J|A|C|S
ARTICLES
Published on Web 11/08/2005

# NMR and X-ray Crystallography, Complementary Tools in Structural Proteomics of Small Proteins

Adelinda A. Yee,[†,||] Alexei Savchenko,[†,⊥] Alexandr Ignachenko,[†] Jonathan Lukin,[†] Xiaohui Xu,[†,⊥] Tatiana Skarina,[†,⊥] Elena Evdokimova,[†,⊥] Cheng Song Liu,[†,||] Anthony Semesi,[†,||] Valerie Guido,[†,||] Aled M. Edwards,[†,§,⊥] and Cheryl H. Arrowsmith*[,†,§,||]

*Contribution from the Ontario Centre for Structural Proteomics and Ontario Cancer Institute, University Health Network, and Department of Medical Biophysics, and Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada, Northeast Structural Genomics Consortium, and Midwest Center for Structural Genomics*

Received May 31, 2005; E-mail: carrow@uhnres.utoronto.ca

***Abstract:*** NMR spectroscopy and X-ray crystallography, the two primary experimental methods for protein structure determination at high resolution, have different advantages and disadvantages in terms of sample preparation and data collection and analysis. It is therefore of interest to assess their complementarity when applied to small proteins. Structural genomics/proteomics projects provide an ideal opportunity to make such comparisons as they generate data in a systematic manner for large enough numbers of proteins to allow firm conclusions to be drawn. Here we report a comparison for 263 unique proteins screened by both NMR spectroscopy and X-ray crystallography in our structural proteomics pipeline. Only 21 targets (8%) were deemed amenable to both methods based on an initial 2D $^{15}$N-HSQC NMR spectrum and optimized crystallization trials. However, the use of both methods in the pipeline increased the total number of targets amenable to structure determination to 107, with 43 amenable to NMR only and 43 amenable to X-ray crystallographic methods only. We did not observe a correlation between $^{15}$N-HSQC spectral quality and the success of the same protein in crystallization screens. Similar results were found for an independent set of 159 proteins as reported in the accompanying paper by Snyder et al.[1] Thus, we conclude that both methods are highly complementary, and in order to increase the number of proteins suited for structure determination, we suggest that both methods be used in parallel in screening of all small proteins for structure determination.

Genomic sequence information is being generated at an unprecedented rate. Structural proteomics (often called structural genomics) aims to solve the structures of gene products as a means to better understand their functions and the relationship between sequence, structure, and function. The major bottleneck in the structural proteomics pipeline remains the ability to generate recombinant proteins in a stable, soluble, and suitably concentrated form for X-ray crystallography and NMR spectroscopy. Various strategies have been used to improve protein solubility such as refolding,[2,3] variation of fusion tags,[4] screening of mutational variants,[5−7] and the use of orthologous proteins from a range of organisms,[8] but no single method is universally successful. Furthermore, only a fraction of the subset of proteins that can be purified in concentrated form yield to structural biology methodologies.

Currently, X-ray crystallography and NMR spectroscopy are the two most widely used experimental methods for structure determination of proteins. Our previous results[8−10] showed that NMR spectroscopy, despite its protein size limitation, can complement X-ray crystallography in structural proteomics.

We have sought to understand and optimize the application of these two experimental techniques in large-scale structure determination endeavors. The primary step in each method

[†] University Health Network.
[§] Department of Medical Biophysics and Banting and Best Department of Medical Research, University of Toronto.
[||] Northeast Structural Genomics Consortium.
[⊥] Midwest Center for Structural Genomics.

(1) Snyder, D.; Chen, Y.; Denissova, N.; Acton, T.; Aramini, J.; Ciano, M.; Karlin, R.; Liu, J.; Manor, P.; Paranji, R.; Rossi, P.; Swapna, G.; Xiao, R.; Rost, B.; Hunt, J.; Montelione, G. *J. Am. Chem. Soc.* **2005**, *127*, 16505.
(2) Maxwell, K. L.; Bona, D.; Liu, C.; Arrowsmith, C. H.; Edwards, A. M. *Protein Sci.* **2003**, *12*, 2073−80.
(3) Tresaugues, L.; Collinet, B.; Minard, P.; Henckes, G.; Aufrere, R.; Blondeau, K.; Liger, D.; Zhou, C. Z.; Janin, J.; Van Tilbeurgh, H.; Quevillon-Cheruel, S. *J. Struct. Funct. Genomics* **2004**, *5*, 195−204.
(4) Hammarstrom, M.; Hellgren, N.; van Der Berg, S.; Berglund, H.; Hard, T. *Protein Sci.* **2002**, *11*, 313−321.
(5) Maxwell, K.; Mittermaier, A.; Forman-Kay, J.; Davidson, A. *Protein Sci.* **1999**, *8*, 1908−1911.
(6) Pedelacq, J. D.; Piltch, E.; Liong, E. C.; Berendzen, J.; Kim, C. Y.; Rho, B. S.; Park, M. S.; Terwilliger, T. C.; Waldo, G. S. *Nature Biotech.* **2002**, *2*, 927−932.
(7) Waldo, G. S. *Methods Mol. Biol.* **2003**, *230*, 343−359.
(8) Savchenko, A.; Yee, A.; Khachatryan, A.; Skarina, T.; Evdokimova, E.; Pavlova, M.; Semesi, A.; Northey, J.; Beasley, S.; Lan, N.; Das, R.; Gerstein, M.; Arrowsmith, C.; Edwards, A. *Proteins* **2003**, *50*, 392−399.
(9) Christendat, D.; et al. *Nat. Struct. Biol.* **2000**, *7*, 903−909.
(10) Yee A.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1825−1830.


**10.1021/ja053565+ CCC: $30.25** 

involves protein purification and a subsequent screen for solution conditions optimal for data collection (NMR spectroscopy) or crystallization. NMR spectroscopists typically screen a small number of conditions, whereas protein crystallographers commonly screen hundreds.

The appearance of the initial $^{15}$N-HSQC NMR spectrum is usually a very good indicator of whether the sample is amenable to structure determination by NMR spectroscopy. Subsequent steps in the "NMR pipeline" focus on data acquisition and analysis of samples with a high likelihood of yielding a structure and usually take anywhere between 1 and 6 months in most structural genomics contexts. X-ray crystallography, on the other hand, usually requires substantial investment of time after initial crystallization conditions are established in order to optimize the diffraction properties of a crystal. This process can take weeks to months, but once a well-diffracting (<2.5Å) crystal is obtained, and an anomalous scattering atom is incorporated, the structure determination can proceed very quickly, within hours for some datasets of very high-resolution crystals.[11] Advances have been made in both fields in order to minimize these time investments, such as the use of more rapid data collection strategies,[12−14] automated data analysis,[15−17] nanotechnologies,[18−20] and the use of empirical correlations in order to increase the crystallization success rate.[21,22]

Thus, both fields are evolving rapidly, leading to the dilemma of whether one method should be used as the primary structural strategy for small proteins, and the other a secondary method (for example if both methods yield structures of the same proteins), or should both methods be used simultaneously (for example, if most small proteins "yield" to only one technique, but not both). To address this question, in 2002 we performed a limited study of 23 pairs of *E. coli* and *T. maritima* orthologous targets in which 46 unique proteins were screened by both methods.[8] The results indicated that the two methods are highly complementary with limited redundancy. Thus, to increase the number of structural samples from a given set of protein targets, both methodologies should be used in parallel.[8] Furthermore, the study showed that NMR was not a good predictor of the protein sample's behavior in crystallization trials, while failing to generate a crystal "hit" in the initial crystallization screen had no correlation with NMR behavior of a protein. The results from this limited study prompted us to modify our pipeline for small proteins so that both NMR and crystallization trials are used in parallel for initial evaluation of each protein. Thus, the number of samples that could be used to validate this approach has now swelled to 263 and although the percentage of the samples that have been successfully processed by both NMR and X-ray crystallography is higher than that in the original study, the overall conclusions remain the same. This paper focuses on the comparative analysis of the NMR screening and crystallization trial results of structural proteomics targets obtained as of April 2005.

## Materials and Methods

All target genes were subcloned from genomic DNA into a pET-derived vector with N-terminal hexa-histidine followed by either a thrombin or TEV protease cleavage site. All proteins were expressed in *E. coli* strain BL21-(Gold DE3), and in the case of the archaeal and eukaryote proteins the cells were cotransformed with a plasmid (pMgk) encoding three transfer RNAs for rare *E. coli* codons.[23]

**NMR Screening.** Cells were grown in either 1 L of M9 or 0.5 L of 2X M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source and supplemented with $ZnCl_2$, thiamine, and biotin. The cells were grown at 37 °C to an OD600 of 0.6 or 1.0, respectively, and protein expression was induced with 1 mM isopropyl beta-D-thiogalactoside. The temperature was reduced to 15 °C, and the cells were allowed to grow overnight before harvesting. Frozen cell pellets were thawed in 500 mM NaCl, 20 mM Tris, 5 mM imidazole (pH 8) and lysed by sonication. The proteins were extracted from the lysates by batch $Ni^{2+}$ affinity chromatography (Qiagen). The $Ni^{2+}$ affinity beads were washed 3 times with 5 column volumes of 500 mM NaCl, 20 mM Tris, 30 mM imidazole (pH 8), and the proteins were eluted with 5 column volumes of 500 mM imidazole in this same buffer. The purified proteins were concentrated, and buffer was exchanged by ultrafiltration and dilution/reconcentration. The NMR screening buffer was among the following: (a) 25 mM $Na_2PO_4$, pH 6.5, 450 mM NaCl; (b) 10 mM MOPS, pH 6.5, 450 mM NaCl; or (c) 10 mM NaAc, pH 5.0, 300 mM NaCl. All NMR buffers contain ~20 uM $Zn^{2+}$, 10 mM DTT, 1 mM benzamidine, 1 × inhibitor mixture (Roche Molecular Biochemicals), 0.01% $NaN_3$. For targets with a theoretical pI in the range of 6.0 to 7.0, buffer (c) was used (pH = 5.0); otherwise a pH 6.5 buffer was the generic NMR buffer.

All $^1$H−$^{15}$N heteronuclear single quantum coherence (HSQC) spectra were acquired at 25 °C by using a Varian INOVA 500- or 600-MHz spectrometer or a Bruker Avance500. The total number of $t_1$ increments was 64, with 8−64 scans per increment depending on the concentration of the sample being screened. The data were processed by using the NMRPIPE software package for those acquired from the Varian instruments and XWINNMR for those acquired from a Bruker instrument.

The HSQC spectra were classified into categories according to their suitability for NMR structure determination. The "good" HSQC spectra are indicative of folded proteins and contained well-dispersed peaks of roughly equal intensity and roughly the expected number of amide NH peaks. These spectra suggested that the process of determining their three-dimensional structures should be relatively straightforward. The "promising" spectra had features such as too few (~<80% of expected) or too many (~>110% expected) peaks and/or broad but dispersed signals. These proteins appeared to be folded, but the spectra suggested that optimization of either the protein construct or the solution conditions would be needed to yield a sample appropriate for NMR structure determination. The "poor" spectra had very low signal-to-noise relative to the amount of protein in the sample, and/or a single cluster of very broad peaks in the center of the spectrum. This category likely includes proteins that form soluble large entities under the NMR

(11) Minor, W.; Tomchick, D.; Otwinowski, Z. *Structure* **2000**, *8*, R105−R110.
(12) Kim, S.; Szyperski, T. J. *Biomol. NMR* **2004**, *28*, 117−130.
(13) Snell, G.; Cork, C.; Nordmeyer, R.; Cornell, E.; Meigs, G.; Yegian, D.; Jaklevic, J.; Jin, J.; Stevens, R. C.; Earnest, T. *Structure* **2004**, *12*, 537−545.
(14) Leslie, A. G.; Powell, H. R.; Winter, G.; Svensson, O.; Spruce, D.; McSweeney, S.; Love, D.; Kinder, S.; Duke, E.; Nave, C. *Acta Crystallogr., Sect. D* **2002**, *58*, 1924−1928.
(15) Gronwald, W.; Kalbitzer, H. R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *44*, 33−96.
(16) Terwilliger, T. C. *Methods Enzymol.* **2003**, *374*, 22−37.
(17) Ness, S. R.; de Graaff, R. A.; Abrahams, J. P.; Pannu, N. S. *Structure* **2004**, *12*, 1753−1761.
(18) Kuil, M. E.; Bodenstaff, E. R.; Hoedemaeker, F. J.; Abrahams, J. P. *Enzyme Microb. Technol.* **2002**, *30*, 262−265.
(19) Bodenstaff, E. R.; Hoedemaeker, F. J.; Kuil, M. E.; de Vrind, H. P.; Abrahams, J. P. *Acta Crystallogr., Sect. D* **2002**, *58*, 1901−1906.
(20) Santarsiero, B. D.; Yegian, D. T.; Lee, C. C.; Spraggon, G.; Gu, J.; Scheibe, D.; Uber, D. C.; Cornell, E. W.; Nordmeyer, R. A.; Kolbe, W. F.; Jin, J.; Jones, A. L.; Jaklevic, J. M.; Schultz, P. G.; Stevens, R. C. *J. Appl. Crystallogr.* **2002**, *35*, 278−281.
(21) Kimber, M.; Vallee, F.; Houston, S.; Necakov, S.; Skarina, T.; Evdokimova, E.; Beasley, S.; Christendat, D.; Savchenko, A.; Arrowsmith, C. H.; Vedadi, M.; Gerstein, M.; Edwards, A. M. *Proteins* **2003**, *51*, 562−568.
(22) Kantardjieff, K. A.; Rupp, B. *Bioinformatics* **2004**, *20*, 2162−2168.
(23) Baca, A. M.; Hol, W. G. *Int. J. Parasitol.* **2000**, *30*, 113−118 and references therein.

***Table 1.*** Results of the NMR and Crystallographic Screening of 263 Small Proteins[a]

| HSQC category | number of targets | number of targets with initial crystal hits (%) | structures solved by crystallography[b] (%) | crystallographic "work stopped" (%) | structures solved by NMR[b] (%) | NMR "work stopped" (%) |
|---|---|---|---|---|---|---|
| good | 64 | 21 (32.8) | 3 (4.7) | 11 (17.2) | 8 (12.5) | 26 (41) |
| promising | 33 | 6 (18.2) | 3 (9.1) | 2 (6.1) | n/a[c] | n/a |
| poor | 128 | 33 (25.8) | 14 (10.9) | 10 (7.8) | n/a | n/a |
| unfolded | 3 | 0 (0) | 0 (0) | 0 (0) | n/a | n/a |
| limited solubility | 35 | 4 (11.4) | 2 (5.7) | 1 (2.8) | n/a | n/a |

[a] Percentage is calculated relative to the number of targets from each category. [b] The PDB accession numbers are summarized in Table 2 (Supporting Information). [c] These targets are not suitable for structure determination by NMR

screening conditions such as higher order oligomers (tetramers or larger) or nonspecific aggregates, or proteins that are undergoing conformational fluctuations in the intermediate range of the NMR time scale. This category is not readily amenable to NMR structural analysis. The "unfolded" HSQC category displayed intense peaks but with little dispersion in the amide $^1$H dimension, most likely reflecting proteins that were soluble yet largely disordered. The "limited solubility" category includes spectra for which no HSQC signal was observed after 64 scans on a conventional probe or 32 scans on a low-temperature probe. This represents targets that have very low expression or that have poor solubility in the NMR screening buffer.

**Crystal Screening.** Large-scale expression of recombinant proteins was performed by growing cells in 1 L of LB with appropriate antibiotics in either a 6-L flask or a custom-baffled 4-L flask. The sample was induced at an OD600 of 0.6−0.8 with 0.4 mM IPTG after growth at 37 °C and grown overnight at 15 °C. The harvested cell pellet was resuspended into 40 mL with binding buffer, supplemented with 1 mM each of PMSF and benzamidine, flash-frozen in liquid nitrogen, and stored at −70 °C. The purification procedure used buffers containing 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, and 5, 30, and 250 mM imidazole for the binding, wash, and elution buffers, respectively. The harvested cells were lysed by adding 0.5% Nonidet P-40 to the thawed sample before sonication. The lysate was passed by gravity through a DE52 column in series with a $Ni^{2+}$ column. Contaminating proteins were removed by washing the $Ni^{2+}$ column with 50 column volumes of wash buffer. The bound protein was removed with elution buffer as qualitatively determined by the Bradford assay. The sample was then brought to a final concentration of 0.5 mM EDTA, followed by the addition of a final concentration of 0.5 mM DTT. The His-tag was removed by cleavage with recombinant His-tagged TEV protease (60 $\mu$g TEV/mg recombinant protein) or thrombin. The cleavage step was done concurrently with dialysis in binding buffer without imidazole at 4 °C overnight. The cut His tag and His-tagged TEV protease were removed from the purified recombinant protein by passage through a second $Ni^{2+}$ column. The sample was prepared for crystallization screening by a second dialysis in 10 mM HEPES, pH 7.5, 500 mM NaCl, followed by concentration using ultrafiltration.

Proteins that can be concentrated to at least 3 mg/mL were subjected to a sparse-matrix crystallization screen consisting of at least 96 conditions[21] at room temperature and/or at 4C.

## Results and Discussion

Since inception, our group at the Ontario Centre for Structural Proteomics has included both NMR spectroscopy and X-ray crystallography in our structural proteomics pipeline. Initially, we divided the targets according to protein size whereby smaller proteins go solely through the NMR spectroscopy pipeline and the rest go through the X-ray crystallography pipeline. Over time, we have modified our pipeline so that all small proteins are screened by both methods. As of April 2005, our group has attempted to purify over 3970 unique protein constructs, with about half of these (2017) yielding soluble protein. A total of 1175 were screened by crystallography, and 1105, by NMR. Of these, 263 targets have been screened by both NMR spectroscopy and X-ray crystallography, with sizes ranging from 44 to 379 residues. These proteins were classified by NMR as follows: good (24.3%), promising (12.5%), poor (48.7%), unfolded (1.1%), and 13.3% had limited expression or solubility in the NMR screening buffer (Table 1). Figure 1 shows examples of our HSQC classifications.

All of our NMR and crystallography data are collated under a common database. In our current laboratory procedures, all small targets are screened by NMR regardless of initial crystal screening results. Those that give good HSQC spectra are posted for spectroscopists to select for structure determination. All targets that give initial crystal hits are pursued for crystal optimization, regardless of the HSQC classification. When an NMR spectroscopist has selected a target for structure determination, then crystal optimization work is stopped. A decision to "stop work" for both methods is made when a structure of the protein or an ortholog of the protein is deposited in the PDB.

Well dispersed $^1$H and $^{15}$N resonances in an HSQC spectrum are a good indication that the protein adopts a globular fold and is tumbling freely and isotropically in a given NMR screening buffer. Although the HSQC classification is currently a subjective decision, there is a clear distinction between the "good" and the "poor" HSQC category. The "promising" category is a "gray" area between these two categories and may have features of both.

Of the 64 targets with good HSQC spectra, only 21 showed positive crystal screen results. So far, 3 have been solved by X-ray crystallography, 2 have been "claimed" and subsequently solved by NMR, and work was stopped for 11 targets due to significant homology to structures recently deposited in the PDB. The remaining 5 targets are still in the crystal optimization pipeline and also on the list of good HSQC samples available for NMR analysis. Of the remaining 43 "good HSQC" targets that did not crystallize, NMR structures have been solved for 6, and 14 now have homologues in the PDB and will likely not be pursued further.

Almost half of the targets we screened for NMR spectroscopy gave poor HSQC spectra. These proteins have good solubility in the NMR screening buffer, but they showed very poor dispersion in the HSQC spectra and the signals are very broad. The NMR spectral line width is highly dependent on the molecule's tumbling rate in solution and therefore on its molecular weight. So a protein exhibiting a poor HSQC could suggest that the protein is forming a higher molecular weight entity in a given buffer. These higher molecular weight entities

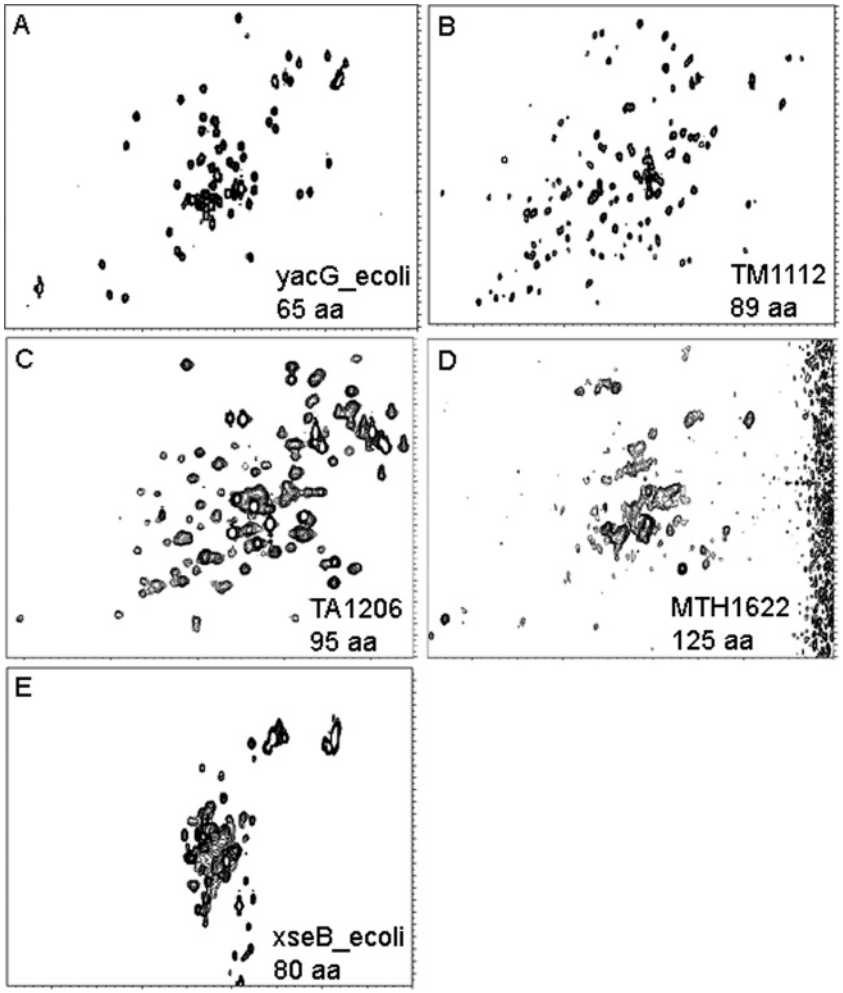


***Figure 1.*** Sample $^{1}H-^{15}N$-HSQC spectra of different classifications: (a) Good. This target failed to give any initial crystal hits, and the structure was solved by NMR spectroscopy, PDB_ID 1LV3. (b) Good. This target had an initial crystal hit, but the structure was solved by NMR spectroscopy, PDB_id 1LKN. (c) Promising. This structure was solved by X-ray crystallography, PDB_id 1QW2. (d) Poor. This structure was solved by X-ray crystallography, PDB_id 1PBJ. (e) Unfolded. This target failed to give initial crystal hits. The *x*-axis shows the $^{1}H$ chemical shift (ppm), and the *y*-axis shows the $^{15}N$ chemical shift (ppm).

could be in the form of nonspecific aggregates or discrete higher order homo-oligomers. Resonance line broadening and consequent poor sensitivity can also be due to dynamic properties of proteins that are fluctuating between multiple conformations at an intermediate rate on the chemical shift time scale. Unfortunately, $^{15}N$-HSQC spectra alone cannot distinguish between these different situations.

Initial crystal screens showed 33 positive crystal hits from the "poor" HSQC class, and out of these, 14 targets were optimized and, subsequently, structures were solved. Optimization of 10 targets was stopped because a homologue was deposited in the PDB.

In the promising HSQC category, 18% of the targets gave initial crystal hits and so far 9% of the targets were refined and structures were solved by X-ray crystallography.

A possible explanation for poor quality spectra of proteins that yielded well diffracting crystals could be that these proteins are oligomeric. However, examination of the apparent oligomeric states of the proteins solved by crystallography suggests that the slow tumbling and consequent broadened line widths of homo-oligomers can only account for a subset of the poor HSQC spectra (Figure 2). We estimate that symmetric oligomers

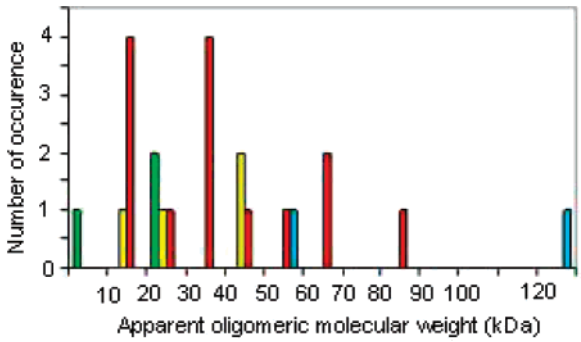


***Figure 2.*** Apparent size of the proteins screened by NMR spectroscopy but for which structures were solved by X-ray crystallography. (Green) Those that gave good HSQC spectra; (yellow) gave promising HSQC spectra; (red) those that gave poor HSQC spectra; and (blue) those that had limited solubility in the NMR screening buffer. The apparent oligomeric molecular weight was the product of the monomeric molecular weight and the apparent oligomeric state as observed in the crystal, based on the biological unit suggested in the PDB deposition. For those structures in which no remark on the biological unit was mentioned in the PDB deposition, the oligomerization state was based on the number of chains in the crystallographic asymmetric unit.

of up to ~40 kDa can still be identified as good or promising using conventional (non-TROSY) HSQC spectra, although their

structure determination would require partial and/or selective deuteration.[24]

No crystal hits were observed for the unfolded HSQC category. Collectively, 164 targets were deemed unsolvable by NMR, but 39 of these gave crystal hits and are currently being optimized for crystallization and 17 structures have been successfully solved by X-ray crystallography. Likewise, if we were to use only crystallography, we would have missed the 43 targets that gave good HSQC spectra but did not crystallize.

The limited solubility category warrants a separate discussion because for these targets, under our current NMR screening protocol, we failed to see sufficient signal intensity to classify their HSQC. Our choice of NMR buffer to screen the protein is based on the protein's theoretically calculated isoelectric point (pI). It is known that solubility of a protein tends to decrease in solutions with a pH near its pI. The targets with good, promising, and poor HSQC spectra had an average pI of 6.2−6.3 with only about 30% having a pI greater than 6.5. Interestingly, the targets that have limited solubility in the NMR screening buffer have a higher average pI of 7.49 than the rest of the categories with 68% having a pI greater than 6.5. Figure 3 shows plots of the target's theoretical pI for the different categories for comparison.

The soluble targets included in this study were screened in a single NMR buffer condition as compared with at least 96 initial crystallization buffer conditions with pH values between 3.5 and 8.5. It is interesting to note that, for the targets that had good HSQC spectra and for which we have detailed crystallization records, only half of the crystallization conditions had a pH close to that of the NMR buffer. These data suggest that the conditions that favor good HSQC spectra are not necessarily related to those that promote crystal formation. This can be rationalized by the fact that good HSQC spectra require that proteins are not associated with one another (except for stable oligomers) and are tumbling isotropically, while crystallization requires highly ordered self-association. It stands to reason that solution conditions that favor the former may not be consistent with the latter.

We also examined all the targets that gave crystal hits, with respect to their theoretical pI (magenta/yellow data points in Figure 3 and Table 3, Supporting Information). Similar to the study of Cavanes et al.,[25] we find that very few proteins with a high pI (pI > 10) crystallized. Furthermore, proteins in the pI range 4−7 were slightly more likely to form crystals (~27%) than those in the pI range 7−10 (~22%). However, the total number of proteins in the higher pI ranges is small, and these differences may not be statistically significant. Our current initial crystal screening conditions span pH's between 3.5 and 8.5. We were able to obtain well diffracting crystals and subsequently structures for proteins with pI's of 4.4 to 8.8. One of the major differences between our work and that of Cavanes et al. is that our data are restricted to small proteins among many organisms, whereas Cavanes et al. focused on several hundred proteins from a single thermophilic organism.

NMR spectroscopy and X-ray crystallography are the two methods used within structural proteomics projects. Here we show that by using both methodologies in a coordinated fashion it is possible to almost double the number of structures of small proteins that can potentially be solved, compared to either method alone. Structure determination by X-ray crystallography

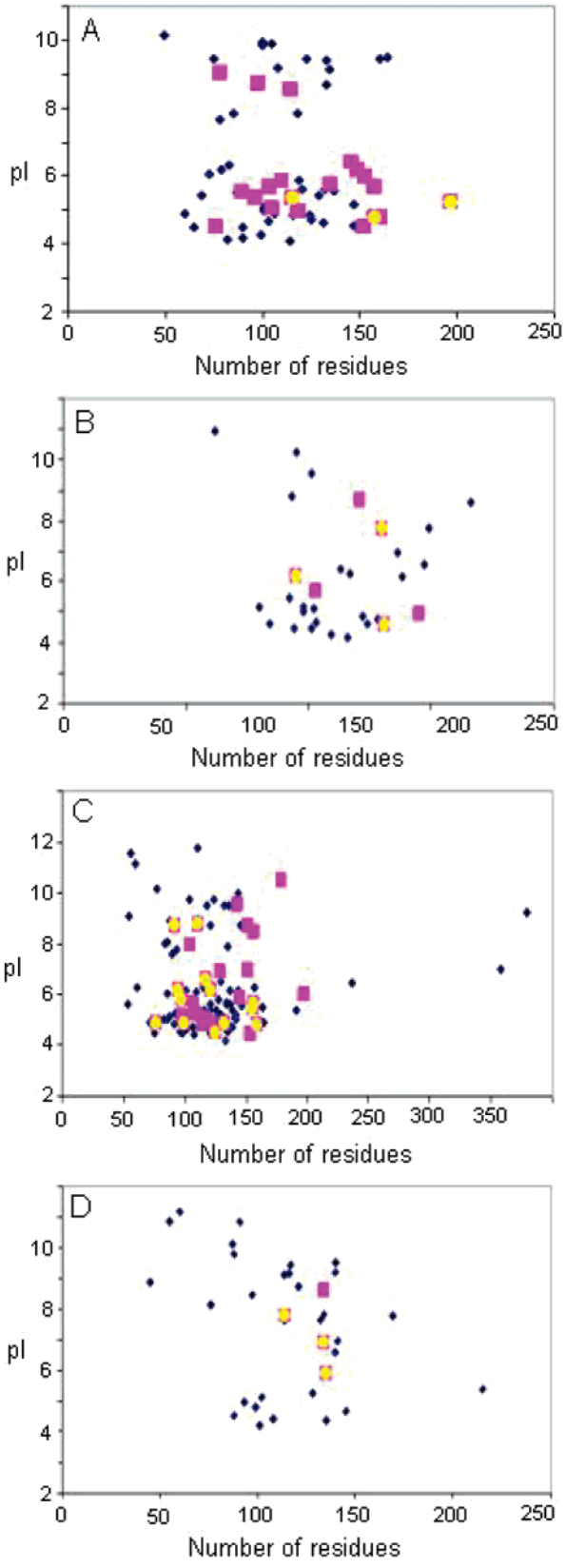


***Figure 3.*** Protein molecular weight and isoelectric point distribution of the targets in this study as classified according to their NMR HSQC spectra. (a) Good; (b) Promising; (c) Poor; (d) Limited solubility. Those that gave initial crystal hits are highlighted with magenta, and those that were refined and structures were solved are highlighted by yellow circles.

(24) Tugarinov, V.; Hwang, P. M.; Kay, L. E. *Annu. Rev. Biochem*. **2004**, *73*, 107−146.
(25) Cavanes, J.; Page, R.; Wilson, I.; Stevens, R. *J. Mol. Biol*. **2004**, *344*, 977−991.

is faster than that by NMR spectroscopy provided that a well diffracting crystal is available. However, screening for conditions that yield well diffracting crystals is still a trial and error method that takes significant resources. In contrast, screening by HSQC NMR is quick, and the results are highly indicative of whether the solution structure can be determined by NMR. Other biophysical methods such as circular dichroism, dynamic light scattering, and one-dimensional NMR have been reported as tools to prescreen targets for crystallization. Our data and that of the accompanying paper[1] suggest that although the HSQC spectrum is a good indicator of a protein's foldedness and its amenability to NMR structure determination, it is not a good indicator of a protein's crystallizability or its ability to be solved by X-ray diffraction. This result also agrees with the recent data from a set of *A. thaliana* targets in which no correlation was observed between HSQC quality and the results of crystallization screens.[26] These results may be compared to recent data from the Joint Center for Structural Genomics in which the diffraction quality of protein crystals was found to correlate with the classification of its one-dimensional $^{1}$H NMR spectrum.[27] The primary difference between our approach and the 1D screening approach is that the latter is primarily evaluating the "foldedness" or globularity of a protein, while the HSQC approach not only evaluates foldedness but also evaluates the protein's suitability for NMR structure determination. There are other technologies, such as Circular Dichroism Spectroscopy, or susceptibility to limited proteolysis, that can be used to assess the degree of foldedness or globularity of a protein sample. However, only NMR can also be used to obtain high-resolution structural information. Based on this, we conclude that a coordinated, parallel strategy whereby all targets are screened by both 1D and 2D-$^{15}$N-HSQC NMR and crystallography will provide a more efficient strategy for structure determination of small proteins than a tandem strategy whereby NMR or other biophysical methods are used as a prescreen for crystallography or vice versa.

**Acknowledgment.** We thank all members of the Ontario Center for Structural Proteomics for their contributions, cloning, expression, and preparation of the proteins analyzed in this paper. We thank G. Montelione and colleagues for sharing drafts of the accompanying manuscript (Snyder et al.) prior to publication. This research was supported by the Ontario Research and Development Challenge Fund; Genome Canada; the Protein Structure Initiative of the National Institutes of Health (P50-GM62413 and GM62414-01) through the Northeast Structural Genomics Consortium and Midwest Center for Structural Genomics. C.H.A. holds a Canada Research Chair in structural genomics, and A.M.E. holds the Banbury Chair of Medical Research.

**Supporting Information Available:** Table 2: PDB accession numbers of the structures included in this studies. Table 3: Breakdown of the number of targets according to the NMR classification, theoretical pI, and crystal screening results. Reference section: Complete list of authors for cited refs 8, 9, and 26. This material is available free of charge via the Internet at http://pubs.acs.org.

JA053565+

(26) Tyler, R.; et al. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 633−643.
(27) Page, R.; Peti, W.; Wilson, I.; Stevens, R.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1901−1905.