

**Northeast Structural Genomics Consortium**  
**Structural Proteomics In the NorthEast (SPINE)**  
<http://spine.nesg.org>

SPINE (Structural Proteomics In the NorthEast) is the NESG's inter-institutional project coordination database. SPINE is based on a three-tier software architecture. It is a web-based data management application built upon a MySQL database driven by a series of Perl scripts and Perl modules hosted by an Apache web server. Initially developed to manage the NESG protein production pipeline, SPINE has evolved to include components designed to coordinate activities and projects in the several laboratories across the consortium. The Oracle-based PLIMS (Protein Laboratory Information System) manages the majority of the NESG's protein production pipeline at Rutgers University.

The SPINE data model (Bertone et al., 2001; Goh et al., 2003) closely mirrors the current pepcDB data model, with database tables for tracking target, construct, expression, purification, biophysical characterization, X-ray and NMR data, and protein structure data. Spine generates weekly targetDB and pepcDB data deposition files which are retrieved by the PSI-KB. Spine also generates meta-data required for archiving NESG constructs at the PSI-MR, and provides key NESG protein production pipeline, sample distribution, and structure determination statistics. It is the primary tool for moving information from the NESG project to public databases.

### Consortium Coordination

As a web-based application, SPINE is also the primary tool for communicating information between the laboratories of the NESG consortium. The SPINE database tracks *protein samples* created for structure screening and structure determination throughout the consortium via PST (Protein Sample Tube) ids. The

SPINE  
Structural Proteomics in the NorthEast

Tools | Editor login | Modeling families | Target families | Aggregation screening | PSTs | Shipments | Data mining | NMR screening | Statistics

pst record for pst: CgR26A.005

- target record: [CgR26A]
- construct record: [CgR26A-41-180-21.3]
- expression record: [CgR26A-41-180-21.3-NCa]
- purification record: [CgR26A-41-180-21.3-NCa-GF]

PST label: CgR26A-41-180-6xHis  
Current location: University at Buffalo  
PST buffer: NMR 6.3-200 mM NaCl  
PST volume: 0.3 mL  
PST protein concentration: 0.94 mM  
PST tube format: NMR tube

Sequence (purification tags are colored blue, mutant residues are colored red)  
MTETVLAESPEFYQDNVTDYTGQISSDITNIQAIDDVKASEQRKVFVY  
FLSSFDGVDPEFTWQQALQANGGNVLIYALAPEERQYGIQGGTQWTDAL  
DAANNAAFQALSQEDWAGSALALAEVSGSSSSSSSSSSSLEHRRHHH

Sequence stats (149 residues)  
Met: 0 Gly: 11 Cys: 0 Pro: 3 Asn: 6  
Ile: 7 Leu: 10 Val: 9 Trp: 3 Gln: 12

Available construct sequencing results:  
CgR26A-41-180-21.3-NC5a-21999F1.ab1  
CgR26A-41-180-21.3-NC5a-22027R1.ab1

(Strip this sample)  
Last shipment sent to: Buffalo (2009-08-27)

Create new structure record: (HSQC) (NMR) (Xray)

Edit Delete

entire sample creation and evolution history are available via PST ids (Figure 1).

Researchers across the NESG consortium use SPINE to retrieve protein sample data, track the progress of construct and fermentation requests, and for various data mining projects.

**Fig. 1.** SPINE's Protein Sample Tube (PST) records provide details about NESG protein samples, and link to the sample's Target, Construct, Expression, Purification, Crystallization, NMR Screening, and 3D Structure SPINE records.

SPINE organizes samples created for NMR screening into 96 sample blocks, programs NMR screening robots, and reports and archives the results of NMR screening runs. Targets that yield well dispersed  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, which can be viewed by co-investigators across the consortium, are enriched with  $^{15}\text{N}$ ,  $^{13}\text{C}$  isotopes and distributed to the NESG NMR groups. Spine also tracks NMR sample distribution and NMR structure progress (Figure 2). Spine organizes and prepares shipments of selenomethionine labeled targets to Hauptman-Woodward Medical Research Institute (HWI) for high throughput crystallization trials. Successfully Identified crystallizations conditions are reported to both SPINE the NESG's crystallization database (PROTEUS).

### Data Integration

Spine is the central archive of all reagent and structure data generated by the NESG. It is closely intergrated with ZebraView (Wunderlich et al., 2004)– the public access point to the NESG protein target list. Target, clone and expression data generated in the sample production pipeline (Acton et al., 2005) are reported to Spine from PLIMS. Predicted (e.g. Zn-binding motifs and disorder predictions) and experimental (e.g. gel filtration with static light scattering) biophysical data are also stored for each protein target. Crystallization progress is reported to Spine from the PROTEUS database, while NMR progress across the

consortium is tracked directly by SPINE. Atomic coordinates, structure validation reports, and raw structural data including FID, chemical shift, constraint, and crystallographic structure factor data, are also archived in SPINE using HarvestDB, the protein structure deposition interface to Spine. Target data and experimental data sets such as the publicly-accessible NESG RDC library and NMR FID libraries are also made freely available to the scientific community through SPINE. The resulting archive, containing data for more than 20,000 cloned constructs, is a valuable resource for datamining the structural genomics process (see for example Goh et al., 2004, Price et al., 2009)

**SPINE**  
Structural Proteomics in the NorthEast

Tools | Editor login | Modeling families | Target families | Aggregation screening | PSTs | Shipments | Data mining | NMR screening | Statistics | Upload | Reports | Structure gallery

**NMR progress**

Sort assignments by: status

Show statuses from C - Assigned to NMR Group / Ordered CN Sample to P - Target is problematic

target apparent size codings:  
 25 - 110 aa  
 111 - 160 aa  
 >160 aa or oligomer

Click on a target id to update the form below the progress table. This form is used to update assigned target information. This form can also be used to add targets to the table.  
 Structure counts are based on the [structure gallery](#) which is populated via [HarvestDB](#).

Montelione			Szyperski			Kennedy			Arrowsmith			Prestegard		
4 targets			3 targets			5 targets			2 targets			1 targets		
(5 pdb ids)	/	18 Yr 5 goal structures	(3 pdb ids)	/	18 Yr 5 goal structures	(5 pdb ids)	/	18 Yr 5 goal structures	(3 pdb ids)	/	12 Yr 5 goal structures	(1 pdb ids)	/	4 Yr 5 goal structures
<a href="#">OR16</a>	102	C Monomer	<a href="#">BuR95A</a>	55	C Monomer	<a href="#">DhR29B</a>	79	C Monomer	<a href="#">att12</a>	112	E Monomer	<a href="#">GmR33A</a>	109	C Dimer
<a href="#">BcR147A</a>	84	C Monomer	<a href="#">BhR97A</a>	98	C Monomer	<a href="#">CrR16A</a>	111	C Monomer	<a href="#">TR17</a>	140	E Monomer	<a href="#">RcR213D</a>	100	D Dimer
<a href="#">HR3057H</a>	55	D Monomer	<a href="#">ErR9A</a>	109	C Monomer	<a href="#">LpR91B</a>	101	D Dimer	<a href="#">SFR81</a>	89	F Monomer	<a href="#">GcR123A</a>	57	D Dimer
<a href="#">LurR17A</a>	101	D Monomer	<a href="#">cvt2</a>	100	D Monomer	<a href="#">TaR102</a>	95	E Monomer	<a href="#">att13</a>	118	F Monomer	<a href="#">HR309A</a>	165	P Monomer
<a href="#">HR3486E</a>	84	D Monomer	<a href="#">DhR1A</a>	129	D Monomer	<a href="#">AtT11</a>	124	E Dimer	<a href="#">AtT1</a>	142	F Monomer			
<a href="#">MvR76</a>	158	D Monomer	<a href="#">CgR26A</a>	114	D Monomer	<a href="#">DNR9C</a>	62	E Dimer	<a href="#">PsT1A</a>	122	P Monomer			
<a href="#">LkR112</a>	140	D Monomer	<a href="#">CrR148A</a>	68	F Monomer	<a href="#">CvT4</a>	82	P Monomer	<a href="#">PaT3B</a>	104	P Dimer			
<a href="#">SgR46</a>	112	D Monomer	<a href="#">HR4403C</a>	149	P Monomer	<a href="#">BR194</a>	80	P Monomer	<a href="#">TR825</a>	88	P Monomer			
<a href="#">NeR103A</a>	82	D Monomer	<a href="#">GmR223A</a>	91	P Monomer									
<a href="#">HvR112</a>	110	E Dimer												
<a href="#">HR6276</a>	56	E Monomer												
<a href="#">OR3</a>	106	E Dimer												
<a href="#">HR4531</a>	728	E Monomer												

**Figure 2.** The NMR Target Progress Table is one of SPINE's project coordination tools, providing details of all active NESG NMR targets. SPINE's NMR progress table tracks the NMR targets assigned to the NESG NMR groups at several institutions, and their progress in the NMR structure determination process (preparation, data collection, refinement, problematic, and In PDB).

Bertone, P.; Kluger, Y.; Zheng, D.; Edwards, A.M.; Arrowsmith, C.H.; Montelione, G.T.; Gerstein, M. *Nucleic Acids Res.* 2001, 29: 2884 - 2898. SPINE: An integrated tracking database and data mining approach for prioritizing feasible targets in high-throughput structural proteomics.

Goh, C.-S.; Lan, N.; Echols, N.; Douglas, S.; Milburn, D.; Bertone, P.; Xiao, R.; Ma, L.-C.; Zheng, D.; Wunderlich, Z.; Acton, T.; Montelione, G.T.; Gerstein, M. *Nucleic Acids Res.* 2003, 31: 2833 - 2838. SPINE 2: A system for collaborative structural proteomics within a federated database framework.

Wunderlich, Z.; Acton, T.B.; Liu, J.; Kornhaber, G.; Everett, J.; Carter, P.; Lan, N.; Echols, N.; Gerstein, M.; Rost, B.; Montelione, G.T. *PROTEINS: Struct. Funct. Bioinformatics* 2004, 56: 181-187. The protein target list of the Northeast Structural Genomics Consortium.

Acton, T.B.; Gunsalus, K.C.; Xiao, R.; Ma, L.-C.; Aramini, J.M.; Baran, M.C.; Chiang, Y.-W.; Climent, T.; Cooper, B.; Denissova, N.; Douglas, S.M.; Everett, J.K.; Ho, C.K.; Macapagal, D.; Paranjli, R.K.; Shastry, R.; Shih, L.-Y.; Swapna, G.V.T.; Wilson, M.; Wu, M.; Gerstein, M.; Inouye, M.; Hunt, J.F.; Montelione, G.T. *Meth. Enzymology* 2005, 394:210-243. Robotic v cloning and protein production platform of the Northeast Structural Genomics Consortium.

Goh, C.-S.; Lan, N.; Douglas, S.; Wu, B.; Bertone, P.; Echols, N.; Milburn, D.; Montelione, G.T.; Zhao, H.; Gerstein, M. *J. Mol. Biol.* 2004, 336: 115-130. Mining the structural genomics pipeline: Identification and analysis of protein properties that affect high-throughput experimental analysis.

Price, W.N.; Chen, Y.; Handelman, S.K.; Neely, H.; Manor, P.; Karlin, R.; Nair, R.; Liu, R.; Baran, M.; Everett, J.; Tong, S.N.; Forouhar, F.; Swaminathan, S.S.; Acton, T.; Xiao, R.; Luft, J.R.; Lauricella, A.; DeTitta, G.T.; Rost, B.; Montelione, G.T.; Hunt, J.F. *Nature Biotechnology* 2009, 27: 51 - 57. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data