Synchrotron protein crystallography (PX) has been effective for the determination of the structures of biological macromolecules, especially complexes, membrane proteins and large molecular assemblies. Highly bright X-rays provide the special needs for those small or weakly diffracting crystals toward their structures at atomic resolution. The following research highlights are selected from the life sciences and biological macromolecular structures published by the user communities of year 2016. Six reports include moenomycin biosynthesis of prenyltransferase MoeN5 by Rey-Ting Guo, fidelity modulation of DNA polymerase λ by Wen-Jin Wu and Ming-Daw Tsai, marine lectin with the anti-cancer activity by Shang-Te Danny Hsu and Shih-Hsiung Wu, cyclic di-GMP receptor MshEN domain with nucleotide binding by Shan-Ho Chou, dimeric endonuclease G in a complex with DNA by Hanna S. Yuan, and the powerful phasing method with molecular averaging by Chun-Jung Chen.

At the recently operated TPS, a new beamline for micro-beam protein crystallography has been open to the domestic and international user community of structural biology since September, 2016. Combining this advanced PX beamline at TPS with other existing PX and bio-related beamlines, such as SAXS, EXAFS, CD and IR, at TLS, increasingly complicated and challenging structures of biological macromolecules can be resolved at high resolution in the near future. (by Chun-Jung Chen)
Lectins are proteins that bind carbohydrates and are observable in various species, such as animals, plants and bacteria. As lectins are capable of recognizing the specific functional group of a carbohydrate on a surface of potential pathogens, they play vital roles in the innate immune response. Regarding carbohydrate binding, lectins apply varied ways, including hydrogen bonds, van der Waals forces, hydrophobic interactions and metal coordination, to bind their target sugar moieties.

According to the various combinations of lectin-carbohydrate binding, lectins have been classified into many superfamilies, including galectins, C-type, I-type, P-type, pentraxins and so on. In this report, a novel lectin, from sea mollusk *Crenomytilus grayanus*, was purified and characterized. The *C. grayanus* lectin (CGL) is proposed to belong to the galactose-binding lectin family because of its great galactosamine (GalNAc) and galactose (Gal) specificity. The protein sequence, however, is not similar to that of galectins or other lectins except MytiLec (88% sequence identity). Structure modelling shows that CGL has a common β-trefoil fold that is found in cytokins, agglutinins and actin-cross-linking proteins. From an evolutionary point of view, these proteins have varied ligand-binding preferences and biological functions. To study CGL in detail, a research team led by Shih-Hsiung Wu (Academia Sinica and National Taiwan University) thus solved the structure of CGL in apo form and in a complex with various ligands, including galactose, galacosamine and globotriose, using Taiwanese beamline SP 12B2 at SPring-8.

Based on the overall structure, CGL has a typical β-trefoil fold consisting of 13 β-strands and three α-helixes. Figure 1(a) reveals that CGL can fold into three subdomains, each of which possesses one carbohydrate-binding site. To investigate that the homodimer form does not result from crystal packing, size-exclusion-chromatography multi-angle light scattering (SEC-MALS) and analytical ultracentrifugation analysis with a sedimentation velocity mode (SV-AUC) were performed; both experimental data clearly indicate that CGL has a dimeric form in solution.

To elucidate the ligand-binding mode, the authors subsequently solved the structures of two complexes, CGL halts the growth of a tumor cell through recognizing that globotriose, Gb3, exists abundantly on a cancer cell surface.
CGL-Gal (1.56 Å) and CGL-GalNAc (1.70 Å). Based on the CGL-GalNAc structure, three GalNAc-binding sites were discovered in each CGL monomer; their superposition within the same monomer shows that all these sites utilize almost the same amino acids to bind ligands. Several water molecules are involved in the water-mediated ligand interaction (Fig. 1(b)). Similar binding modes were found in the CGL-Gal structure. To elucidate the carbohydrate-binding events, a NMR titration experiment was conducted. Consistent with the crystallographic conditions, the concentration between ligands (Gal and GalNAc) and CGL was 5 mM versus 0.1 mM. On the basis of the chemical-shift perturbation result, two conclusions were that the residues involved in ligand binding were observed to undergo significant motion, and that the overall binding modes and binding sites of both carbohydrates are the same because of the similar patterns of chemical shifts (Fig. 1(c)). On comparison of NMR and crystallographic data, these findings are mutually consistent.

According to previous work, MytiLec recognizes an important sugar, globotriose (Gb3), that exists abundantly on the surface of a cancer cell. Interestingly, CGL shares 88% sequence identity with MytiLec, which means that CGL might have the ability to bind the Gb3. The high-resolution CGL-Gb3 structure proves the hypothesis: the binding mode is primarily through the terminal galactose moiety (Fig. 1(d)). The authors subsequently selected MCF7, the breast cancer cell line with abundant Gb3 on the cell surface, to test whether CGL could inhibit the cell viability of MCF7. A detachment of MCF7 cells was certainly observed upon addition of CGL (Figs. 2(a) and 2(b)). A 3D fluorescence-imaging reconstruction with fluorescently labelled CGL shows that CGL almost bound on the surface of MCF7 cells (Figs. 2(c)-2(e)). A cytotoxicity assay indicated that CGL has a dose-dependent anti-cancer activity (e.g. 33% cell death with 200 μg/mL CGL) (Fig. 2(f)). At this point, a lectin-based biosensor could be developed for cancer diagnosis. (Reported by Chun-Hsiang Huang)

This report features the work of Shih-Hsiung Wu and his co-workers published in J. Am. Chem. Soc. 138, 4787 (2016).

SP 12B2 BM – Protein X-ray Crystallography
• Protein Crystallography
• Biological Macromolecules, Protein Structures, Life Sciences

| References |
Cyclic dimeric GMP (c-di-GMP), a second messenger, is involved in vital cellular events, such as the formation of a biofilm, the biogenesis of pili and secretion of pathogenic factors in various bacteria. Although the mechanisms of c-di-GMP biosynthesis and degradation have been studied thoroughly, the nature of c-di-GMP receptors and the process of c-di-GMP-mediated regulation remain elusive.

Recent work on MshE (from *Vibrio cholerae*) and its homologue PA14_29490 (from *Pseudomonas aeruginosa*) showed both to be potent c-di-GMP-binding ATPase. Both proteins utilize N- and C-terminal domains for c-di-GMP and ATP binding, respectively. Arg9 and Gln32 of MshE serve notably as c-di-GMP binding residues, but the two residues were not observed in any previously described canonical c-di-GMP binding motif. To discover whether MshE uses a special binding mode for c-di-GMP binding, a collaborative team led by Shan-Ho Chou (National Chung Hsing University) hence solved the complex structure of the N-terminal domain of MshE (*Vc*MshEN) with c-di-GMP using a 1.37-Å data set at TLS 15A1 of NSR-RC and SP 44XU of SPring-B.2

**Figure 1(a)** shows that each MshEN structure comprises two subdomains, a four-helix domain (α1-α4, called MshEN_N) and an α/β domain (α5-α7 and β1-β3, called MshEN_C). To compare the two subdomains, the entire interaction between MshEN and c-di-GMP was contributed from MshEN_N except for Asp108 (**Fig. 1(b)**). c-di-GMP is thus bound mainly within the MshEN_N, in contrast, the role of MshEN_C in c-di-GMP binding requires further investigation.

Next, to test whether the MshEN domain is specific for c-di-GMP binding, the authors altered the ligand from c-di-GMP to c-di-AMP. The big difference between them is that O6 atom (H-bond acceptor) in c-di-GMP converts into a bulky amino group (H-bond donor), which causes a steric hindrance with the Asp12 amide proton of MshEN_N (**Fig. 1(c)**). Isothermal calorimetry (ITC) experiments proved that the MshEN domain cannot bind c-di-AMP because of the steric clashes (**Figs. 1(d) and 1(e)**). Differential scanning fluorimetry (DSF) was also performed to test whether other nucleotides containing adenine and guanine bases, including cAMP, ATP, ADP and AMP, and cGMP, GMP, GDP and GTP, are effective ligands for the MshEN domain. The DSF results confirmed clearly that the MshEN domain binds with no other nucleotide containing adenine or guanine.

Sequence alignment of the c-di-GMP-binding motifs showed that 12 residues, including R9, L10, G11, L25, L29, Q32, R38, L39, G40, L54, L58 and Q61, are highly conserved and play a key role in c-di-GMP binding. To investigate which conserved residues provided the greatest contribution to the c-di-GMP binding, the authors prepared a series of variants, such as single, double or triple MshEN_N mutants, for further testing using ITC and DSF. The ITC results revealed that hydrophobic interactions are important in c-di-GMP binding, because the mutated hydrophobic residues resulted in a binding affinity less than one tenth (**Figs. 1(f) and 1(g)**). To compare with preceding work, the mutated charged residues, including Arg9, Asp12 and Gln32, showed smaller effects on the c-di-GMP binding; the binding affinity seemed to decrease less than five-fold.3 In contrast, the DSF results indicated that most hydrophobic residues with single, double

---

**MshE: An Important Cyclic di-GMP Receptor and Its Involvement in Formation of a Biofilm of *Vibrio cholerae***

*Mutating the highly conserved residues of MshE significantly diminish c-di-GMP binding and biofilm formation.*
Fig. 1: Overall structure, binding mode and ITC assay of MshEN. (a) Two-subdomain structure, with c-di-GMP (shown in stick) bound primarily within the MshEN_N domain. (b) Interaction network of a MshEN-c-di-GMP complex. Water oxygen is shown as red sphere. (c) Hydrophobic interactions in c-di-GMP binding. Substitution of atom O6 with an amino group of c-di-GMP would cause steric hindrance. (d) Tight binding between MshEN and c-di-GMP by ITC. (e) No detectable binding of c-di-GMP was observed with ITC. (f)-(g) L10A and G11L resulted in significantly decreased c-di-GMP-binding affinity with ITC. [Reproduced from Ref. 2]

and triple mutation were involved in the stability of this complex structure. In conclusion, the strong binding between MshEN and c-di-GMP results from a unique interaction mode.

The authors subsequently ascertained the biological importance of the hydrophobic residues participating in c-di-GMP binding through analysis of various phenotypes, including pilin production, biofilm formation and motility, in relation to the MshE function. Based on these results, the defective phenotypes resulting from lack of mshE were restored by the expression of wild-type mshE in trans. Some mutant-type mshE, including L10A, G11I and L10/G11A, significantly affected, however, the ability of ΔmshE to perform normal phenotypes. (Fig. 2). Taken together, the results in vitro and in vivo indicated that binding of c-di-GMP is crucial for MshE activity. (Reported by Chun-Hsiang Huang)

This report features the work of Shan-Ho Chou and his co-workers published in Nat. Commun. 7, 12481 (2016)
Fig. 2: Analysis of motility, pilus production and biofilm formation. *P ≤ 0.05, **P ≤ 0.01, *** P ≤ 0.001. EV: empty vector. WT: wild-type. ΔmshE: mshE deletion mutant. [Reproduced from Ref. 2]

TLS 15A1 Biopharmaceutical Protein Crystallography
• Protein Crystallography
• Biological Macromolecules, Protein Structures, Life Sciences

| References |
MoeN5: An Important Prenyltransferase Involved in Moenomycin Biosynthesis

Based on the structural basis of MoeN5, it can facilitate the design of new antibiotics.

Moenomycins, first described about 60 years ago, are potent antibiotics because of their ability to inhibit effectively the biosynthesis of bacterial cell walls. Considering the global increase in antibiotic resistance, Moenomycin A (MmA, 1; Fig. 1), a member of the moenomycin family, has become a renewed target for the further development of MmA analogues because of its poor pharmacokinetic property.

In the MmA biosynthesis, 17 enzymes participate in the complicated process. Figure 1 clearly shows that MoeN5 catalyses the reaction of geranyl diphosphate (GPP, 6) with the cis-farnesyl group in phosphoglycolipid (FPG trisaccharide, 5) to produce (C25) moenocinyl-sidechain-containing lipid 7. (C25) moenocinyl trisaccharide 7 subsequently undergoes further transformation to give rise to the formation of 1. Among the biosynthesis paths, the authors in this report have solved the crystal structure of MoeO5 (involved in reaction 3→4); they also attempted to determine the X-ray structure of MoeN5. To elucidate the mechanism of the uncommon head-to-middle prenyltransferase, they proceeded to determine the structures of MoeN5 in complexes with GPP, FPG and a series of model glycolipids.

No similar related structures were identified after a Basic Local Alignment Search Tool (BLAST) search. A collaborative team led by Rey-Ting Guo (Tianjin Institute of Industrial Biotechnology) hence solved the structure of MoeN5 using single-wavelength anomalous diffraction (SAD) with a 3.3 Å Se-Met data set at TLS 13B1 and TLS 15A1. After an iterative model refinement, several missing regions were lacking. It was difficult to propose a catalytic mechanism because of the poor resolution. To overcome this problem, they attempted to use four fusion protein tags, including thioredoxin from Escherichia coli, SUMO (small ubiquitin-like modifier) from Saccharomyces cerevisiae, Sac7d (chromosomal protein 7d) from hyperthermophiles Sulfol-
obus acidocaldarius and Sso7d (chromosomal protein 7d) from Sulfolobus solfataricus, to obtain improved crystal forms. Among
the four constructs, the MoeN5-Sso7d construct was crystallized in two space groups; both crystals yielded the improved data sets,
2.29 Å for the C222 form and 2.8 Å for the I222 form. The full structure of MoeN5 was eventually interpretable because the elec-
tron-density map was improved significantly with the two superi-
or data sets.

Two features were observed in the overall structure of MoeN5-
Sso7d – MoeN5 has an all α-helix structure, and two monomers in
an asymmetric unit form a dimer (Fig. 2(a)). For clarity, size-ex-
clusion chromatography with multi-angle laser scattering (SEC-
MALS) was performed to prove that the dimer form is a natural state unaffected by the addition of Sso7d. Use of the PDBFold
server revealed that the MoeN5 structure shared some similarities
with head-to-tail trans-prenyl synthases, including a (C10) ge-
nyl transferase from Thermotoga maritima, a (C10) geranylgera-
nyl diphosphate synthase from Sinapis alba, and a polyisoprenyl diphosphate synthase from Shi-
gella flexneri. These synthases contain two Asp-rich (DDXXD or
DXXXD) domains; abbreviation names FARM and SARM stand
for the first and second Asp-rich domain, respectively (Fig. 2(b)).
In the same manner, MoeN5 also had two domains. The collective
evidence indicated that the active site is expected to be near the two
Asp-rich domains.

The authors subsequently pro-
ceded to locate the two sub-
strate-binding sites, S1 (for GPP
6
, binding) and S2 (for FPG-tri-
saccharide 5

, binding), through
crystallization, ligand soaking
and structural analysis. Based on
the results, the GPP was located
exactly in the S1 position (Fig.
2(c)), but the FPG case for the
S2 position was unsuccessful. To
address this problem, the authors
executed an alternative strategy
to solve the complex structures of
MoeN5 with a series of glycolipid
models, which turned out to be
useful in mapping the binding
pocket. This structural information
indicated that FPG-trisaccharide
adopted a bent C-terminus con-
formation for further catalytic
reaction upon substrate binding
(Fig. 2(d)). The proposed catalytic
mechanism comprises four steps:
i: phosphoglycolipid 5 (blue) binds
to S2 with its bent (C6-11) side
chain and GPP 6 is shown (yellow);
ii: diphosphate is released
from GPP 6 and attack by C6,7/
C10,11 in 5 yields the cyclopentyl
carbocation 13; iii: 13 rearranges
to cyclohexyl carbocation 14; iv:
removal of a methyl proton in 14
by PPI yields product 7 (Fig. 2(e)).

In summary, the authors have
achieved the first structure of
MoeN5, an important prenyl-
transferase participating in MmA
biosynthesis. This structural infor-
mation can, importantly, provide a
future design of MmA analogues
as new lead compounds; the use
of small fusion tags is applicable
to other cases and is expected to
improve the quality of diffraction
data. (Reported by Chun-Hsiang
Huang)
The mechanism of DNA polymerase (pol) fidelity is of fundamental importance in chemistry and biology. Whereas the pols responsible for DNA replication are required to perform catalysis with great fidelity, those involved in DNA repair or mutagenic functions typically exhibit less fidelity. Although high-fidelity pols have been well studied, much less is known about how some pols achieve medium or low fidelity with functional importance. A structural basis for an atypical dG:dGTP mismatch (GG mismatch) incorporated catalyzed by the most error-prone DNA polymerase from Africa swine fever virus (Pol X) was reported. Pol X catalyzes the dG:dGTP mismatch on prebinding syn-dGTP in the absence of DNA; the syn-dGTP then form an anti:syn dG:dGTP Hoogsteen base pair with the template dG of the DNA. His115 was found to be the critical residue in stabilizing the syn-dGTP; mutation of His115 to an alanine residue abolished the syn-dGTP (only anti-dGTP observed), and resulted in a fidelity increase by 330 fold. This unprecedented finding of pre-binding MdNTP was in contrast to the paradigm in DNA polymerases, which states that DNA binding precedes that of MdNTP.

Wen-Jin Wu, Ming-Daw Tsai and their co-workers of Academia Sinica extended their investigation to human DNA polymerase λ (Pol λ). They examined how Pol λ achieves its moderate fidelity by determining 12 crystal structures of apo-Pol λ, MdNdTNP binary complexes, MnMdNdTNP binary complex (in which dNTP refers to dGTP dATP, dTTP and dGTP), dG:dATP mismatched ternary complex, apo-L431A mutant, binary complexes of L431A:MdCtTP, L431A:MdTTTP and L431A:dGTP, dG:dCTP matched ternary complex but with L431A mutant. X-ray diffraction data were collected at TLS 15A1, TLS 13B1, TLS 13C1 and SP 44XU. The authors also performed pre-steady-state kinetic analyses to determine the rate of dNTP incorporation, apparent Kd values of incoming dNTP to Pol-DNA binary complexes, catalytic specificity and fidelity. They showed that apo-Pol λ already exists in the closed conformation (Fig. 1(a)), unprecedentedly with a preformed MdNdTNP binding pocket (Figs. 1(c)-1(e)), and binds MdNdTNP readily in the active conformation in the absence of DNA (Fig. 1(d)). A large conformational change occurs upon the binding of the gapped DNA substrate containing a 5’-phosphate in the downstream primer (Fig. 1(f)).
This report features the work of Wen-Jin Wu, Ming-Daw Tsai and their co-workers published in J. Am. Chem. Soc. 138, 2389 (2016).

Their catalytic functions with varying fidelities (Fig. 2).
(Reported by Chun-Jung Chen)

Fig. 1: Comparison of apo-Pol λ with other relevant structures.  
(a) Structure of apo-Pol λ.  (b) Comparison between apo-Pol λ (green) and apo-Pol β (magenta).  (c) Overlay of apo-Pol λ (green) and Pol λ:MnMgdCTP binary complex (yellow, with dCTP shown in sticks).  (d) Expansion of (c) with the dCTP binding residues shown.  (e) Surface representation of apo-Pol λ with dCTP binding residues shown in (d) colored and labeled.  (f) Overlay of apo-Pol λ (green) and Pol λ:DNA binary complex (protein magenta, DNA grey) showing a large conformational change upon substrate-gapped DNA binding.  Asterisk symbols indicate the β-strand 8 location; a black star indicates the location of downstream primer 5'-phosphate.  [Reproduced from Ref. 2]

Fig. 2: X-family DNA polymerases utilize varied conformations to perform their specific functions and to regulate the DNA repair fidelity. For example, the dNTP-binding site is not formed in apo-Pol β; DNA-binding is required to form a dNTP-binding site. In contrast, the dNTP-binding site is ready in apo-Pol λ, and it can follow either the canonical DNA-binding first path or our newly discovered dNTP-binding first path.  Apo-Pol μ adopts a closed form with a DNA-binding site ready, which is absent in apo-Pol β or Pol λ.  For apo-Pol β, neither dNTP site nor DNA site is ready.  [Reproduced from Ref. 2]
Role of Dimeric Endonuclease G in DNA Degradation

The complex structures of CPS-6 with DNA provide a clue how endonuclease degrades DNA and exhibits endonuclease activity as a homodimer.

Endonuclease G (EndoG) is a highly conserved mitochondrial endonuclease in eukaryotes that plays a role in mitochondrial DNA biogenesis. EndoG is also a pro-apoptotic enzyme that promotes cell death in a caspase-independent path. During apoptosis, a fraction of EndoG is translocated from mitochondria into a nucleus to degrade chromosomal DNA. In mammals, EndoG is highly expressed in cardiomyocytes. EndoG-deficient mice have elevated levels of reactive oxygen species in cardiomyocytes leading to cardiac hypertrophy independent of blood pressure.

CPS-6 is a homologue of EndoG identified in C. elegans displaying a nuclease activity similar to that of EndoG. CPS-6 was identified to be involved in an elimination of the paternal mitochondrial DNA following fertilization during early embryogenesis. A decreased nuclease activity was observed for both CPS-6 and EndoG when the dimeric conformation was dissociated into monomer under oxidative conditions, but how EndoG/CPS-6 binds and cleaves DNA as a homodimer and why it loses its activity when it is oxidized and dissociates into monomers remain unknown.

Hanna S. Yuan of Academia Sinica and her co-workers constructed a CPS-6 double mutant, H148A/F122A, which retained its ability to bind DNA but had no activity in DNA cleavage. CPS-6 H148A/F122A mutant was crystallized in space group \( P2_1 \) with one dimer per asymmetric unit. The crystal structure of CPS-6 was determined by molecular replacement with CPS-6 H148A (PDB entry: 3S5B) as a search model. X-ray diffraction data were collected at TLS 15A1. They made a further attempt to co-crystallize CPS-6 H148A/F122A and a single-stranded DNA and determine the complex crystal at 2.3 Å resolution in different space group \( P2_12_12 \).

The overall structure of dimeric CPS-6 H148A/F122A bound to two single-stranded DNA in one asymmetric unit is illustrated in Fig. 1. Only five of eight nucleotides (5'-TTTTTTGT-3') were visible in the omitted density map; they were all assigned as thymidine (T1 to T5) based on the shape of each base (Fig. 2). These five nucleotides could represent a digested 5-nt DNA molecule or an intact 8-nt DNA that has disordered ends (Fig. 2).

The two DNA molecules are located remotely from each other and face opposite directions. Inspection of the CPS-6-DNA complex structure reveals that CPS-6 binds DNA mainly at the phosphate backbones via hydrogen-bonded interactions. The DNA bases point away from the protein surface and in no way interact with CPS-6, thereby avoiding sequence-specific interactions.

Based on the structural and functional information obtained, a DNA hydrolysis mechanism was proposed. The active site is located in the ββα-metal motif that is bound with a Mg\(^{2+}\) ion. His148 acts as a general base to activate a water molecule, which in turn makes an in-line attack...
on the scissile phosphate. The Mg-bound water functions as a general acid to provide a proton to the cleaved DNA product. The bound DNA substrate is thereby hydrolyzed in a sequence-independent manner to produce cleaved products with a 5'-end phosphate and a 3'-end OH group.

Why EndoG has a diminished nuclease activity when it is dissociated into monomers is intriguing, as each protomer can bind and degrade DNA separately. To answer this question, they further constructed two obligatory monomeric CPS-6 mutants (P207E and K131D/F132N), which degrade DNA with a diminished activity because of a DNA-binding affinity poorer than that of wild-type CPS-6. Unexpectedly, the P207E mutant exhibited predominantly a 3’-to-5’ exonuclease activity, indicating a possible activity change from endonuclease to exonuclease. The dimer conformation of CPS-6 is thus essential to maintain its optimal DNA-binding and endonuclease activity. Compared to other non-specific endonucleases, which are typically monomeric enzymes, EndoG is a unique dimeric endonuclease, of which the activity can be modulated with oxidation to induce a dimer-to-monomer conformational change.

In summary, these results provide a molecular basis to explain how EndoG degrades DNA substrates without a sequence preference and why EndoG exhibits optimal endonuclease activity as a homodimer. The authors suggest that stabilizing the dimeric conformation of EndoG might provide a way to promote its endonuclease activity and to combat diseases induced by oxidative stress. (Reported by Chun-Jung Chen)

This report features the work of Hanna S. Yuan and her co-workers published in Nucleic Acid Res. **44**, 10480 (2016).

**TLS 15A1 Biopharmaceuticals Protein Crystallography**
- Protein Crystallography
- Biological Macromolecules, Protein Structures, Life Sciences

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>

**Molecular Averaging Is Powerful for Crystal Structures**

A new method, molecular averaging in real space, is developed to evaluate effectively the phasing power and to enhance the success of determining new protein structures.
In 2015, a research team of Chun-Jung Chen from the Life Science Group, NSRRC, applied this method through *ab initio* molecular averaging to crystal data of a novel grouper nervous necrosis virus-like particle (GNNV-LP) and obtained the first crystal structure. In 2016, they reported the detailed description of the method applied to the GNNV-LP structure, a quantitative discussion of the feasibility of the method on introducing a novel conventional index. This challenging trial and its success of *ab initio* phasing push the boundary of structure determination of biological macromolecules, which will benefit the community of structural biology.

The GNNV-LP data and crystal data of its protrusion domain were recorded at Protein Crystallography beamlines, including TLS 15A1, SP 12B2 and SP 44XU. The GNNV-LP data were collected at resolution 3.6 Å, and the protrusion domain at 1.2 Å. The first *ab initio* method is to use non-crystallographic symmetry averaging (NCSA). From the GNNV-LP data, it assumed 30 redundant copies of a capsid protein molecule that formed viral icosahedral particles in the crystallographic asymmetric unit. This icosahedral orientation was derived from the self-rotation function of the amplitude data. Capsid protein particle positions and the solvent region in the unit cell were guessed from the form of related virus of the same genera: alpha-nodavirus. Beginning from a naive shell model, the electron density of the molecular copies was averaged; the density at the solvent region was flattened. In reciprocal space, amplitude data were added little by little. Between the density maps of real space and reciprocal space was linked with Fourier transform (FT) or inverse-FT. The basic cycle was calculated iteratively from low to high resolution. Figure 1 shows a schematic diagram of the basic cycles. The final density map for a novel GNNV-LP was obtained (Fig. 2).

In the obtained map of electron density, there was still a region of unclear density that was for the protrusion domain located on the surface of the viral particle. The protrusion domain has a flexible structure in its viral states. They crystallized this protrusion domain solely and obtained high-resolution crystal data of four kinds of the space group and cell sizes. Secondly, they applied the molecular averaging procedure to the equivalent molecular densities among the three crystal data as shown in Fig. 3. The procedure was called cross-crystal averaging (CCA).

As an initial density, they used the envelope obtained from the unclear protrusion region of the GNNV-LP map. After an iterative calculation of averaging as the first NCSA case, the phases were much improved to produce high-resolution structures for the protrusion domain and the entire GNNV-LP structure (Fig. 3).

From this experience, the authors introduced a new index named a free fraction ($ff$) to show the phasing power of the molecular averaging. When the copy number of NCS is $n$, $ff$ is expressed as

$$
ff = 1 - \frac{S}{nk + n_1S_1P_1 + n_2S_2P_2 + \ldots + n_mS_mP_m}
$$

### Basic cycle

![Basic cycle diagram](image)

**Fig. 1**: Schematic diagram of a basic iteration cycle of the averaging method. [Reproduced from Ref. 2]
in which $S$ is the fraction of the solvent region. As $ff$ shows the fraction of the unconstrained region in the total cell, $(n-1)$ copies and the solvent region are constrained in the averaging. Taking the benefit of the simplicity of this formula, $ff$ is readily expanded to the multi-crystal situation of CCA.

If the equation is generalized to $m$ pieces of crystal, crystal $k$ has $n_k$ NCS, $S_k$ and $P_k$ are fractions of the solvent and protein regions, respectively; $ff$ becomes

$$ff = 1 - \sum_{k=1}^{m} n_k + n_1 \frac{S_1}{P_1} + n_2 \frac{S_2}{P_2} + ... + n_m \frac{S_m}{P_m}$$

The first successful NCSA case was $ff = 0.024$, the second CCA case was $ff = 0.038$. Adding test calculations with varying $ff$, a value less than about 0.1 was found to have a great phasing power to show the interpretable clear density map.

The authors discussed the meaning of $ff$. Given amplitudes and their phases, one can obtain an entire density map in the crystal cell. With only amplitude data in a typical case, the number of the data is decreased by half. From a point of view of information content, to obtain the entire density in the cell more than half part of the density should be known and $ff$ should be less than 0.5. The region $0.5 < ff < 1$ is considered to be an over-fitting condition in which the number of data is less than the number of parameters to be determined. The authors noted that $ff$ is another expression of a data parameter ratio. $ff$ is more useful than a data parameter ratio because the values of $ff$ can be used without considering the resolution limit or the grid spacing of the density map. They obtained a criterion $ff < 0.1$; it corresponds to a data parameter ratio > 5, which is a reasonable value with another ab initio method for a successful condition such as sub-atomic resolution.

Their new index $ff$ can show a quantitative phasing power. By its simple calculation, it indicates not only the feasibility of ab initio phasing trials but also how much density modification work is required to remove the initial model biases. This method of molecular averaging in real space can effectively evaluate the phasing power to enhance the success of determining new structures.


### TLS 15A1 Biopharmaceuticals Protein Crystallography
### SP 12B2 BM – Protein X-ray Crystallography
### SP 44XU Macromolecular Assemblies
- Protein Crystallography
- Biological Macromolecules, Protein Structures, Life Sciences

<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>