equivalent to a probability and that probability theory is the only consistent method for assigning probabilities to events that are unavailable from the data but necessary in order to describe the problem adequately. In Bayesian theory, such nuisance parameters are treated in the same way as the coordinates: They are estimated from the experimental data by replacing $X$ with $(X, \xi)$ in Eq. 1. Assuming independence of $X$ and $\xi$, the joint posterior density for all unknown parameters is

$$p(X, \xi|D, I) \propto p(D|X, \xi, I) p(X|I) p(\xi|I)$$

Equation 2 provides a unique rule to determine any quantity that is not accessible by experiment.

To demonstrate the practical feasibility of the ISD approach, we infer the molecular structure of the Fyn SH3 domain (59 amino acids length). Experimental distances between amide protons were derived from a series of NOESY spectra on a [15N, 2H] enriched protein (6). The data set is sparse: It comprises 154 measurements, of which on average only one per amino acid provides long-range structural information. The forward model $I = I_{\text{ISPA}}$ defined by the ISPA does not account for experimental errors and systematic effects like spin diffusion (7) and internal dynamics (7); hence, observed intensities will deviate from theoretical predictions. A log normal distribution (5) describes these deviations and introduces a second nuisance parameter $\sigma$ that quantifies their magnitude. Thus, we have two nuisance parameters, $\xi = (\gamma, \sigma)$.

Although given in analytically closed form (5), it is practically impossible to evaluate the posterior density $p(X, \gamma, \sigma|D, I)$ over all conformational space. Therefore, in our view, structure calculation comprises posterior simulation, which samples only regions that carry a considerable amount of probability mass. We have
developed a Markov chain Monte Carlo (MCMC) algorithm based on the replica-exchange method (8) to simulate the joint posterior density of a structure determination problem (5, 9) (Fig. 1 and fig. S1).

The most pronounced features of the posterior density can be represented in a set of conformational samples. Although this looks at first glance like a conventional structure ensemble, the rationale behind our approach to obtain conformational samples is very different. The uncertainty of atomic positions is directly influenced by the uncertainty of nuisance parameters and by the quality of the data. Effects not described in the ISPA, such as protein dynamics, tend to increase the deviations between predicted and measured peak intensities. This is reflected in an increase of the error σ and consequently leads to a loss in structural precision. However, unless the forward model incorporates experimental information on protein dynamics, we cannot discriminate motion from imprecisions due to experimental errors or lack of data.

Compared with conventional structure ensembles, our conformational samples are much better defined and systematically closer to the structure obtained with x-ray crystallography (10) (Fig. 2). A comparison of the 20 most probable conformations with the x-ray structure yields a backbone heavy atom rmsd (root mean square deviation) of 1.84 ± 0.20 Å for all residues and 1.36 ± 0.19 Å for the secondary structural elements. This is a considerable improvement over conventional techniques used in (6), where an ensemble with an overall rmsd of 2.86 ± 0.33 Å and an rmsd of 2.01 ± 0.28 Å for secondary structure elements was obtained. This improvement originates in the calculation of structures by random sampling, which searches conformational space more exhaustively and suppresses topologically unlikely conformations. Misfolds such as mirror images can only be realized in a small number of ways; thus, they are entropically suppressed and do not show up in a statistical ensemble. Discriminating such conformations on the basis of the hybrid energy is more difficult, in particular if the data are sparse.

A probabilistic structure ensemble is exclusively determined by the data and the working hypotheses that enter the analysis (which are in the presented example the ISPA, the log-normal error distribution, and our choice of the force field). Modifications will, of course, lead to changes in the structures. The atom positions, for example, are sensitive to the parameters and the functional form of the force field used in the conformational prior density. This also holds for conventional approaches, which are based on analogous assumptions. However, in addition, conventional methods require empirical rules to treat nuisance parameters, because they cannot be determined from the hybrid energy alone. Cross-validation (11, 12) and maximum likelihood methods (13), for example, have successfully been applied in NMR and crystallographic refinement to determine certain nuisance parameters such as the weight w_data. The ISD approach goes...
beyond these techniques. Once the working hypotheses are made, Eq. 2 provides definite rules to determine any nuisance parameter, including its uncertainty, directly from the data (Fig. 3). Therefore heuristics and other subjective elements are superfluous.

Because conventional structure ensembles depend on user-specific parameter settings and on the minimization protocol, it is difficult if not impossible to assign statistically meaningful error bars to atomic coordinates. In contrast, stochastic samples drawn from the joint posterior density \( p(X; \gamma, \sigma(D,J)) \) are statistically well defined and can directly be used to calculate estimates of mean values and standard deviations (14). As a special case, we can derive an average structure with atom-wise error bars and are thus able to define an objective figure of merit for NMR structures (Fig. 4).

Bayesian and maximum likelihood approaches have already proven useful for data analysis and partial aspects of structure refinement in NMR spectroscopy and x-ray crystallography (15, 16, 13, 17). Our results suggest that structure determination can be solved entirely in a probabilistic framework.

It is straightforward to apply our approach to other NMR parameters. In case of three- bond scalar coupling constants, for example, an appropriate forward model is the Karplus curve (18) involving three coefficients that are treated as nuisance parameters. However, our method is not restricted to NMR data and can be applied to other structure determination problems. Besides theoretical coherence, a rigorous probabilistic approach has decisive practical advantages. It has no free parameter and is stable for many more than the two nuisance parameters used in the example (19). Hence, tedious and time-consuming searches for optimal values are no longer necessary. Once the forward model to describe the data has been chosen, probability calculus uniquely determines the posterior distribution for all unknowns. It is then only a computational issue to generate posterior samples. Further intervention is not required, and structure determination attains objectivity.

**Fig. 3.** Estimation of nuisance parameters. Posterior histograms compiled from MCMC samples for the scaling factor \( \gamma \) in the ISPA and for the width \( \sigma \) of the log normal error distribution. (A) Posterior histogram \( p(\gamma^{-1/6};D,J) \) for the inverse sixth power of \( \gamma \). This factor corrects interproton distances to match the experimental distances best. (B) Posterior histogram \( p(\sigma(D,J)) \) for the error \( \sigma \). In conventional approaches, this analog to the weight \( w_{\text{data}} \approx \sigma^{-2} \) can only be estimated via cross-validation or must be set empirically.

**Fig. 4.** Conformational uncertainty. MOLMOL "sausage" plot of the mean structure with atom-wise error bars indicated by the thickness of the sausage. The 20 most probable conformations (also shown in Fig. 2A) from the simulation of the joint posterior distribution \( p(X;\gamma;\sigma(D,J)) \) were used to calculate the average structure and its precision. The local precision ranges from 0.6 Å for secondary structure elements to 4.6 Å for loop regions (bottom and right-hand side) and termini (top). The average precision is 1.07 Å. The average precision of the structure ensembles calculated with CNS is 4.93 Å for the flat-bottom harmonic-wall potential and 5.04 Å for the harmonic potential.

**References and Notes**

5. Materials and methods are available as supporting material on Science Online.
26. The authors thank I. D. Campbell for kindly providing the experimental SH3 NMR data. This work was supported by European Union grants QLG2-CT-2000-01313 and QLG2-CT-2002-00988. The 20 most likely structures and the restraint list used in the calculation have been deposited in the Protein Data Bank under accession code 1ZBJ. The structure determination program is available from the authors on request to M.N.

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/309/5732/303/DC1

Materials and Methods

Fig. 51

31 January 2005; accepted 16 May 2005

10.1126/science.1110428
The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis

David D. Bohr, Dan McElheny,* H. Jane Dyson, Peter E. Wright†

We used nuclear magnetic resonance relaxation dispersion to characterize higher energy conformational substates of Escherichia coli dihydrofolate reductase. Each intermediate in the catalytic cycle samples low-lying excited states whose conformations resemble the ground-state structures of preceding and following intermediates. Substrate and cofactor exchange occurs through these excited states. The maximum hydride transfer and steady-state turnover rates are governed by the dynamics of transitions between ground and excited states of the intermediates. Thus, the modulation of the enzyme landscape by the bound ligands funnels the enzyme through its reaction cycle along a preferred kinetic path.

It has long been recognized that dynamic fluctuations in protein conformation play a central role in enzyme catalysis (1–3). Protein dynamics are implicated in events such as substrate or cofactor binding and product release, and the chemical event itself involves an inherently dynamic process, with changes in atomic coordinates required along the reaction coordinate (4). Although there is considerable evidence from both theory and experiment that many enzymes are inherently flexible, the fundamental mechanisms by which protein fluctuations couple to catalytic function remain poorly understood.

Escherichia coli dihydrofolate reductase (DHFR) has been used extensively as a model enzyme for investigating the relations between structure, dynamics, and function. Theoretical and experimental investigations suggest that protein fluctuations play a direct role in catalysis by DHFR [see (5) for a recent review]. The enzyme catalyzes the reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) by using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor. The kinetic mechanism involves rebinding of NADPH to assist the release of the THF product. The enzyme (E) cycles through five major intermediates: E:NADPH, E:NADPH:DHF, E:NADP⁺:folate, E:THF, and E:NADP⁺:THF (6) (Fig. 1A). The structures of all of the kinetic intermediates, or models of the intermediates, have been determined by x-ray crystallography, and the conformational changes that occur during the catalytic cycle have been delineated (7, 8). The major sites of conformational change include the active-site loop (residues 9 to 24, termed the Met20 loop) and the substrate-binding pocket (7) (Fig. 1B). In the holoenzyme E:NADPH and the Michaelis complex E:NADPH:DHF (modeled by the ternary E:NADP⁺:folate complex), the Met20 loop adopts a closed conformation, where it packs against the nicotinamide ring of the cofactor bound within the active site. In the three product complexes, the Met20 loop adopts an occluded conformation, where it sterically hinders the nicotinamide ring from binding in the active site; therefore, the nicotinamide ring is outside the pocket in these complexes. The substrate-binding pocket undergoes a similar transition, closing more tightly when both substrate and cofactor are present and opening to release products (7).

Protein dynamics can be evaluated experimentally by nuclear magnetic resonance (NMR) spin relaxation techniques (9). Carr-Purcell-Meiboom-Gill (CPMG)-based R2 relaxation experiments monitor motion on the μs to ms time scale that is generally the most relevant for protein conformational change (10). Through these methods, the transverse relaxation rate, R2, can be decomposed into two contributions (9). For two-site chemical exchange between a ground state (A) and an excited state (B), R2 relaxation dispersion is a function of the exchange rate constant kex (kex = kA → B + kB → A), the populations of states A and B (pA and pB, respectively), and the chemical shift difference between states A and B (Δω) (11), thus giving information regarding the kinetics and thermodynamics of protein motion (12–15) and providing insight into the structure of the higher energy state (13, 15, 16).

The 15N R2 relaxation dispersion measurements for the Michaelis complex model E:NADP⁺:folate (15) indicated that many of the residues that exhibit exchange contributions to relaxation are directly or indirectly associated with the Met20 loop (Fig. 2A). These residues show characteristic chemical shift differences between closed and occluded complexes, and their resonances have been previously categorized as active site loop conformation markers (17). Likewise, chemical shift perturbation studies identified cofactor-binding and substrate-
A product-binding (substrate-/product-binding) marker resonances associated with residues that cluster around the cofactor- and substrate-binding pockets, respectively (17) (Fig. 1B). A comparison of the dynamic chemical shift differences (Δδ values) determined from fits of the $R_2$ relaxation dispersion data to the equilibrium chemical shift differences (ΔS values) between the closed complex E:NADP±:folate and the occluded complex E:DHNA:DHNA:folate showed a remarkable linear correlation (15). Thus, the higher energy state contributing to $R_2$ relaxation in the closed E:NADP±:folate complex represents an occluded conformation similar to that found in the E:NADPH±:THF product ternary complex.

A complete set of $^{15}$N and $^1$H $R_2$ relaxation dispersion data have now been obtained for DHFR complexes that represent all of the kinetic intermediates populated in the steady-state catalytic cycle. Dispersion data measured at two frequencies were fitted to the general two-site exchange equations; the methods, dispersion curves, and fitted parameters for all complexes are provided (tables S1 to S4 and Figs. S1 to S5). Some of these data are shown in Figs. 2 and 3.

Analysis of amide $^{15}$N and $^1$H $R_2$ relaxation dispersion measurements for the holoenzyme E:NADPH revealed exchange processes for many residues located in or around the substrate binding site (Fig. 2B). Dispersive behavior was also observed for several residues in the active site loop and the loop (residues 116 to 132) between β strands F and G (the FG loop), but no relaxation dispersion. Some of these data are shown in Figs. 2 and 3A). A complete set of $^{13}$C and $^1$H $R_2$ relaxation dispersion data have now been obtained for DHFR complexes that represent all of the kinetic intermediates populated in the steady-state catalytic cycle. Dispersion data measured at two frequencies were fitted to the general two-site exchange equations; the methods, dispersion curves, and fitted parameters for all complexes are provided (tables S1 to S4 and Figs. S1 to S5). Some of these data are shown in Figs. 2 and 3.

Analysis of amide $^{15}$N and $^1$H $R_2$ relaxation dispersion measurements for the holoenzyme E:NADPH revealed exchange processes for many residues located in or around the substrate binding site (Fig. 2B). Dispersive behavior was also observed for several residues in the active site loop and the loop (residues 116 to 132) between β strands F and G (the FG loop), but no relaxation dispersion was seen for residues in the cofactor-binding site. The localization of the residues showing exchange contributions to relaxation around the substrate-binding pocket suggests that the higher energy conformation sampled by E:NADPH plays an important role in capturing the substrate. Indeed, there is a strong linear correlation between the Δδ values derived from the relaxation dispersion curves and the ΔS values derived from the chemical shift differences between E:NADPH and E:NADPH:THF, or between E:NADPH and E:NADPH±:folate, representing the previous step or the next step in the cycle, respectively (Fig. 3A).

A complete set of $^{15}$N and $^1$H $R_2$ relaxation dispersion data have now been obtained for DHFR complexes that represent all of the kinetic intermediates populated in the steady-state catalytic cycle. Dispersion data measured at two frequencies were fitted to the general two-site exchange equations; the methods, dispersion curves, and fitted parameters for all complexes are provided (tables S1 to S4 and Figs. S1 to S5). Some of these data are shown in Figs. 2 and 3.

Analysis of amide $^{15}$N and $^1$H $R_2$ relaxation dispersion measurements for the holoenzyme E:NADPH revealed exchange processes for many residues located in or around the substrate binding site (Fig. 2B). Dispersive behavior was also observed for several residues in the active site loop and the loop (residues 116 to 132) between β strands F and G (the FG loop), but no relaxation dispersion was seen for residues in the cofactor-binding site. The localization of the residues showing exchange contributions to relaxation around the substrate-binding pocket suggests that the higher energy conformation sampled by E:NADPH plays an important role in capturing the substrate. Indeed, there is a strong linear correlation between the Δδ values derived from the relaxation dispersion curves and the ΔS values derived from the chemical shift differences between E:NADPH and E:NADPH:THF, or between E:NADPH and E:NADPH±:folate, representing the previous step or the next step in the cycle, respectively (Fig. 3A). This result implies that the E:NADPH complex samples a higher energy substrate state in which the empty substrate-/product-binding pocket adopts a conformation similar to that of the ligand-bound state. A similar observation has been reported for ribonuclease A (RNase A). As a result of conformational fluctuations, the free enzyme samples a higher energy state whose structure resembles the ligand-bound form (18). Although many residues in the Met10 and FG loops experience exchange contributions, the derived Δδ values do not correlate with the ΔS values between the closed and the occluded conformations (Fig. S6); the active site loop conformation in the excited state is currently unknown.

---

Fig. 2. Relaxation dispersion data for each intermediate in the catalytic cycle of DHFR. (Left) Model structures for various intermediates of the DHFR catalytic cycle (7). The backbone is colored red, green, or blue as in Fig. 1B. Residues for which conformational exchange is observed are indicated with spheres, which are colored red, green, and blue for resonances that report on active site loop conformation, cofactor binding, and substrate/product binding, respectively, and gray for resonances that are not identified with any of these categories. Yellow indicates resonances that show broadening, but for which data quality was insufficient to obtain reliable $R_2$ relaxation dispersion results. (Right) Representative $^{15}$N $R_2$ relaxation dispersion curves for each complex. A full set of data for all residues that show detectable relaxation dispersion is included (figs. S1 to S5). Error bars indicate estimated uncertainties in $R_2$ ($15$).

A) E:NADP±:folate (PDB 1RX2) and NMR data at 303 K. (B) E:NADPH (PDB 1RX1) and NMR data at 303 K. (C) E:THF (PDB 1RX5) and NMR data at 300 K. (D) E:NADP±:THF (PDB 1RX4) and NMR data at 300 K. (E) E:NADPH±:THF (PDB 1RX6) and NMR data at 300 K. Relaxation dispersion data were collected and analyzed at two external magnetic field strengths (1H: 500 MHz and 800 MHz), but only 800 MHz data are shown for clarity. Red curves report on the active site loop conformation marker Gly23; blue curves, on the substrate-/product-binding marker Asp97; and green curves, on the cofactor-binding marker Ser77 (A) to (C) and (E) or Met62 (D). The green curve for Ser77 in (E) (green) has been offset for clarity (right axis). This figure was generated in part by using MOLMOL (27).

---

Fig. 3. Correlation between Δδ values obtained from relaxation dispersion measurements and differences ($\Delta \delta$) between chemical shifts in the ground states of adjacent complexes in the catalytic cycle. (A) Δδ for E:NADPH plotted against Δδ [equal to $\delta_{E:NADPH} - \delta_{E:NADPH±:THF}$] (circles); or $\delta_{E:NADPH} - \delta_{E:NADPH±:folate}$ triangles (slope = 0.95, $R^2$ = 0.97); (B) Δδ for E:THF plotted against Δδ [equal to $\delta_{E:THF} - \delta_{E:NADPH±:THF}$] (circles); or $\delta_{E:THF} - \delta_{E:NADPH±:folate}$ triangles (slope = 1.1, $R^2$ = 0.99); (C) Δδ for E:NADPH±:THF plotted against Δδ [equal to $\delta_{E:NADPH±:THF} - \delta_{E:NADPH±:THF±:folate}$] by using 10 mM (circles) or 50 mM (triangles) NADP+ (slope = 1.0, $R^2$ = 0.99); and (D) Δδ for E:NADPH±:THF plotted against Δδ [equal to $\delta_{E:NADPH±:THF} - \delta_{E:NADPH±:THF±:folate}$] by using 10 mM (circles) or 50 mM (triangles) NADP+ (slope = 0.91, $R^2$ = 0.94). The data points for the Gly37 amide, which hydrogen-bonds directly to the cofactor, are enclosed in a circle and were not included when determining the line of best fit. (E) Δδ for E:NADPH±:THF plotted against Δδ [equal to $\delta_{E:NADPH±:THF} - \delta_{E:NADPH±:THF±:folate}$] (slope = 0.97, $R^2$ = 0.98). Residues are colored red, green, and blue to indicate residues reporting on the active site loop conformation, cofactor binding, and substrate/product binding, respectively. Solid symbols indicate that the sign of Δδ could be determined from a comparison of HSQC and heteronuclear multiple-quantum coherence spectra at an external magnetic field strength of 1H 500 MHz (28), and open symbols indicate residues where only the absolute values for Δδ and Δδ are reported. Error bars indicate uncertainties in Δδ estimated by Monte Carlo simulation (25).
In direct contrast to the E:NADPH complex, residues surrounding the cofactor-binding cleft display exchange contributions to relaxation in the E:THF complex (Fig. 2C). A linear correlation is observed between $\Delta \alpha$ and $\Delta \delta_{(\text{E:THF} - \text{ENADP}^+ : \text{THF})}$ or $\Delta \delta_{(\text{E:THF} - \text{ENADPH} : \text{THF})}$ values (Fig. 3B), suggesting that the higher energy conformation contributing to $^{15}$N $R_2$ relaxation in E:THF resembles the product ternary complexes. Any relaxation dispersion observed for residues lining the substrate-binding pocket or in the FG loop can generally be traced to local differences in conformation between E:THF and product ternary complexes (table S2 and fig. S7).

In the E:NADP$^+ : $THF complex, conformational changes are observed in the active site loops and the ligand-binding pockets (Fig. 2D). The $\Delta \alpha$ values for residues surrounding the cofactor-binding cleft and the active site loops correlate with different $\Delta \delta$ values (Fig. 3, C and D). The $\Delta \alpha$ values for residues in the active site loops correlate to $\Delta \delta_{(\text{E:THF} - \text{ENADPH} : \text{folate})}$ (Fig. 3C), showing that the occluded E:NADP$^+ : $THF complex samples a higher energy state in which the active site loops are in a closed conformation, resembling the conformation of the E:DHF:NADPH Michaelis complex (modeled by E:NADP$^+ : $folate) that immediately precedes it in the catalytic cycle. For many of the residues surrounding the cofactor-binding cleft, a linear correlation is observed between $\Delta \alpha$ and $\Delta \delta_{(\text{E:THF} - \text{ENADPH} : \text{THF} - \text{E:folate})}$ and/or $\Delta \delta_{(\text{E:THF} - \text{ENADP}^+ : \text{THF} - \text{E:folate})}$ (Fig. 3D), revealing the presence of an additional excited state in which the conformation of the adenosine-binding site is similar to that in the binary E:THF product complex. The excited protein substates do not reflect physical dissociation of cofactor or chemical changes. The population of E:THF in equilibrium with the ternary product complex E:NADP$^+ : $THF is estimated to be 0.4% on the basis of rate constants determined from pre–steady-state analysis (6), whereas the excited state population from relaxation dispersion experiments is much larger ($p_B > 2.3\%$). Repeat experiments at fivefold higher NADP$^+$ concentration, where the population of the E:THF complex is estimated to be $\sim 0.08\%$, showed identical $R_2$ relaxation dispersion for the residues surrounding the adenosine-binding site (table S3), which rules out cofactor dissociation as the origin of the exchange contributions to the $R_2$ relaxation rates. In addition, the x-ray structures (7) show that the Gly$^{96}$ amide forms a hydrogen bond to the phosphate group of the cofactor, which leads to a large change in the $^{15}$N chemical shift [3.5 to 4.0 parts per million (ppm)] upon binding of NADP$^+$ to the E:THF or E:folate complexes. However, the $\Delta \alpha$ for Gly$^{96}$ is much smaller (<1.35 ppm) (Fig. 3D), implying that the hydrogen bond remains largely intact and that conformational exchange is not modulated by cofactor dissociation. The closed excited-state conformation of the active site loops also cannot be a consequence of hydride transfer, because the rate constant of the back reaction is too slow at this pH ($\sim 0.03$ s$^{-1}$) for this process to contribute measurably to $R_2$ relaxation dispersion (6).

The ground-state conformations of E:NADPH:THF and E:NADP$^+ : $THF are very similar, as evidenced by their nearly identical $^{15}$N heteronuclear single-quantum coherence (HSQC) spectra, yet the two complexes exhibit very different $R_2$ relaxation dispersion. The E:NADPH:THF complex (Fig. 2E) exhibits more pronounced dispersive behavior for residues surrounding the substrate-/product-binding

---

**Fig. 4.** The dynamic energy landscape of DHFR catalysis. Ground state (larger) and higher energy (smaller) structures of each intermediate in the cycle, modeled on the published x-ray structures (7), are shown color-coded according to the scheme in Fig. 1A, with NADPH and NADP$^+$ shown in gold and substrate, product, and analsogs shown in magenta. For each intermediate in the catalytic cycle, the higher energy conformations detected in the relaxation dispersion experiments resemble the ground-state conformations of adjacent intermediates; their interconversion rates, also obtained from the relaxation dispersion experiments, are shown with black arrows. Rate constants for the interconversion between the complexes, measured by pre–steady-state enzyme kinetics at 298 K, $p_B = 6$ (6) are indicated with red arrows. $R_2$ relaxation dispersion measurements were made at pH = 6.8 (E:NADP$^+ : $folate) or pH = 7.6 (E:NADPH:THF, E:NADP$^+ : $THF, E:NADPH, and E:THF) at 281 K (E:NADPH), 300 K (E:NADPH:THF, E:NADP$^+ : $THF, and E:THF), or 303 K (E:NADP$^+ : $folate).
transition, which the cofactor- or product-binding pocket is inaccessible to the technique. However, the excited-state conformations that we observe, together with the ground-state conformation, will constitute the lowest-energy members of the conformational ensemble of each intermediate. These results imply that the most functionally relevant conformations also possess the lowest energy of all potential conformations. In this view, ligands dictate not only the ground-state conformation but also the most accessible higher energy substates. As ligands change, through binding or dissociation processes or through chemistry, the energy landscape and the populations of the accessible states change in response. Thus, the dynamic energy landscape (26) efficiently funnels the enzyme through its catalytically competent conformations along a preferred kinetic path, where the number and heights of the energetic barriers between consecutive conformations have been minimized.

References and Notes
20. J. C. Tsai, S. Kumar, B. Y. Ma, R. Nussinov, Protein Sci. 8, 3183 (1999).
38. J. C. Tsai, S. Kumar, B. Y. Ma, R. Nussinov, Protein Sci. 8, 3183 (1999).
We introduce a method for optically imaging intracellular proteins at nanometer spatial resolution. Numerous sparse subsets of photoactivatable fluorescent protein molecules were activated, localized (to ~2 to 25 nanometers), and then bleached. The aggregate position information from all subsets was then assembled into a superresolution image. We used this method—termed localized (to \( \ell \)) stimulated emission depletion (L-STED)—to image specific target proteins in thin sections of lysosomes and mitochondria; in fixed whole cells, we imaged vinculin at focal adhesions, actin within a lamellipodium, and the distribution of the retroviral protein Gag at the plasma membrane.

Transfected cells expressing fluorescent proteins (1) contain information that is accurate at the molecular level about the spatial organization of the target proteins to which they are bound. However, the best resolution that can be obtained by diffraction-limited conventional optical techniques is coarser than the molecular level by two orders of magnitude. Great progress has been made with superresolution methods that penetrate beyond this limit, such as near field (2), stimulated emission depletion (3), structured illumination (4, 5), and reversible saturable optical fluorescence transients microscopy (6), but the goal remains a fluorescence technique capable of achieving resolution closer to the molecular scale.

Early results (7) in single-molecule microscopy (8) and the spatiotemporal isolation of individual exciton recombination sites in a semiconductor quantum well (9) led to a proposal for a means of molecular resolution fluorescence microscopy a decade ago (10). In brief, individual molecules densely packed within the resolution limit of a given instrument [as defined by its point-spread function (PSF)] are first isolated from one another on the basis of one or more distinguishing optical characteristics. Each molecule is then localized to much higher precision by determining its center of fluorescence emission through a statistical fit of the ideal PSF to its measured photon distribution (Fig. 1). When the background noise is negligible compared with the molecular signal, the error in the fitted position is \( \sigma \) = \( \sqrt{(N/\gamma)} \), where \( \gamma \) is the standard deviation of a Gaussian approximating the true PSF (\( \gamma = 200 \) nm for light of wavelength \( \lambda = 500 \) nm) and \( N \) is the total number of detected photons (11, 12). Given that it is possible to detect many more than \( 10^3 \) photons from a single fluorophore before it bleaches, single-molecule localization to nearly 1-nm precision has already been demonstrated (13–15) and applied to studies of molecular motor dynamics (13).

Multiple emitters within a single diffraction-limited region (DLR) have been isolated from one another by either spectral (16, 17) or temporal means, the latter exploiting the photobleaching (14, 17) or blinking (18) of the emitters. However, the number of emitters isolated per DLR (typically 2 to 5) has been too small to give resolution within the DLR that is comparable to existing superresolution techniques, and it is far from the molecular level. Here, we developed a method for isolation of single molecules at high densities (up to \( \sim 10^5 \mu m^{-2} \)) based on the serial photoactivation and subsequent bleaching of numerous sparse subsets of photoactivatable fluorescent protein (PA-FP) molecules (19–24) within a sample. We then applied the method to image specific target proteins in thin (~50–80-nm) sections and near the surfaces of fixed cultured cells, resolving the most precisely localized molecules therein at separations (~10 nm) approaching the molecular level.

The method and typical data subsets are shown in Fig. 1. Cultured mammalian cells expressing PA-FP-tagged target proteins were prepared by transient transfection, fixed, and processed on cover slips either as whole cells or in cryosections cut from a centrifuged pellet of cells (25). Such cover slips were then placed in a custom microscope chamber (fig. S1) designed to minimize thermal and mechanical drift (fig. S2) (25). They were continuously excited by a laser at a wavelength (\( \lambda_{exc} = 561 \) nm) near the excitation maximum of the activated form of the expressed PA-FPs. Finally, to minimize both autofluorescence and detector noise, they were imaged by total internal reflection fluorescence (TIRF) microscopy (13, 26) onto an electron-multiplying charge coupled device (EMCCD) camera that can detect single photons.

Initial image frames typically consisted of sparse fields of individually resolvable single molecules on a weaker background presumably dominated by the much larger population of PA-FP molecules still in the inactivated state. When necessary, excitation and thus bleaching was maintained until such sparse fields were obtained. Additional image frames were then captured until single-molecule bleaching resulted in a mean molecular separation considerably larger than that required for isolation (Fig. 1, A and C). At that point, we applied a light pulse from a second laser at a wavelength (\( \lambda_{act} = 405 \) nm) capable of activating the remaining inactive PA-FPs, at a duration and intensity chosen so that the overall density of activated PA-FPs was increased back to a higher, but still resolvable, level (Fig. 1, B and D). This process of photoactivation, measurement, and bleaching was then repeated (movie S1) for many cycles over \( \sim 10^4 \) to \( \sim 10^5 \) image frames (depending on the expression level and spatial distribution of the PA-FPs) until the population of inactivated, unbleached molecules was depleted. At typical frame rates of ~0.5 to 1.0 s, between 2 and 12 hours were required to acquire a complete image stack that could be distilled to a single superresolution image containing \( \sim 10^2 \) to \( \sim 10^3 \) localized molecules. We continued to explore methods (such as brighter molecules, higher excitation power, and higher activation density) to speed this process.

When the \( xy \) frames from any such image stack are summed across time \( t \), the molecular signals overlap to produce a diffusion-limited image (Fig. 1, E and F) similar to that obtained by conventional TIRF, in which all molecules emit simultaneously (fig. S3). However, when the data are plotted in a multidimensional volume \( xyt \) (Fig. 1, center), the signal from each molecule \( m \) is uniquely isolated and can be summed at each pixel and across all of it.