

Appendix A.***List of E. coli Methionyl-tRNA Synthetase Mutants Identified Through Library Screening to Show Activity Toward Azidonorleucine In Vivo***

A list of all methionyl-tRNA synthetase (MetRS) mutants identified from clones isolated through various screens of the LYH.1.0 saturation mutagenesis library for activity toward azidonorleucine (AnI) is presented here. Only variants carrying mutations at the three mutagenised sites (L13, Y260, and H301) are listed. The populations each mutant has appeared in are also indicated. These populations are explained below. From 150 active clones, a total of 41 distinct MetRS mutants were identified. A complete list of these mutants are provided in the following page.

Table A.1

List of various populations that were obtained by screening the LYH.1.0 library

Population	Information
1.2	See LYH.1.2 in Figure 2.6.
2.2	See LYH.2.2 in Figure 2.6.
2.3	Revealed by screening 2.2 at 0.1 mM AnI
3.2	See LYH.3.2 in Figure 2.6.
3.3	Revealed by screening 3.2 at 0.3 mM AnI
4.3	Revealed by screening 3.2 at 1.0 mM AnI
5.3c	The mixture of LYH.1.1, LYH.2.1 and LYH.3.1 (See Figure 2.6.) screened two rounds with 1.0 mM AnI and Met
5.3d	See LYH.5.3d in Figure 2.7.
5.4b	The mixture of LYH.1.1, LYH.2.1 and LYH.3.1 (See Figure 2.6.) screened three rounds with 0.3 mM AnI and Met
6.2	See LYH.6.2 in Figure 3.?

Table A.2*List of MetRS mutants identified through library screening to activate AnI in vivo*

L13	Y260	H301	Occurrence	Populations clones are isolated from:							
A	C	L	1	1.2							
A	Q	L	2	1.2	6.2						
A	V	L	3	6.2							
C	L	L	2	3.2							
C	V	L	1	1.2							
D	L	I	1	1.2							
F	G	L	1	6.2							
F	V	L	1	3.3							
G	C	L	2	1.2							
G	I	L	3	1.2							
G	L	L	2	3.2	5.3d						
G	L	V	1	6.2							
G	M	L	4	1.2							
G	N	L	3	3.2	3.3	4.3					
G	T	L	4	1.2	3.3	6.2					
G	V	L	4	6.2							
H	C	L	1	1.2							
H	T	M	1	1.2							
L	C	L	1	6.2							
M	A	L	1	1.2							
M	S	L	1	6.2							
N	L	L	62	2.2	2.3	3.2	3.3	4.3	5.3c	5.3d	5.4b
N	V	L	1	1.2							
P	C	L	1	3.2							

L13	Y260	H301	Occurrence	Populations clones are isolated from:		
P	I	L	1	3.2		
P	I	M	1	1.2		
P	L	I	9	2.2	5.3c	5.4b
P	L	L	2	5.3c		
P	L	V	1	3.2		
P	M	L	1	2.2		
P	N	L	5	1.2	3.3	4.3
P	N	V	2	6.2		
P	T	V	1	1.2		
P	V	L	1	5.3d		
S	C	M	1	6.2		
S	L	L	12	2.2	2.3	3.2 5.4b
S	L	V	1	4.3		
S	M	L	4	3.2	3.3	
S	N	L	1	6.2		
S	S	L	2	6.2		
S	T	L	1	1.2		
Total			150			

Appendix B.

MALDI-MS Data Demonstrating the Incorporation of Azidonorleucine into Methionine Positions on Tryptic Fragments of Murine Dihydrofolate Reductase

Murine dihydrofolate reductase (DHFR) expression profiles of cells containing different MetRS mutants, and a brief description of the MALDI-MS data on tryptic digests of DHFR were previously presented in Figure 2.9. Here, this data is presented in greater detail.

DHFR was expressed in M15MA[pREP4/pAJL-61] cells encoding various methionyl-tRNA synthetase (MetRS) mutants. These included mutants isolated from screens carried out at 0.3 mM and 1.0 mM azidonorleucine (AnI), as well as the L13G mutant and the wild-type MetRS. Expression was done in five different conditions: 1) in unsupplemented M9+19aa medium (labeled as “19aa” or “0 mM AnI”), 2) in M9+19aa media supplemented with 0.1 mM, 3) 0.3 mM, or 4) 1.0 mM AnI, or 5) with 40 mg/L methionine. The cell lysates standardized for cell densities show the differing amounts of DHFR expressed at each of these conditions. (Figure B.1) DHFR was purified using Ni-NTA affinity chromatography, and the purified protein was analyzed by SDS-PAGE. (Figure B.2)

Purified DHFR was subjected to digestion by trypsin, and the resulting fragments were examined by MALDI-MS. For each sample, the region of the mass spectrum containing the tryptic fragments IMQEFESDTFFPEIDLK (2437.16 Da) and QNLVIMGR (930.52 Da), both of which contain a single methionine position. Replacement of Met by AnI results in a 23.05 Da mass shift, which is shown on each panel. The expected locations of the for Met- and AnI-containing peptides, as well as other major peaks are labeled by their molecular weights. (Figure B.3)

It is apparent in the SDS-PAGE analysis that the wild-type MetRS (labeled “wt”) supports protein synthesis even in the absence of methionine. This expression is not affected by the presence or absence of 1 mM AnI in the media.

Peptides from these samples indicate the incorporation of another amino acid 2 Da smaller than methionine, which matches the mass of glutamic acid. Similarly, in the absence of methionine or azidonorleucine, the MetRS-L13G mutant incorporates another amino acid 3 Da smaller than methionine, which corresponds to either lysine or glutamine.

Figure B.1

Expression levels of DHFR at different media compositions.

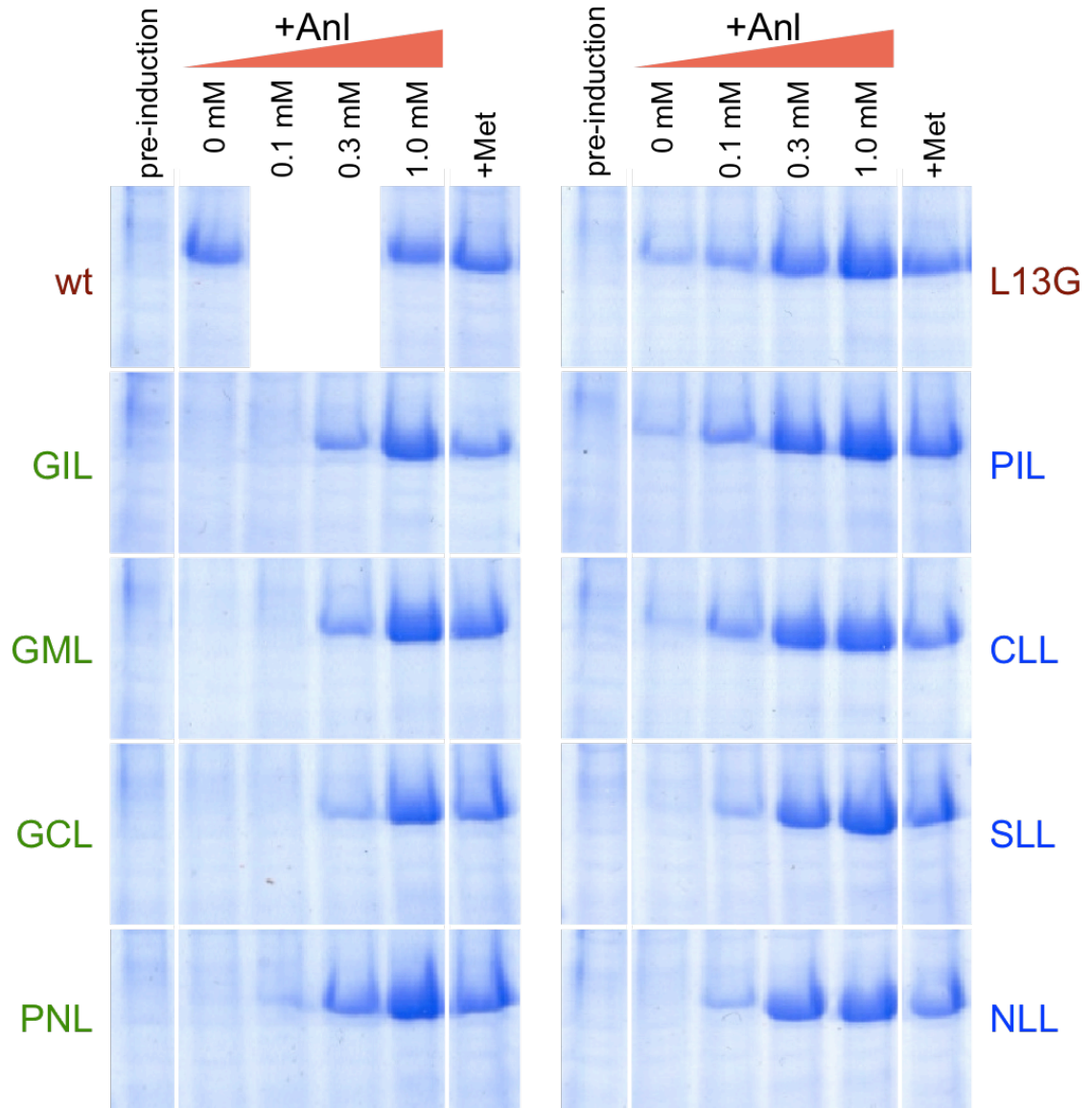


Figure B.2

SDS-PAGE analysis of DHFR induced at different media compositions and purified.

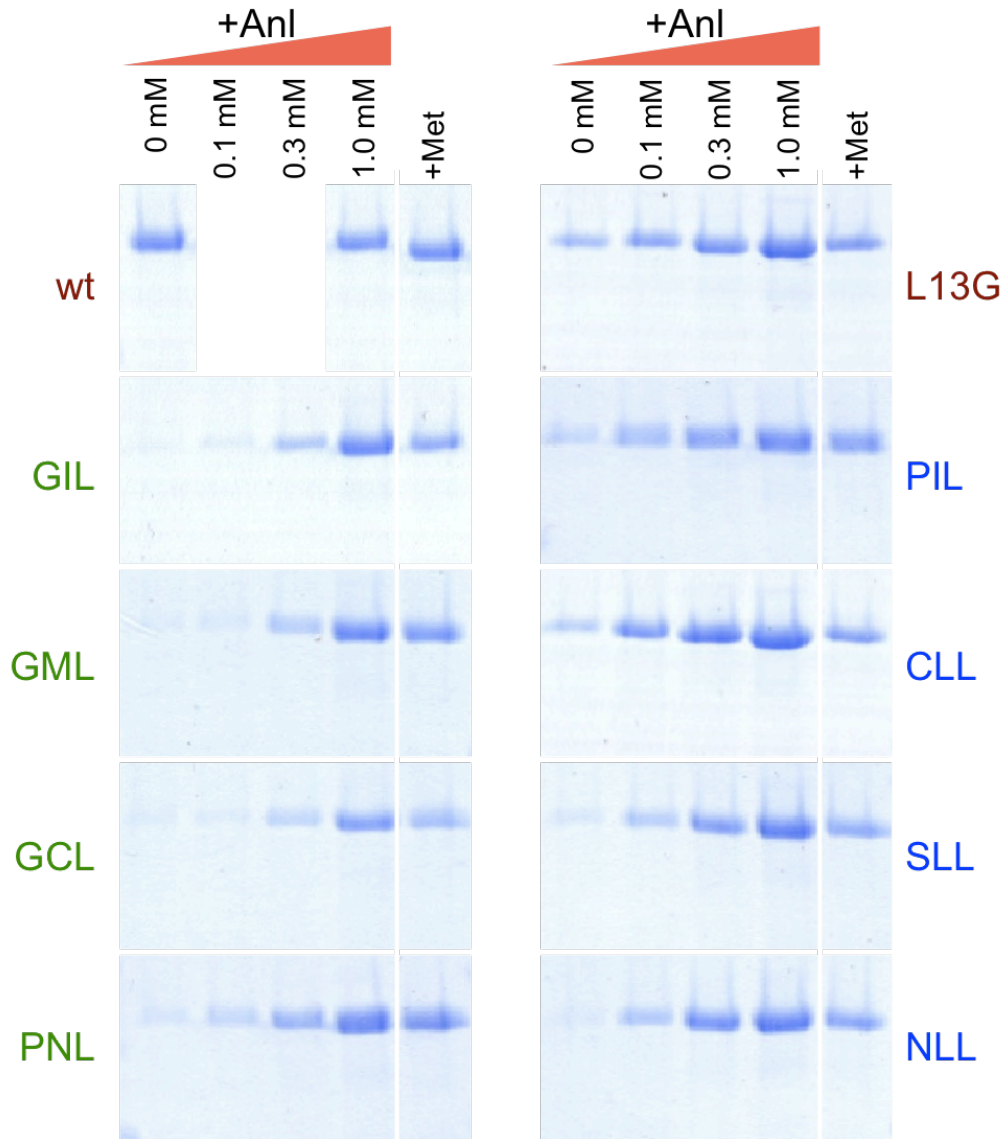


Figure B.3

Mass spectrometric analysis of tryptic fragments of DHFR purified from cells bearing various MetRS variants.

Purified DHFR (Figure B.2) was subjected to digestion by trypsin and the digested peptides were analyzed with MALDI-MS. With the exception of the first panel, each of the following panels show data pertaining to each MetRS variant tested. The initial panel contains data from wild-type MetRS and accompanying controls, showing the locations of Met- and AnI-containing peptides. In addition, data from a digest where DHFR was omitted was also presented (labeled as “Trypsin only”) in order to gauge the contribution of the self-digestion of trypsin to the signal observed. Replacement of Met by AnI increases the mass of each peptide by 23.05 Da, and this shift is presented on each panel by a black arrow, for each peptide analyzed. Red arrows show the differences in mass from the Met-containing peptide to other major peaks that appear nearby, in the absence of Met or AnI.

Figure B.3 (continued)

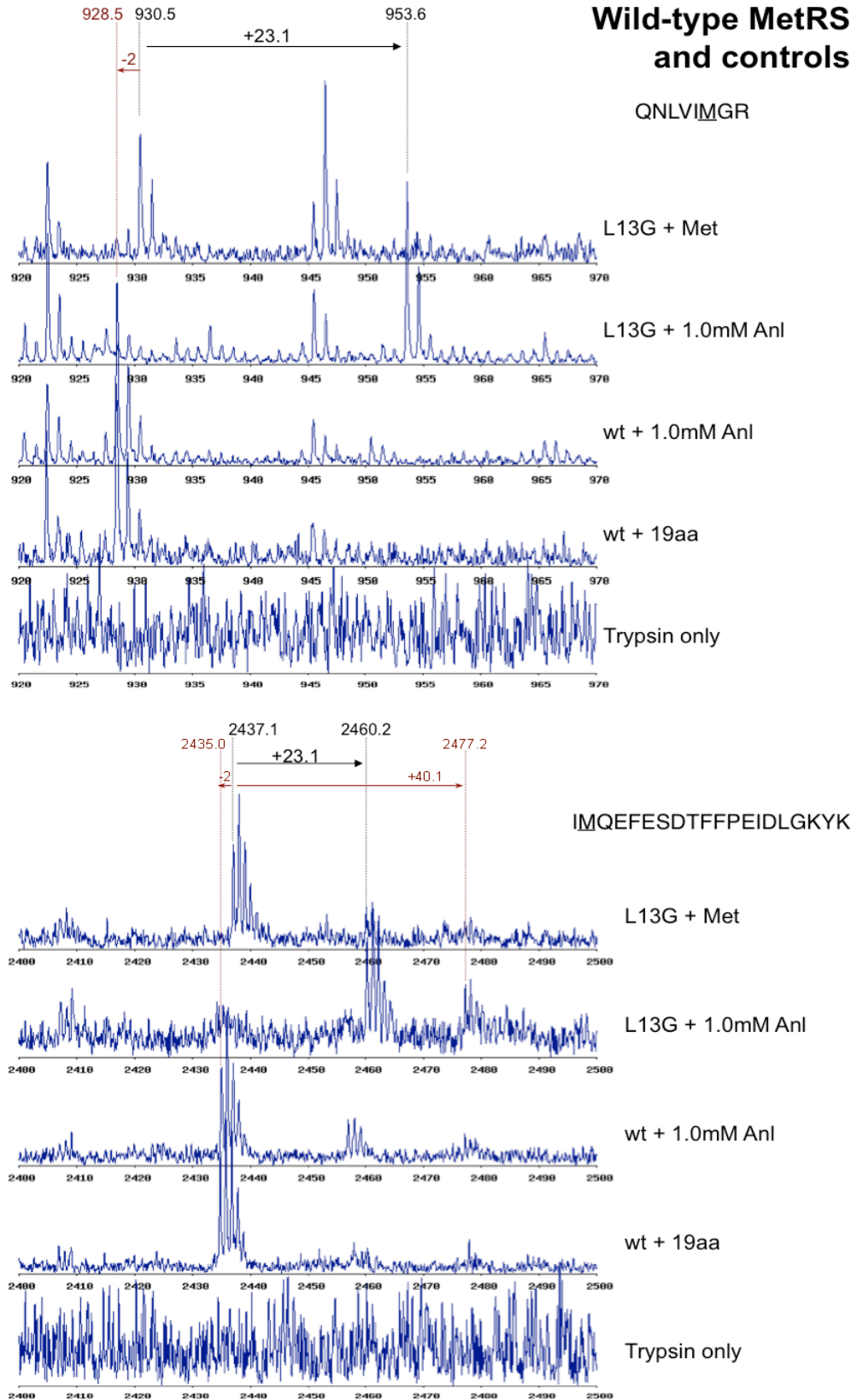


Figure B.3 (continued)

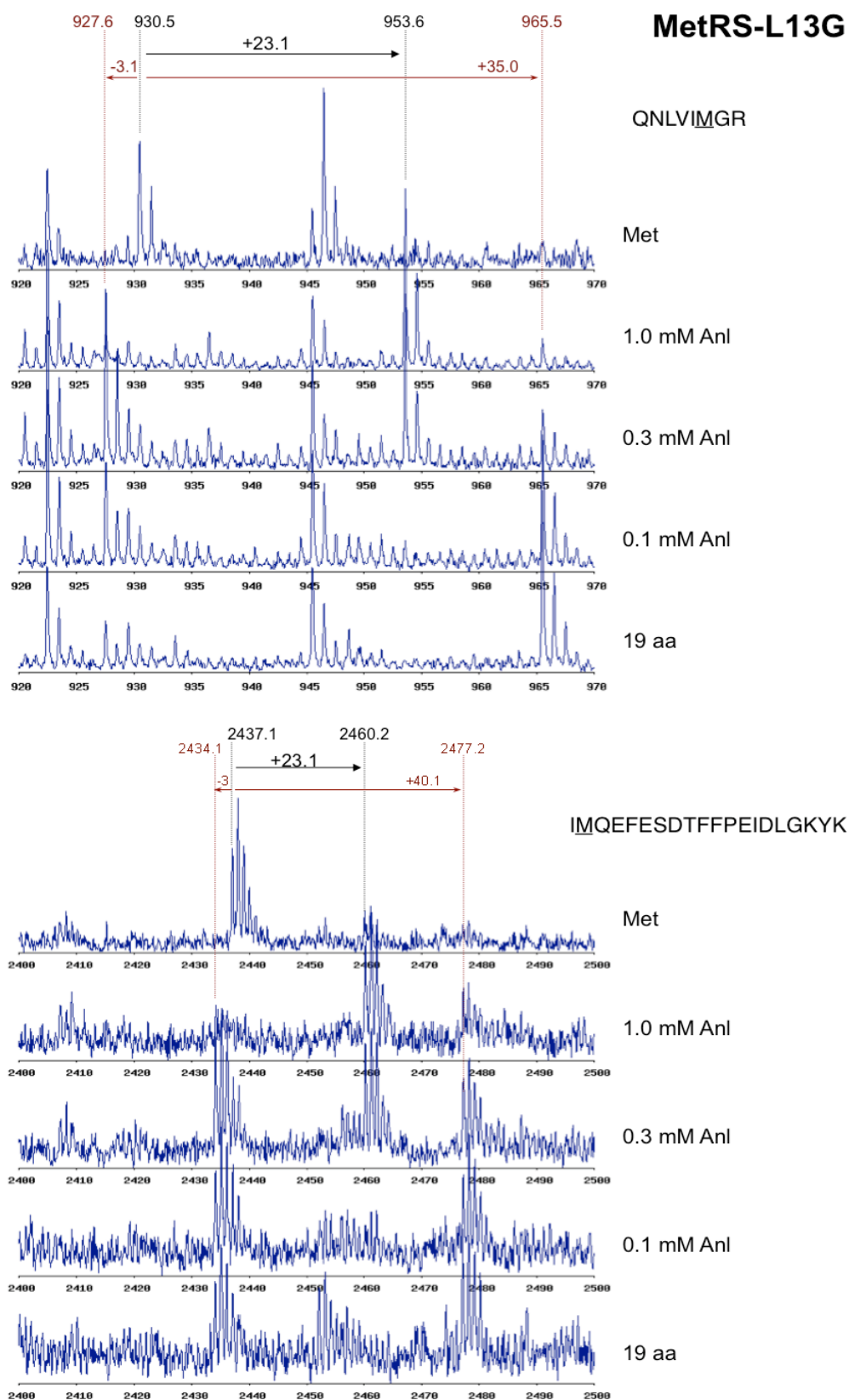


Figure B.3 (continued)

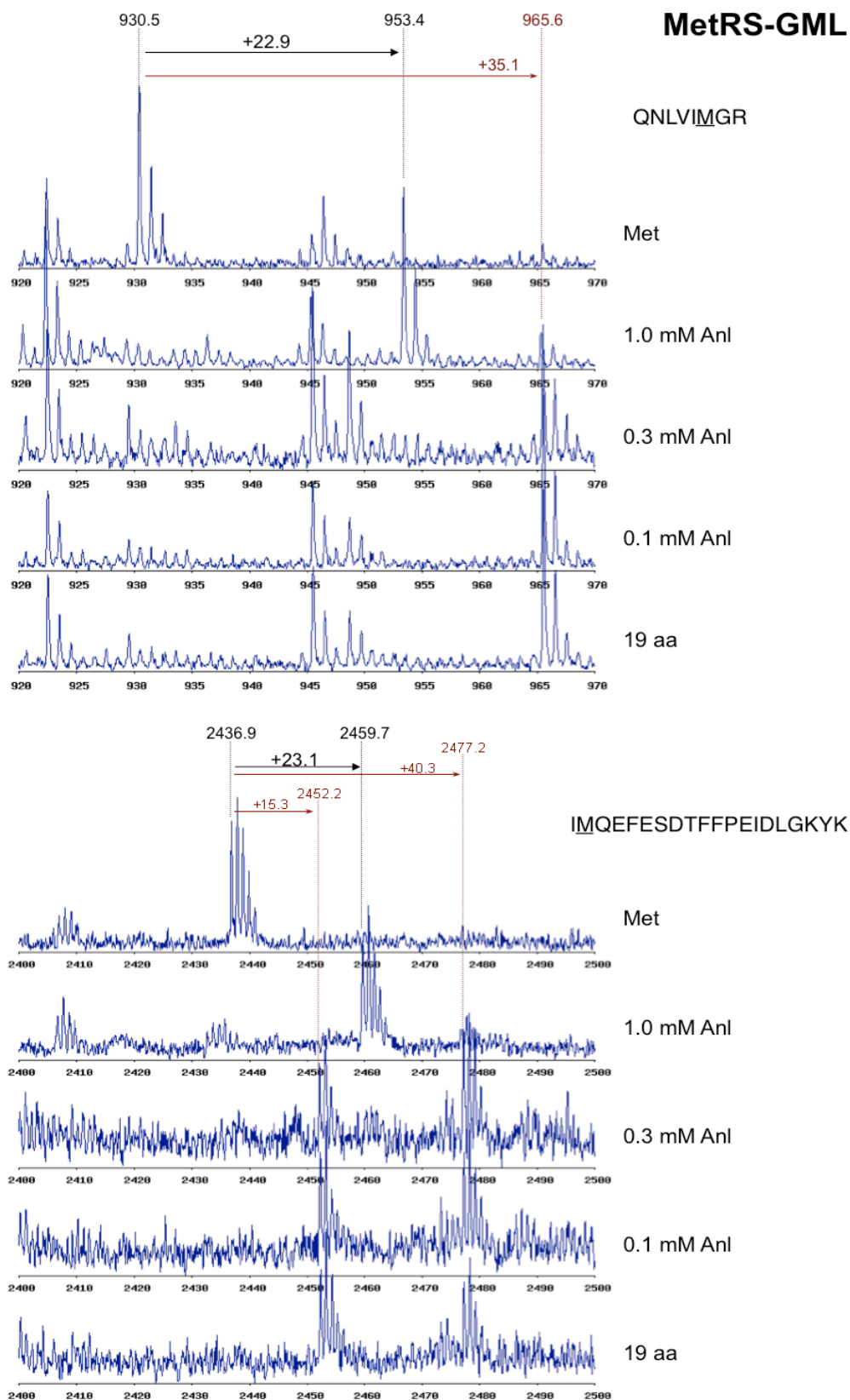


Figure B.3 (continued)

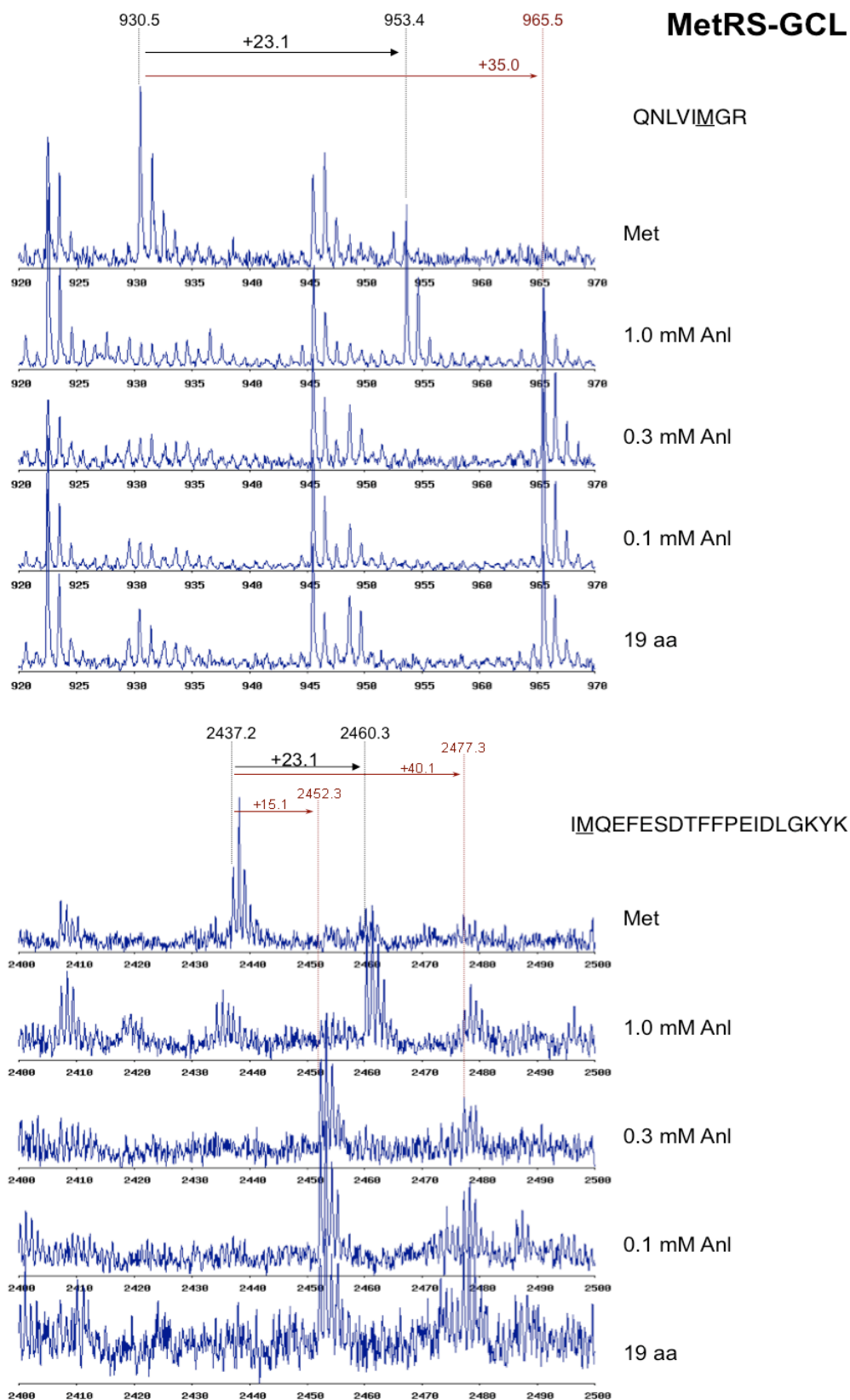


Figure B.3 (continued)

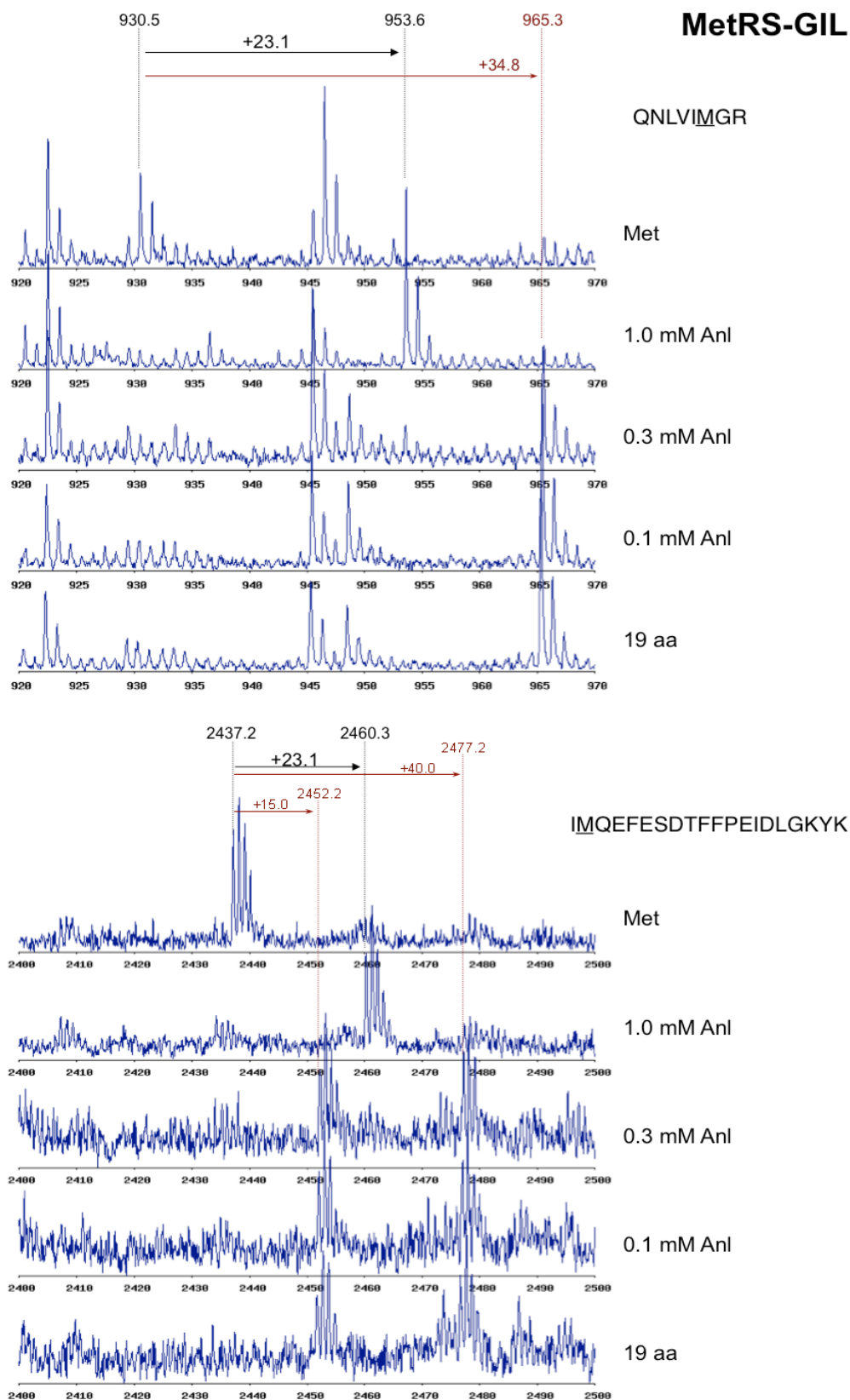


Figure B.3 (continued)

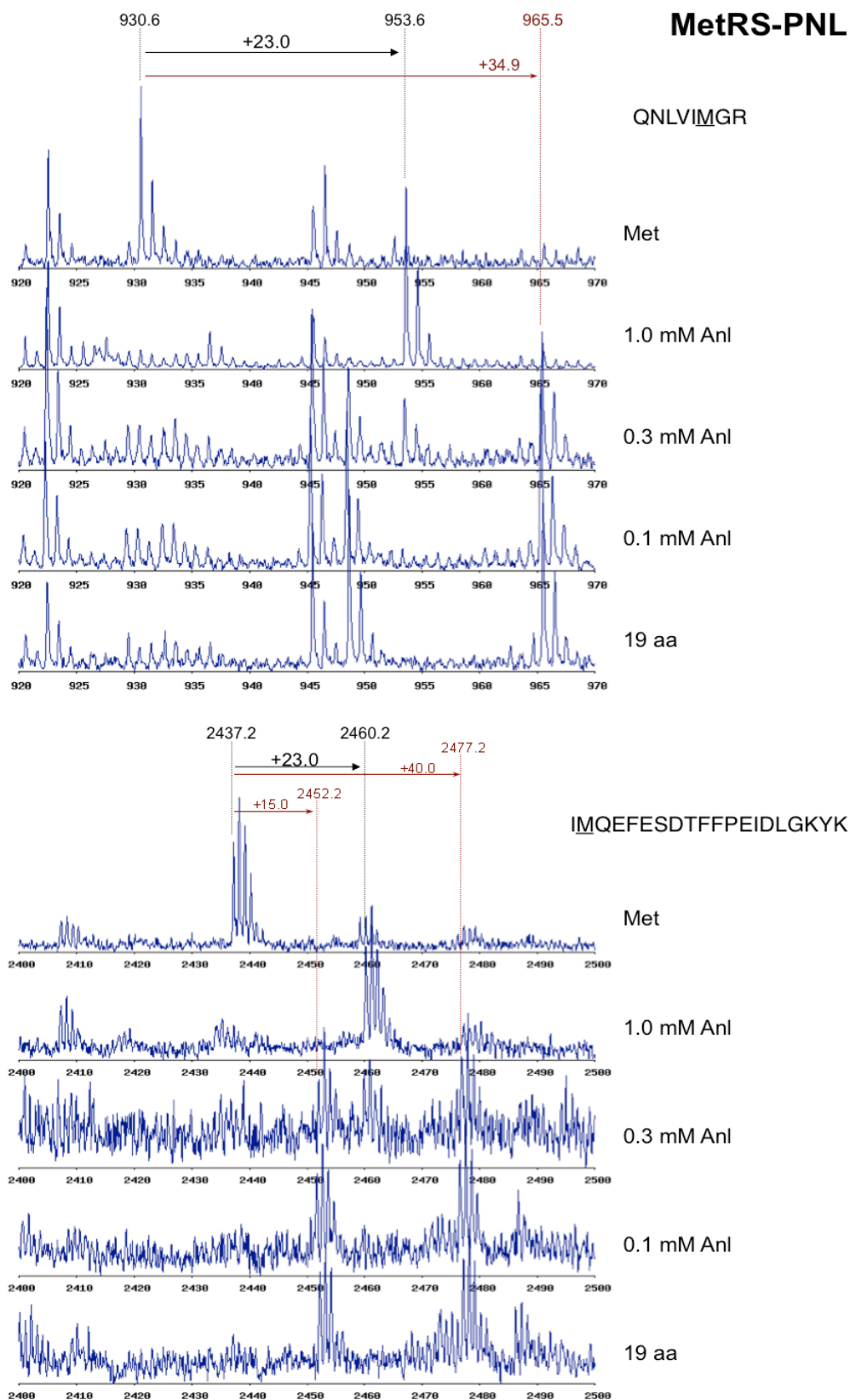


Figure B.3 (continued)

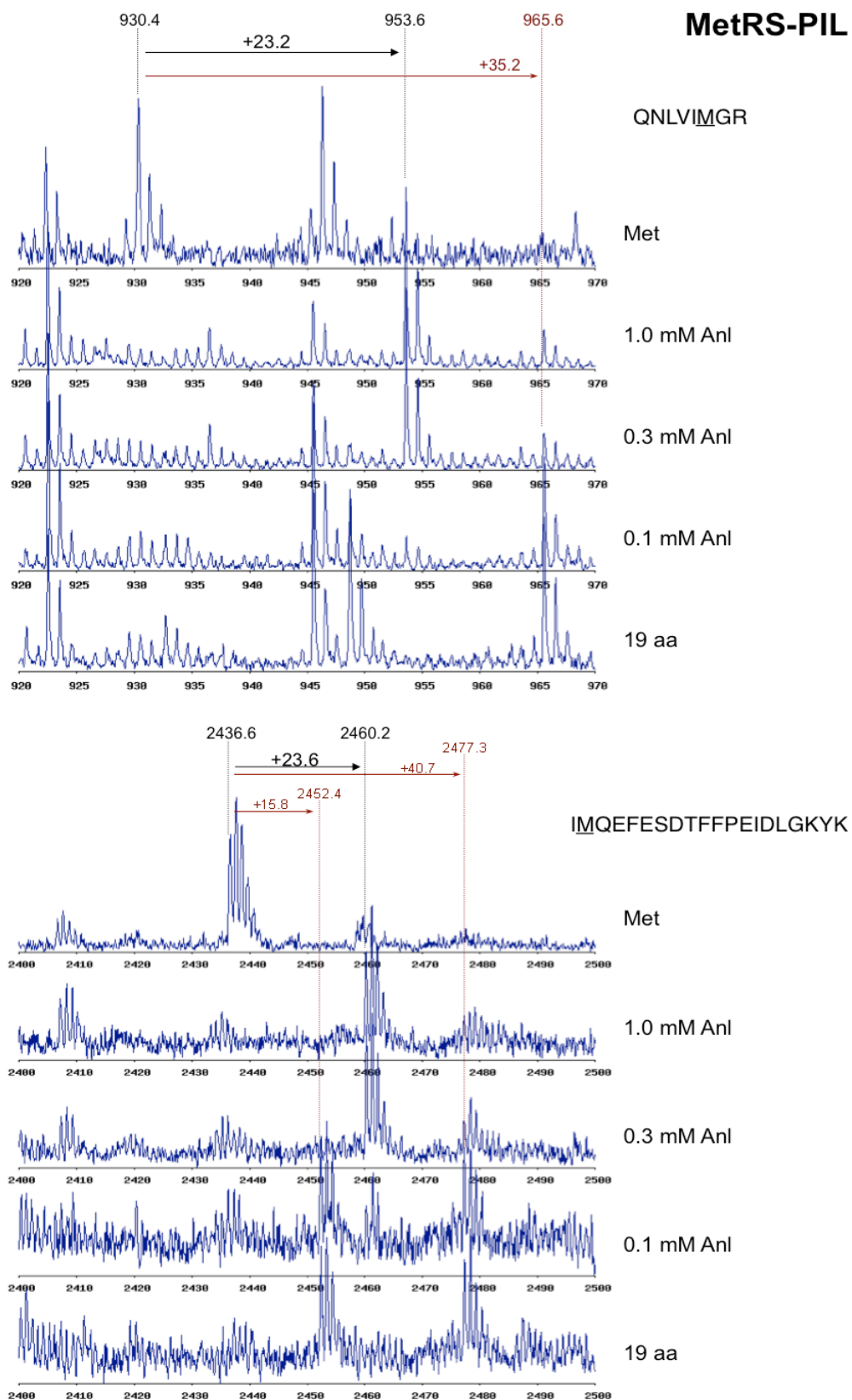


Figure B.3 (continued)

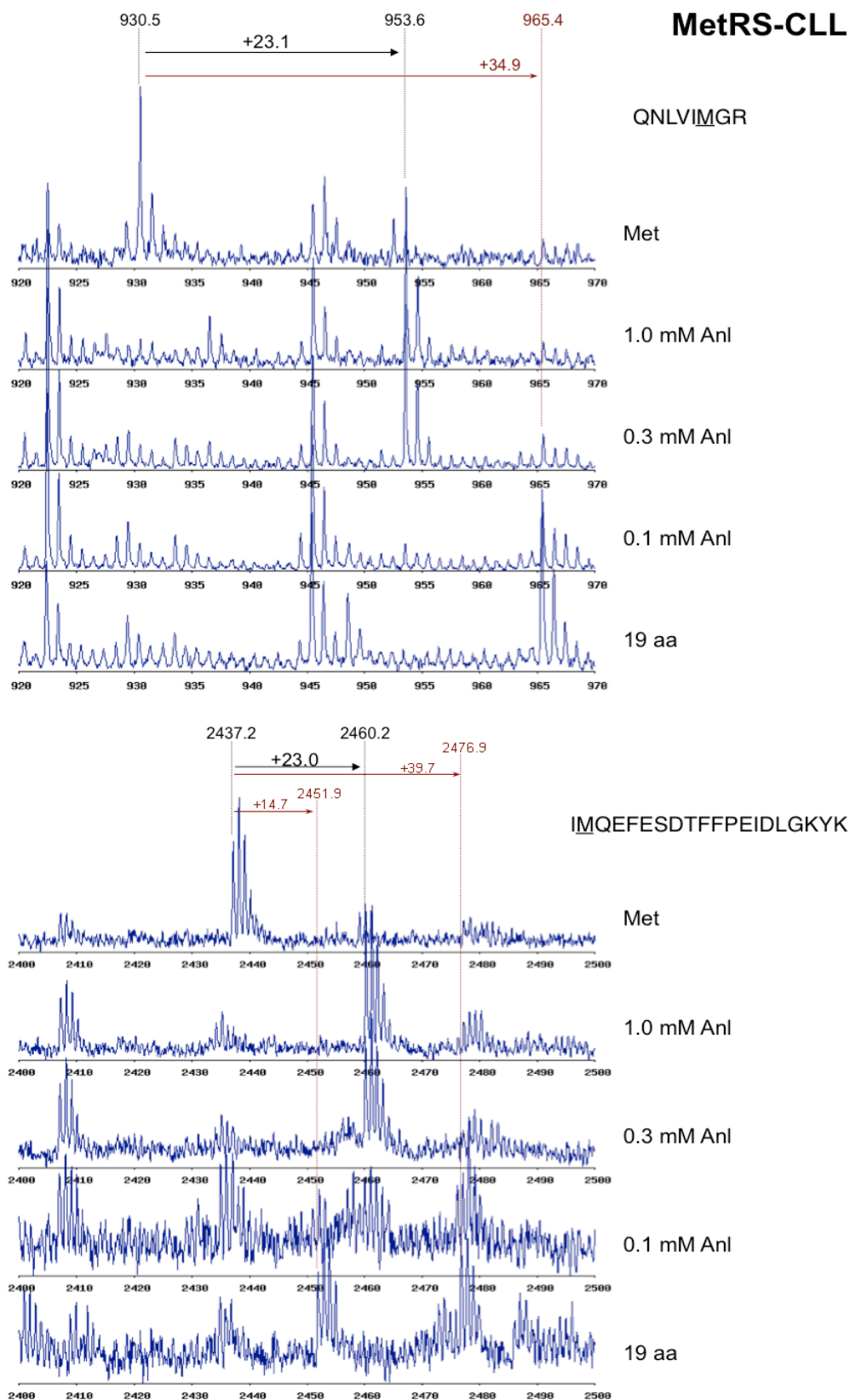


Figure B.3 (continued)

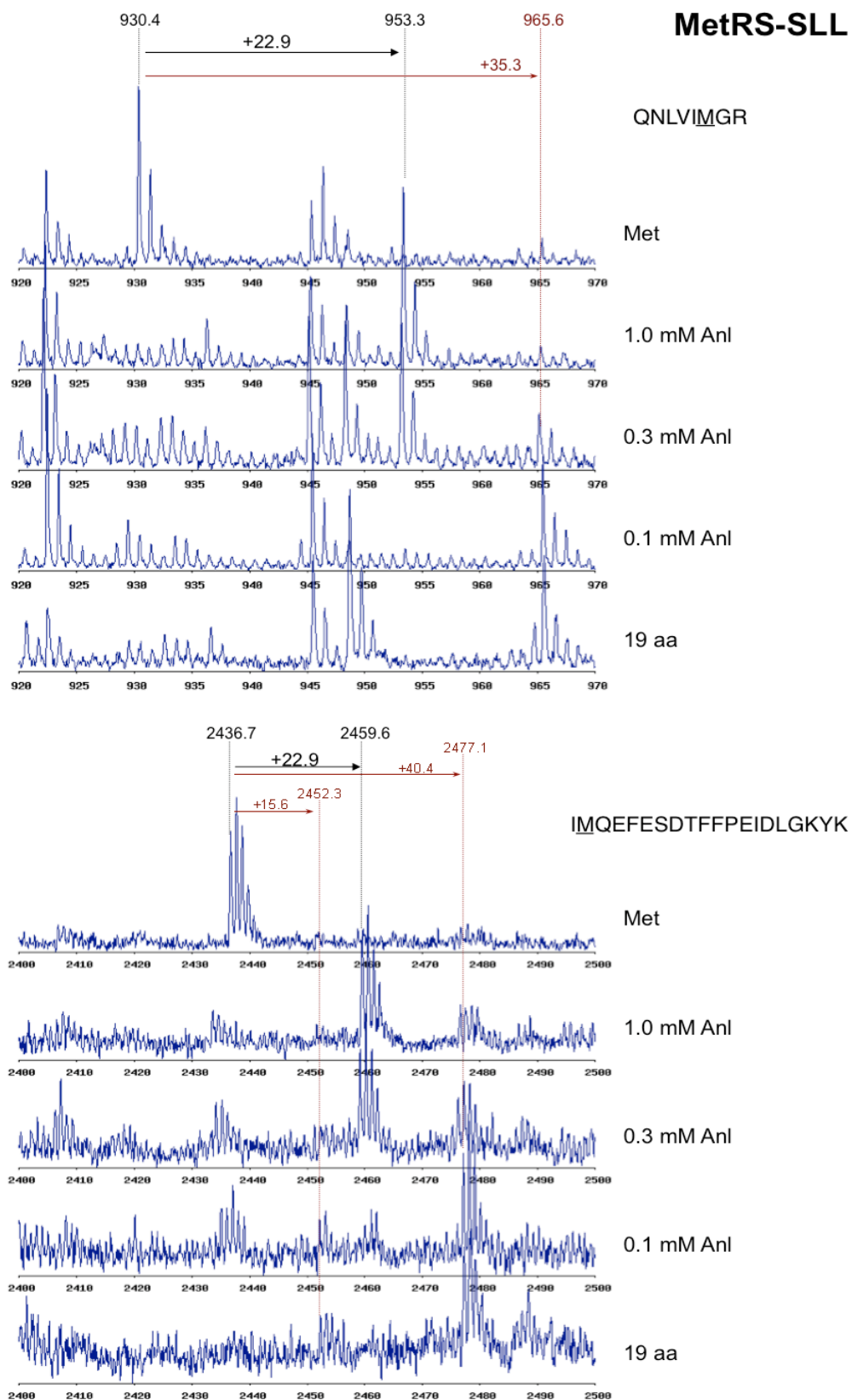
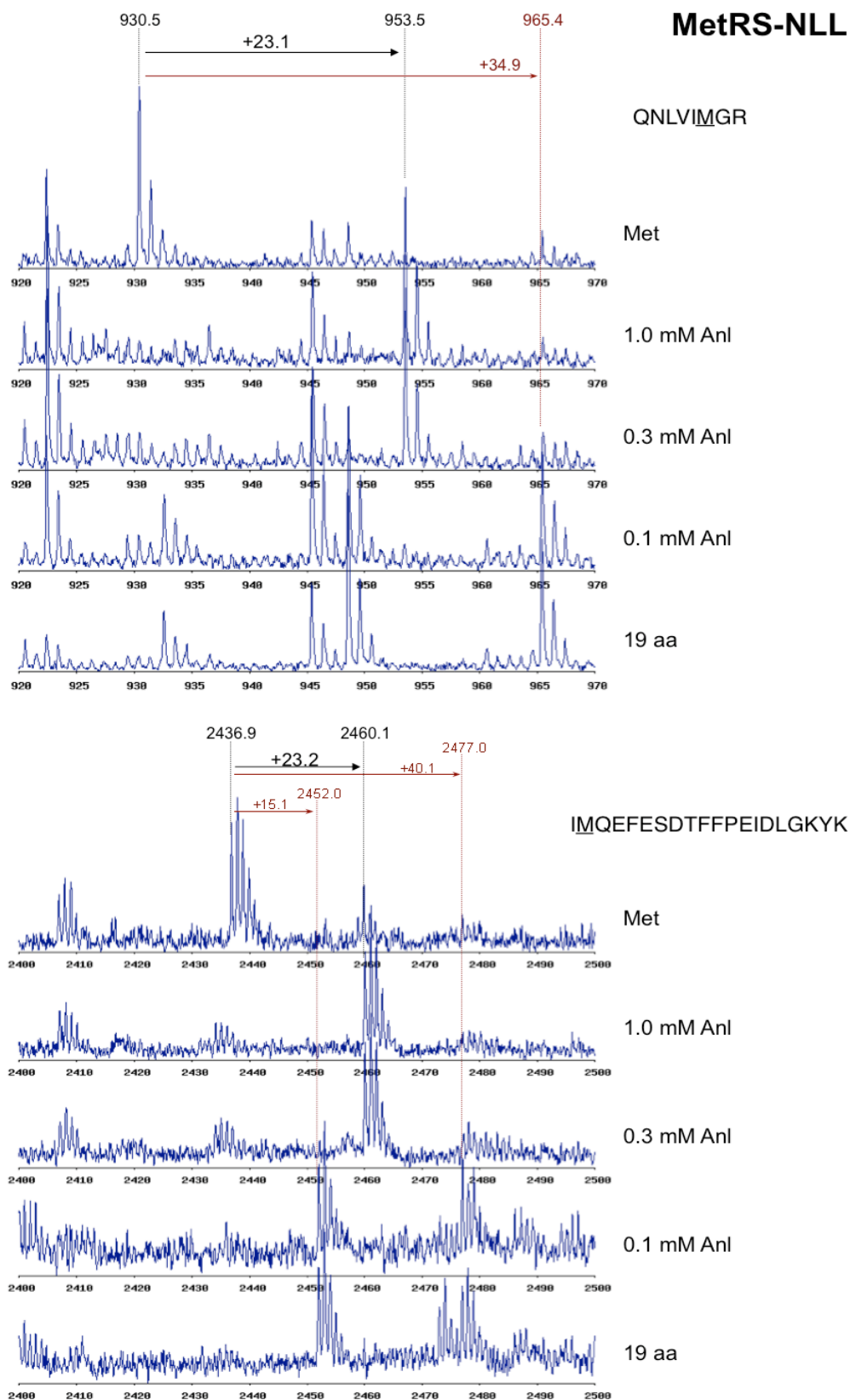


Figure B.3 (continued)



Appendix C.

Summary of Computational Design and Mutation Studies Performed on Aminoacyl-tRNA Synthetases Other than the Methionyl-tRNA Synthetase

Studies discussed in this dissertation all relate to the *E. coli* methionyl-tRNA synthetase (MetRS) and the identification of variants of this enzyme that can activate azidonorleucine. This section presents a summary of work done on aminoacyl-tRNA synthetases, prior to the MetRS study. A description of each system and the results of each study are reported briefly.

Design of the *E. coli* isoleucyl-tRNA synthetase for the incorporation of hexafluorovaline into proteins *in vivo*

- *Hexafluorovaline (Hfv)*: Hfv is a highly fluorinated amino acid, which can be used in the modulation of hydrophobicity of proteins, and stabilization of protein cores.
- *Isoleucyl-tRNA synthetase (IleRS)*: Because the crystal structure of the *E. coli* IleRS is not available, the *Thermus thermophilus* IleRS structure at 2.8 Å resolution was used [1]. The *T. thermophilus* enzyme shares the same binding site residues with the *E. coli* IleRS. (Figure C.1)
- *Design*: The Clash-Opportunity Progressive (COP) design procedure [2] was carried out on three positions to generate a series of one-, two-, and three-fold mutants. Three mutants which show improved VdW interactions with the ligand were chosen for experimental characterization (A, AF and F; Table C.1.a). The F518 was chosen based on the analysis of the IleRS structure.
- *Results*: Hfv incorporation was not observed *in vivo* for any of the mutants, and its activation was not detected *in vitro* for A, F and AF. Isoleucine was activated *in vitro* by A and AF, but at lower rates than the activation by the wild-type IleRS.

Design of the *E. coli* valyl-tRNA synthetase for the incorporation of hexafluorovaline into proteins *in vivo*

- *Valyl-tRNA synthetase (ValRS)*: Because the crystal structure of the *E. coli* ValRS is not available, the *Thermus thermophilus* ValRS structure at 2.9 Å resolution was used [3]. As it is for IleRS, the *T. thermophilus* ValRS also shows extensive similarities to the *E. coli* enzyme, sharing the same binding site residues.
- *Design*: COP was used on positions P41 and I491 and eight double mutants were chosen for experimental characterization. In addition to these, eight additional single ValRS mutants were chosen based on their differential interaction energies between Val and Hfv. (Table C.1.b) Expression profiles for clones carrying these mutants and the MALDI-MS for the expressed proteins were obtained.
- *Results*: None of the 16 ValRS mutants constructed showed activity toward Hfv, although the incorporation of isoleucine or leucine was commonly observed. This indicates the importance of especially the residues P42 and I491 in discriminating between Ile and Val.

Designing the *M. jannaschii* tyrosyl-tRNA synthetase binding site for tyrosine analogs

- *Ligands*: Following the selection experiments of the laboratory of Peter Schultz on the *in vivo* incorporation of O-methyl-tyrosine through a mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj-TyrRS*) [4], Zhang et al. showed that it was possible to correctly identify the mutations in the immediate vicinity of the ligand, at positions Y32 and D158, through computational means using the COP procedure [2]. We continued these efforts to verify the success of the COP strategy using other ligands activated by *Mj-TyrRS* mutants identified by the Schultz laboratory: *p*-acetyl-phenylalanine (Acf) [5], *p*-azido-phenylalanine (Azf) [6], *p*-amino-phenylalanine (Amf) [7], 2-naphthylalanine (Nal) [8]. These ligands were selected for the low degrees of freedom they have in their side chain

coordinates. (Figure C.2)

- *Mj-TyrRS*: A 1.95 Å resolution crystal structure of this enzyme complexed with its native ligand, tyrosine, was used in this study [9]. The structure of the mutant *Mj-TyrRS* specific for *O*-methyl-tyrosine [10] was not available at the time of this study.
- *Design*: The *Mj-TyrRS* binding site was designed over the two positions (Y32 and D158) identified as clashing with the selected ligands. The positions Y32 and D158 are in the 5 positions randomized in the selection experiments. (Figure C.2)
- *Results*: Almost all *Mj-TyrRS* mutants identified by the Schultz laboratory contain four or more mutations. For each mutant-analog pair we introduced the full set of mutations into *Mj-TyrRS*, and determined the binding energy of each mutant with the analog, and with tyrosine as a competitor. We verified that the full mutants show good discrimination for their respective analogs and against tyrosine.

For Acf, the two double mutants containing the experimental mutations at the positions 32 and 158 mutations at positions were ranked 3rd and 12th in the design. A comparison of the model of the LW1 mutant (Y32L-D158G-I159C-L162R) [5] with the Y32L-D158G mutant retrieved from COP can be seen in Figure C.3.a.

For the remaining ligands, it was not possible to retrieve the experimental mutations via COP. Comparing the COP results with models containing the full set of experimental mutations showed that desired sequences cannot be selected at positions 32 and 158 in the absence of more distant secondary mutations, outside the scope of COP. One example of this is the Amf case. Santoro et al. showed that Amf is activated by a *Mj-TyrRS* variant carrying the mutations Y32T-E107T-D158P-I159L-L162A [7]. COP cannot identify the Y32T-D158P mutant for this ligand in the absence of the L162A mutation. The L162A mutation creates space to allow the movement of the backbone around position 158, which in turn allows proline to be placed at that position without any clashes with the ligand. (Figure C.3.b)

A similar observation is made in the designs for Nal. The *Mj*-TyrRS mutant Y32L-D158P-I159A-L162Q-A167V can activate Nal [8]. COP does not consider the Y32L mutation since introducing this mutation in the absence of the L162Q mutation results in L32 to clash with the ligand. The conformation of the L32 residue in the presence and absence of the L162Q mutation can be seen in Figure C.3.c.

- This study suggest a strong coupling between the L162 position and the positions Y32 and D158 in *Mj*-TyrRS. The identity of the residues at such secondary sites may have strong influences determining what is selected at the immediate vicinity of the ligand. Because this behavior is context dependent, COP may or may not agree with the selection experiments, depending on the system considered.

Designing the *E. coli* phenylalanyl-tRNA synthetase for the incorporation of *p*-acetyl phenylalanine into proteins *in vivo*

- *p*-Acetyl phenylalanine (*Acf*): Using the computational design program ORBIT, Datta et al. showed that the T251G-A294G double mutant of the *E. coli* phenylalanyl-tRNA synthetase (*Ec*-PheRS) is able to activate *Acf* [11]. In this study we tested whether if we can capture these mutations using COP. Unlike the *Mj*-TyrRS designs, there are no secondary mutations in this system.
- *Ec*-PheRS: Because a crystal structure of the *E. coli* PheRS is not available, a homology model of the *E. coli* PheRS was constructed [12] using the available structures from *Thermus thermophilus* [13, 14].
- *Design*: COP correctly identified the two positions, T251 and A294, clashing with the ligand. These mutations were designed for the recognition of *Acf*.
- *Results*: We were able to correctly identify the T251G-A294G mutations in *Ec*-PheRS required to activate *Acf* from all possible double mutants at these positions.

Designing the *S. cerevisiae* phenylalanyl-tRNA synthetase for the incorporation of cyano-phenylalanine into proteins *in vivo*

- *para- and meta-cyano phenylalanine (pCnf and mCnf)*: Cyano-phenylalanine is a probe for the internal electric field of proteins for vibrational Stark spectroscopy. (Figure C.4) Although this analog shows about 10-15% incorporation through the yeast PheRS variants, more than 50% incorporation is desired for spectroscopy experiments.
- *S. cerevisiae phenylalanyl-tRNA synthetase (Sc-PheRS)*: Because a crystal structure of Sc-PheRS is not available, a homology model of this enzyme was constructed [12] using the available structures from *Thermus thermophilus* [13, 14]. (Figure C.4.)
- *Design*: No mutation sites could be identified for pCnf via COP. However, for mCnf three mutants were proposed (mutants 1-3; Table C.1.c.). Because all these mutants contain a glycine substitution at position 412, we also tested mutants that have alanine at this site (mutants 4-6).
With no success from mutations 1-6, we expanded the sites considered for mutation, by including position S418 in this set. The combinatorial set of all triple mutants at the N412, T415 and S418 positions were evaluated for their interactions with mCnf. From this study four 3-fold and one 2-fold mutant was selected for experimental testing (mutants 7-11).
- *Results*: *In vivo* incorporation of mCnf was not observed for any of the mutants. We concluded that our homology model of Sc-PheRS may not be sufficiently accurate for a design study.

Introduction of *E. coli*-like mutations to the *S. cerevisiae* PheRS to achieve sufficient *in vivo* incorporation of 2-quinoxalinyllalanine (Qxa)

- *2-quinoxalinyllalanine (Qxa)*: Structurally similar to Nal, Qxa is a reporter of slow time-scale solvation of charges on proteins it is incorporated into.
- *Background*: Experiments [15] suggest a nitrogen at the ortho position of phenylalanine is not accepted well by the Sc-PheRS, but is better permitted by the *Ec*-PheRS:

- 2-Pyridylalanine is accepted by the *Ec*-PheRS [16], but not the *Sc*-PheRS.
- Nal is accepted both by mutant *S. cerevisiae* (T415G) and *E. coli* PheRS (T251G-A294G).
- 2-Quinolinyllalanine (Qla) and Qxa are not accepted by *Sc*-PheRS (T415G).
- Four positions in the *Sc*-PheRS binding site (336, 412, 414, 436) that differ from the corresponding residues in the *Ec*-PheRS were identified. To see if these changes have a role in the differences observed between the two enzymes, residues in *Sc*-PheRS were mutated to their *E. coli* identities. The T415G mutation was included in all mutants constructed since this mutant creates the space required to bind the naphthyl ring.
- Qla, which is commercially available, was used instead of Qxa when evaluating the mutants. Seven *Sc*-PheRS variants containing *E. coli*-like mutations were tested with Qla. The E376Q mutation was also tested to see if it can help decrease tryptophan misincorporation. (Table C.1.c)
- *Results:* Low levels of protein expression was observed with most mutants. Mutants with higher expression levels, which provided good MALDI-MS signals, showed equal or lower levels of Qla incorporation as the T415G mutant. Higher expression rates are needed to establish the behavior of other mutants.

Table C.1

List of aminoacyl-tRNA synthetase mutants tested in design studies in this section

a) IleRS mutants experimentally tested with Hfv

<i>IleRS</i>	A	AF	F	F518
P 46	A	A	P	P
W 518	W	W	W	F
W 558	W	F	F	W

b) ValRS mutants identified by COP and experimentally tested with Hfv

<i>ValRS</i>	SD	SV	AV	GV	CV	AF	AN	AD
P 42	S	S	A	G	C	A	A	A
I 491	D	V	V	V	V	F	N	D

ValRS single-mutants experimentally tested with Hfv

<i>ValRS</i>	1	2	3	4	5	6	7	8
P 42	G	A	C	P	P	P	P	P
W 456	W	W	W	F	W	W	W	W
I 491	I	I	I	I	V	I	I	I
W 495	W	W	W	W	W	F	Y	H

Figure C.1

Comparison of the T. thermophilus IleRS and ValRS binding sites.

A comparison of the binding sites for ValRS (yellow) and IleRS (purple) from *T. thermophilus* is shown to highlight the differences and similarities between the two enzymes. The ligands used in the studies of these enzymes are shown on the right-hand side.

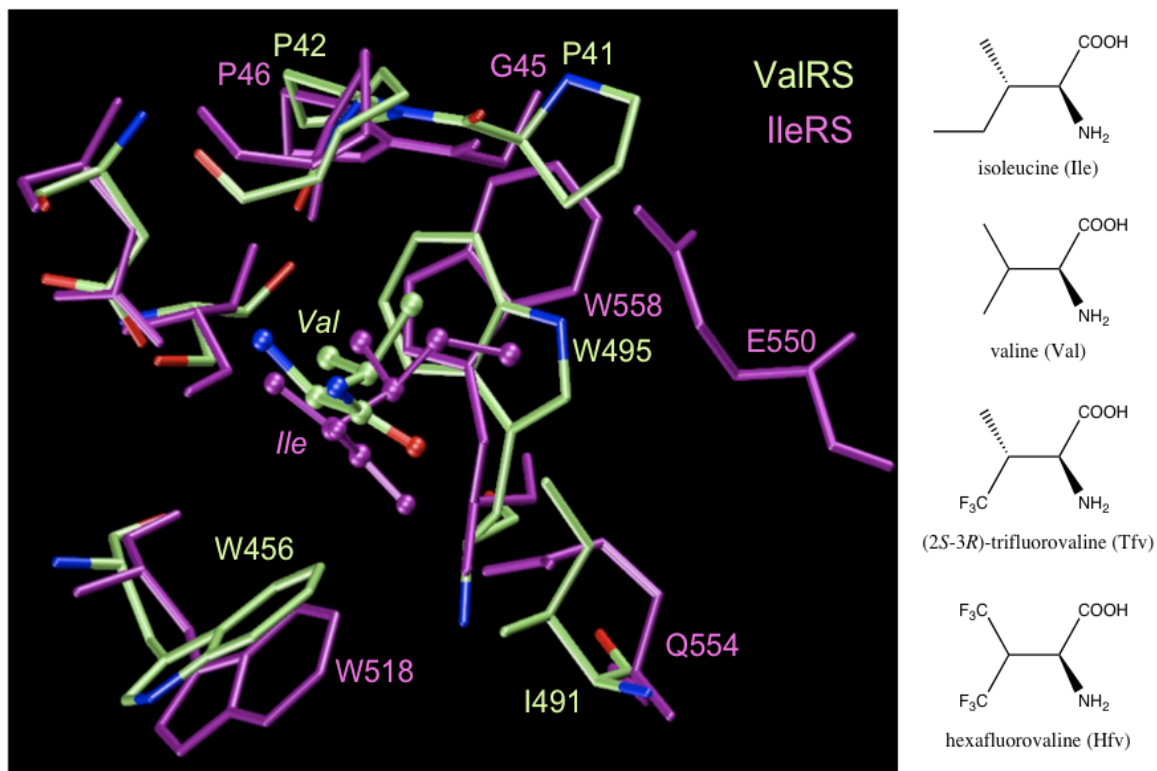


Figure C.2

Randomized positions in the Mj-TyrRS binding site and the ligands studied.

Positions randomized in the *Mj*-TyrRS saturation mutagenesis library used in the selection experiments performed by the Schultz laboratory (cyan) are shown in the presence of the native ligand, tyrosine (yellow). The ligands discussed in this study are presented on the right-hand side.

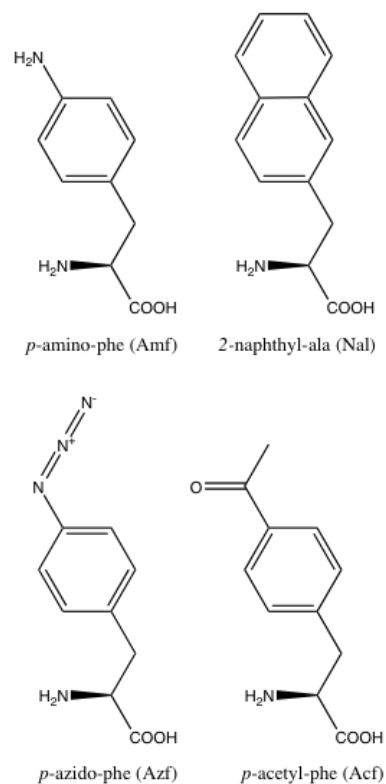
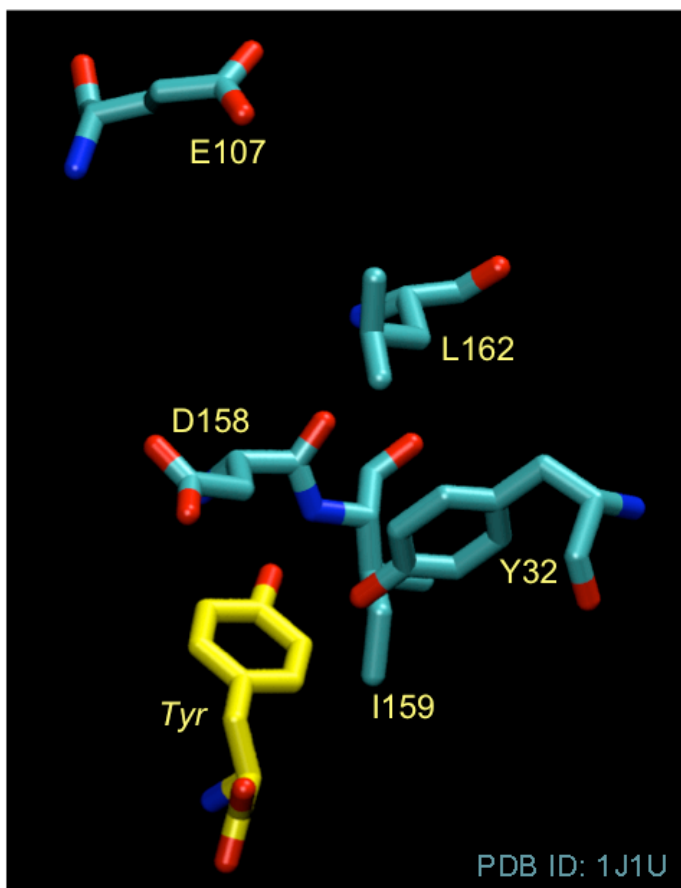


Figure C.3

Comparison of Mj-TyrRS models bearing mutations identified by experiments with double mutants received from the COP design procedure.

Mj-TyrRS models containing the full set of experimental mutations (cyan) were shown to prefer the non-natural ligand over the tyrosine in the binding-energy analyses performed. When residues identified as clashing with the ligand by the COP design procedure (Y32 and D158) were mutated to their experimental identities, the purple models were obtained. These two models compare favorably when mutants for Acf are analyzed. However, the double mutants for Nal and Amf show significant differences in their binding characteristics from the experimental mutants identified for these ligands.

- a) Comparison of the Acf-bound models of the L32-G158 mutant from COP and the four-fold mutant LW1.

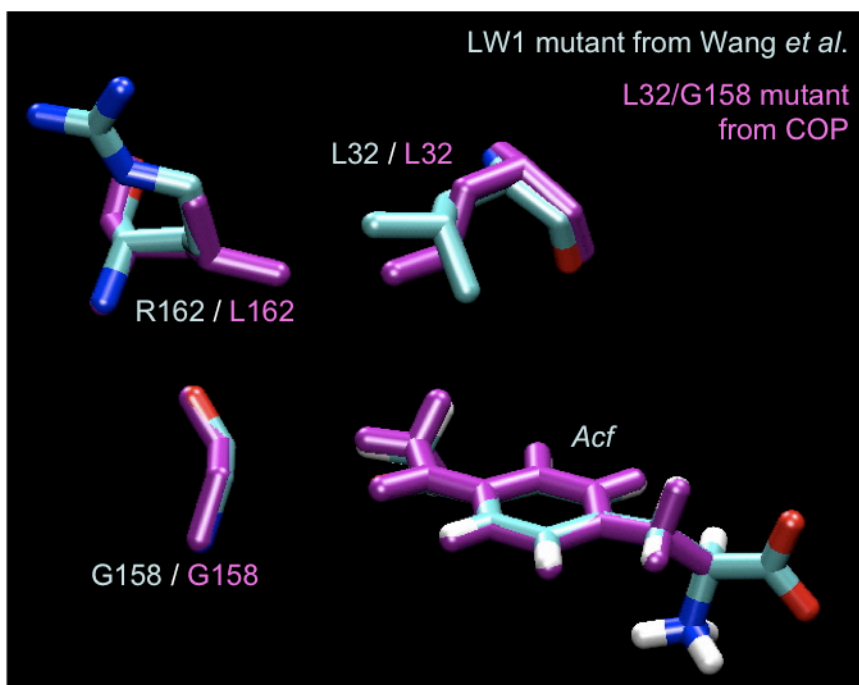
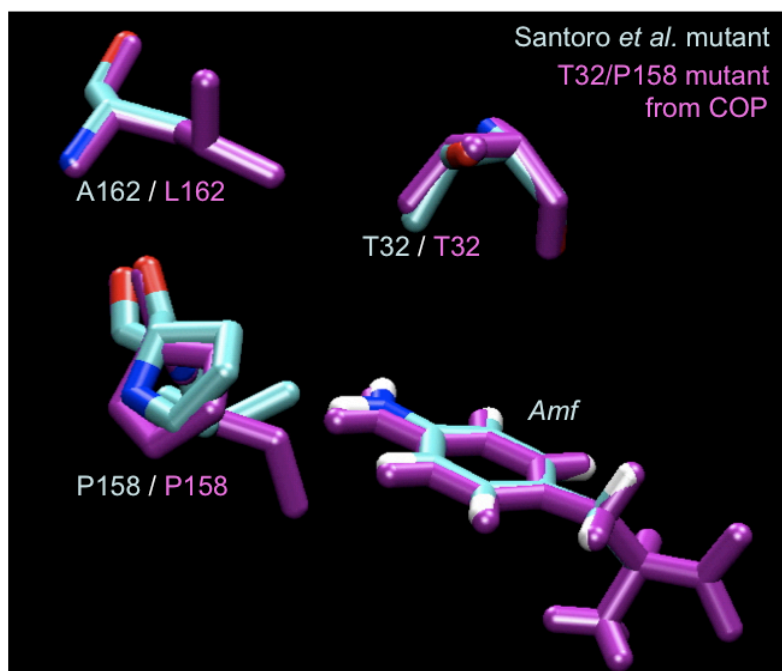


Figure C.3 (continued)

- b) Comparison of the Amf-bound models of the T32-P158 mutant from COP and the five-fold mutant from Santoro et al.



- c) Comparison of the Nal-bound models of the L32-P158 mutant from COP and the five-fold mutant from Wang et al.

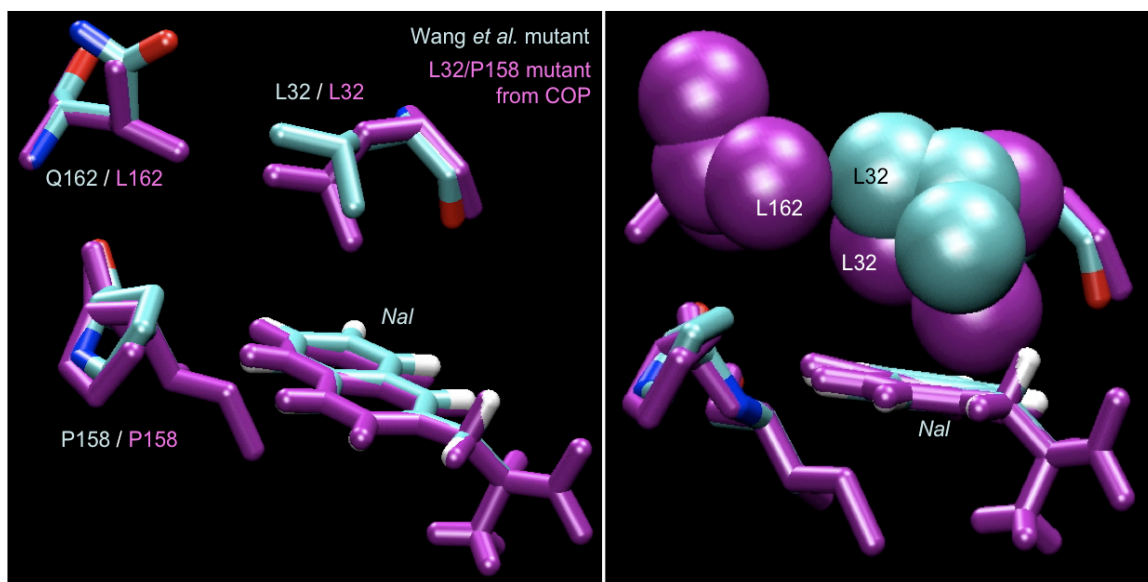
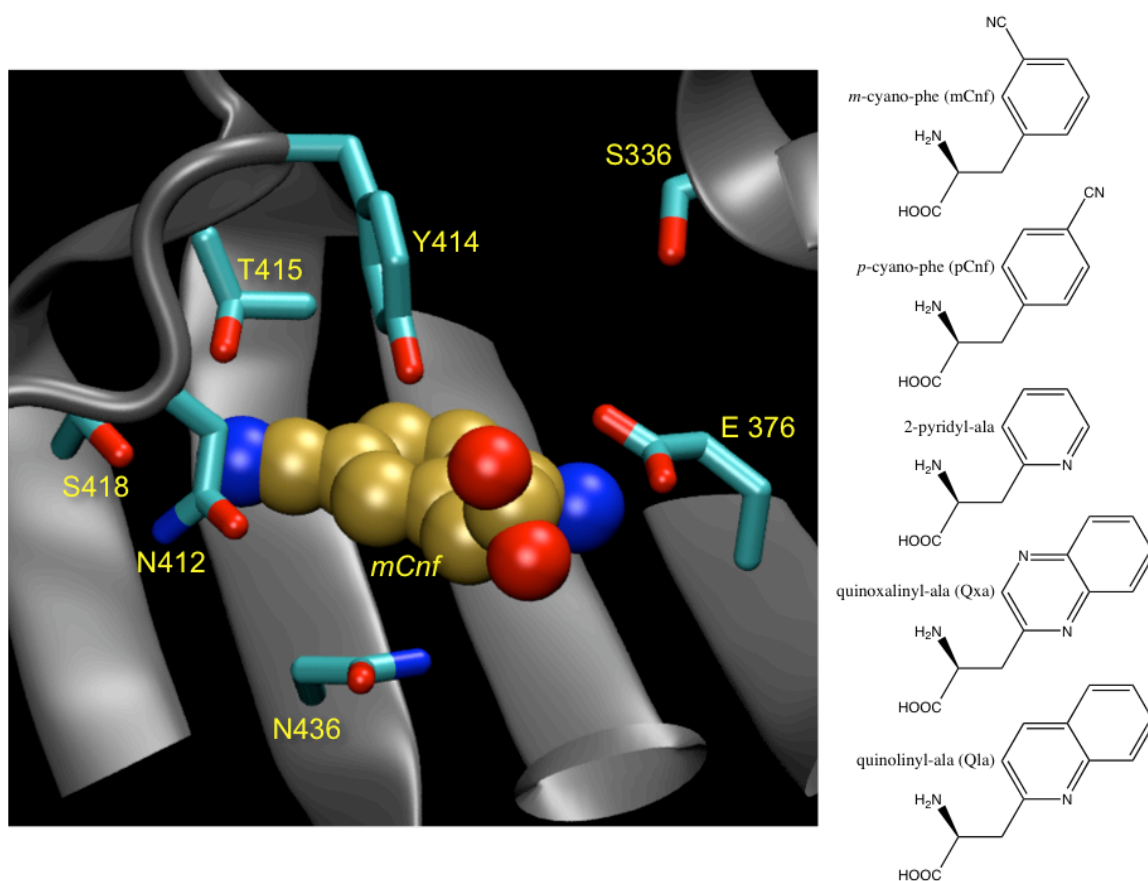


Figure C.4

The homology model of the S. cerevisiae PheRS.

The homology model of the *S. cerevisiae* PheRS is shown complexed with the ligand mCnf. Residues interacting with the ligand are shown in cyan. The ligands discussed in this study are shown on the right-hand side.



References

- [1] Nureki O, Vassylyev DG, Tateno M, Shimada A, Nakama T, Fukai S, Konno M, Hendrickson TL, Schimmel P and Yokoyama S, "Enzyme structure with two catalytic sites for double-sieve selection of substrate," *Science*. **1998**, *280*:578.
- [2] Zhang DQ, Vaidehi N, Goddard WA, Danzer JF and Debe D, "Structure-based design of mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase for incorporation of O-methyl-l-tyrosine," *Proc Natl Acad Sci U S A*. **2002**, *99*:6579.
- [3] Fukai S, Nureki O, Sekine S, Shimada A, Tao JS, Vassylyev DG and Yokoyama S, "Structural basis for double-sieve discrimination of l-valine from l-isoleucine and l-threonine by the complex of tRNA(val) and valyl-tRNA synthetase," *Cell*. **2000**, *103*:793.
- [4] Wang L, Brock A, Herberich B and Schultz PG, "Expanding the genetic code of *Escherichia coli*," *Science*. **2001**, *292*:498.
- [5] Wang L, Zhang ZW, Brock A and Schultz PG, "Addition of the keto functional group to the genetic code of *Escherichia coli*," *Proc Natl Acad Sci U S A*. **2003**, *100*:56.
- [6] Chin JW, Santoro SW, Martin AB, King DS, Wang L and Schultz PG, "Addition of p-azido-l-phenylalanine to the genetic code of *Escherichia coli*," *J Am Chem Soc*. **2002**, *124*:9026.
- [7] Santoro SW, Wang L, Herberich B, King DS and Schultz PG, "An efficient system for the evolution of aminoacyl-tRNA synthetase specificity," *Nat Biotechnol*. **2002**, *20*:1044.
- [8] Wang L, Brock A and Schultz PG, "Adding l-3-(2-naphthyl)alanine to the genetic code of e-coli," *J Am Chem Soc*. **2002**, *124*:1836.
- [9] Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K and Yokoyama S, "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion," *Nat Struct Biol*. **2003**, *10*:425.
- [10] Zhang Y, Wang L, Schultz PG and Wilson IA, "Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-l-tyrosine," *Protein Sci*. **2005**, *14*:1340.
- [11] Datta D, Wang P, Carrico IS, Mayo SL and Tirrell DA, "A designed phenylalanyl-tRNA synthetase variant allows efficient *in vivo* incorporation of aryl ketone functionality into proteins," *J Am Chem Soc*. **2002**, *124*:5652.
- [12] Kekenus-Huskey PM, Unpublished results, **2005**
- [13] Fishman R, Ankilova V, Moor N and Safro M, "Structure at 2.6 Å resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese," *Acta Crystallogr D Biol Crystallogr*. **2001**, *57*:1534.
- [14] Reshetnikova L, Moor N, Lavrik O and Vassylyev DG, "Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue," *J Mol Biol*. **1999**, *287*:555.

- [15] Kwon I, Unpublished results, **2005**
- [16] Kirshenbaum K, Carrico IS and Tirrell DA, "Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues," *Chembiochem.* **2002**, 3:235.

Appendix D.***List of Software Written for and Used in the Modeling, Design, and Study of Methionyl-tRNA Synthetase Variants***

A list of software written to facilitate the study and design of methionyl-tRNA synthetase (MetRS) mutants is presented here. The list contains software written in Perl and Python, distributed into multiple categories. A brief description is provided for each entry.

Table D.1

Software written for the study of MetRS mutants.

The programs listed here can be found on hulk.wag.caltech.edu, inside the directories “/ul/caglar/Python”, “/ul/caglar/Perl”. Most scripts will print usage information when executed without arguments.

- a) List of utilities written to aid in the construction and experimental characterization of MetRS mutants

Name	Description
formatCodingSeq.pl	Reads in a DNA sequence, translates genes and reformats sequences, finds DNA/protein sequences, as well as ORFs, within the entered sequence
getMSrange.pl	Reads in a ABI Voyager mass spectrum data file as text, and saves selected data ranges, detects peaks, and plots spectra using Gnuplot
degenerateCodons.pl	Given a set of desired and/or undesired amino acids, searches for degenerate codons that code the amino acids specified, and lists any that are found
mutationMassShifts.pl	Provides suggestions for the identities for the incorporated amino acids in a residue-specific incorporation experiment to compare with mass spectrum data
primerReverse.pl	Prints the complementary stand and the reverse complement of the entered DNA sequence

Table D.1 (continued)

b) List of utilities for modifying three dimensional models of molecules in BGF or PDB formats

Name	Description
playWithBGF	Makes selections in a PDB or a BGF file and performs operations on the selection
addChainDsg2BGF.pl	Assigns the given chain designator to an atom range specified in a BGF file
addCharges2bgf	Copies charges between two similar BGF files, or allows you to enter them interactively
bgf2jagCoords.pl	Reads atom coordinates from a BGF file and reformats them to be pasted into a JAGUAR input file
copyCoords_bgf2bgf.pl	Copies the atom coordinates between two BGF files that have the same number of atoms
setBGF_DESCRP.pl	Adds or sets the DESCRP line in a BGF file
supplementBGFrecs.pl	Supplements missing fields in a BGF file based on a reference BGF, user-defined defaults or a conversion file
toggleMovableRecs.pl	Sets atoms in a BGF file movable or fixed based on a playWithBGF atom-list file
translateBGF.pl	Translates all coordinates in a BGF file by the given vector
atomicRuler.py	Measures distances, angles or dihedrals between atoms in a BGF file
checkForMissing- AtomsInBGF.py	Reads a BGF file and checks for the number of atoms in each residue against a pre-defined dictionary to see if any residues are missing or have extra atoms

Name	Description
listClashesWith- RotLib.py	Determines clashes between a binding site, and a SCREAM rotamer library generated by MOLECULEGL
mergeBGFs.py	Merges BGF files into a single file, in the order they are entered
reorderBackbone- Atoms.py	Reorders atoms in each residue of a BGF file so that backbone atoms are followed by side chain atoms
resListFromBGF.py	Reads a BGF file and writes a playWithBGF residue-list file (for use with option -R)
sumCharge.py	Sums charges and reports them by residue, chain, or structure
zeroCharge.py	Sets all charges in a BGF file to zero

Table D.1 (continued)

- c) List of programs that combine multiple types of functionalities (usually through wrappers, or other scripts) in order to achieve a certain calculation, e.g. a binding-energy calculation

Name	Description
getVerticalBE.pl	Calculates vertical binding energies given a BGF file with a ligand marked by its chain identifier.
determineSASAon-Ligand.pl	Calculates solvent accessible surface area using the MPSIM final energy output files for the selected atoms
seqMin.pl	Runs sequential MPSIM minimizations setting different sets of atoms movable at each stage.
compileLibFromBGF.py	Prepares a SCREAM-compatible rotamer library file for a given set of BGF structures of a certain molecule, and lists MPSIM energies for each rotamer
getEnerDistPerAA.py	Tabulates energy contribution of selected residues to the total energy using MPSIM energy lists by atom

Table D.1 (continued)

d) List of programs that parse and act on the output of other programs

Name	Description
diffCavan.pl	Calculates the difference of interaction energies reported by cavity_analysis.pl between two files
diffFinEner.pl	Given two MPSim final energy lists, calculates the energy differences between the two files
diffPWB_iaTables.pl	Given two playWithBGF output files containing nonbond interaction analysis tables based on the same selection on two different structures, this script calculates the difference between the two outputs
finEnerBE.pl	Given MPSim final energy lists for the enzyme, ligand and complex this script calculates the binding-energy contribution for each atom
get-tvnE.pl	Calculates average energy from an MPSim dynamics run output (adapted from David Zhang)
mergePWBatomLists.pl	Merges multiple atom-list files written by playWithBGF (option -l) into a single list
jaguarBindingE.py	Calculates and tabulates the binding energy of a complex (AB - A - B) from JAGUAR output files
jaguarOutputToBGF.py	Given a BGF file and a corresponding JAGUAR output file, writes a new BGF file containing the JAGUAR coordinates and Mulliken charges
tabulateMPSimEout.py	Tabulates the final energies from MPSim output files
tabulateVBEout.py	Tabulates binding-energy components from getVerticalBE.pl output directories

Table D.1 (continued)**e) List of programs used in the binding site design calculations**

Name	Description
cbMuts.py	Customizable program for generating combinatorial mutants, optimizing structures, and determining binding energies through the use of SCREAM and MPSIM. (Used in the 2005 STUDY)
evaluateLYHmutant.py evaluateMutant.py	For a specified mutation, runs MPSIM minimizations, interaction energy calculations with playWithBGF, and binding-energy calculations with DELPHI solvation. (Used in the 2008 STUDY)
foreachMutn.pl	Given the amino-acid pool at each mutation site, enumerates all combinatorial mutations and executes the given commands for each mutation.
compileMutationData.py compileMutationVBE.py	Each script reads the specified evaluateMutant.py data output files and tabulates the results from each calculation
shelveCompleted- Mutns.py	Given a calculation prefix, checks the for errors in each evaluateMutant.py output and for completed calculations, files the output file into appropriate directories

Table D.1 (continued)

f) List of wrappers, which provide easy access to various programs

Name	Description
runMPSim.pl	A customizable wrapper for MPSIM for running minimizations, simulated annealing, dynamics and energy evaluations.
runScream.py	A wrapper for SCREAM to perform side chain placement and mutations on BGF structures
convertTov400.pl	A wrapper for John Wendel's bgf2fsm, which converts a BGF file into the v400 format
generateLigand-RotLib.py	Uses MOLECULEGL (version 1) to generate a ligand rotamer library readable by SCREAM
run_delphi.py	A wrapper for DELPHI which calculates the contribution of PB solvation to the energy of a given BGF structure. (Written by John Wendel)

Table D.1 (continued)

g) List of Python modules utilized in Python scripts listed here

Name	Description
BGF.py	Module containing the BgfAtom and BgfFile classes for reading, writing and manipulating BGF files
crunners.py	Module containing adaptor functions or wrappers that run common Perl scripts or standalone programs that are employed for structure optimization, analysis or design
cutils_getopt.py	An extension on the Python's own getopt module to allow it to accept options of type +X
cutils.py	Module containing general utilities for data conversion, file input-output, program execution

Table D.1 (continued)**h) List of general utilities**

Name	Description
addAvg.pl	Calculates the sums and averages of entered or numbers ignoring any non-numerical characters
changeFileNames.pl	Substitutes a given text for a certain text in the file names of the specified files
execInDirs.pl	Script to execute the same command in multiple directories, or recursively in all subdirectories
foreachFile.pl	Script to execute a shell command for each of the files specified in the command line, repetitively.
foreachWord.pl	Script to execute a shell command for each word entered in the command line, repetitively.
memlist.pl	Sorts and tabulates disk usage on the specified files and directories
r3D.pl	Calculates the distance between two points in space
runViaPBS.pl	Writes a PBS file for the entered command and submits it to the job queue
secs2time.pl	Translates seconds into common time
sortTable.pl	Interactive script to sort an arbitrary-size table containing numerical data
tarTextFiles.pl	Creates a tar archive containing all the text files found in the directories or files specified.
tarTextFilesInDirs.pl	
zipTextFiles.pl	Compresses any text files specified using gzip