Excerpts of information given to HAMB, showing the types of information we are able to give HAMB about the known relationships among the attributes and our reasoning behind the information. In addition to known relationships among the attributes, we are also able to provide HAMB with other types of information, such as attributes and values that are uninteresting or uninformative, such as UNKNOWN-CHEMICAL-CLASS, which HAMB uses to avoid reporting potentially uninteresting or uninformative rules.

- Associations derived from chemical properties and classifications of the additives and species. For example, the additive sodium chloride breaks down into the species sodium and chloride, and the species imadazole belongs to the chemical class m.other, has a perceived role of buffer (ROLE-B), has a pK_a in the neutral range, and all species are used to calculate ionic strength.

  (known-association ADD-SODIUM-CHLORIDE SPEC-SODIUM)
  (known-association ADD-SODIUM-CHLORIDE SPEC-CHLORIDE)
  (known-association SPEC-SODIUM SPEC-CHLORIDE)
  (known-association CHM-CLSS-M.OTHER SPEC-IMIDAZOLE)
  (known-association CHM-CLSS-M.OTHER ROLE-B)
  (known-association CHM-CLSS-M.OTHER NEUTRAL)
  (known-association CHM-CLSS-M.OTHER IONIC-STR)

- Net charge is directly related to the species and indirectly to properties related to the species, such as the chemical class

  (known-association CH>=2 SPEC-CALCIUM_[II])
  (known-association CH<= -2 SPEC-SULFATE)
  …

- Because the additive sodium phosphate's perceived role is as a buffer (ROLE-B), there is a known chemical association among the additive sodium phosphate, the perceived role of buffer, and the species sodium and phosphate.

  (known-association ADD-SODIUM-PHOSPHATE)
  (known-association ROLE-B SPEC-SODIUM)
  (known-association ROLE-B SPEC-PHOSPHATE)

- All additives and species with titrateable groups are associated with buffering capacity, since buffering capacity is derived from the properties of species

  (known-association BUFFERING-CAPACITY ROLE-P.AM)
  (known-association BUFFERING-CAPACITY SPEC-Amine)
  (known-association BUFFERING-CAPACITY SPEC-Carbonate)
  …

- Similary, because ionic-strength is derived from the properties of species, ionic strength is chemically related to all additives that dissociate into charged species.

  (known-association IONIC-STR SPEC-AZIDE)
  (known-association IONIC-STR SPEC-EDTA)
  …

- Pclass, sclass, tclass, and qclass form the macromolecule heirarchy described in this paper; therefore, these attributes are related to each other and to the macromolecule id and the molecular weight of the macromolecule.

  (defined-by macmolwgt macmol-id)
  (defined-by macmolwgt macmol-pclass)
  (defined-by macmolwgt macmol-sclass)
  (defined-by macmolwgt macmol-tclass)
  (defined-by macmolwgt macmol-qclass)
  …

- Abstractions of the crystallization method:

  (abstraction crysmeth_scale crysmeth_desc)
  (abstraction crysmeth_type crysmeth_scale)
  (abstraction crysmeth_type crysmeth_desc)
  …

- Abstractions of the form of the crystal:

  (abstraction spgrpid1 spgrps_desc)
  (abstraction spgrps_desc crhabit)
  (abstraction spgrps_desc crform)
  …

- Some of the relationships used to form the macromolecule hierarchy:

  (abstraction macmol-pclass macmol-sclass)
  (abstraction macmol-sclass macmol-tclass)
  (abstraction macmol-pclass macmol-tclass)
  …

A subset of over 300 rules discovered by HAMB. Patterns discovered from HAMB organized according to "rule families".

Sulfhydryls and chelators

Rule	TP	FP	SEN	PPV
Species with strongly basic PKAs are not present ➔ organic sulfhydryls are not present ****	727	0	0.35	1.00
Species with strongly basic PKAs are not present ➔ additives with perceived roles of reducing agent are not present	727	0	0.35	1.00
EDTA is not present ➔ 2-mercaptoethanol is not present	2019	38	0.93	0.98
Organic chelators are not present ➔ 2-mercaptoethanol is not present	2007	37	0.93	0.98
Inorganic alkaline metals are not present and organic sulfhydryls are not present ➔ organic chelators are not present	991	25	0.48	0.98
Species with strongly basic PKAs are not present ➔ organic chelators are not present	712	15	0.35	0.98
Species with ≥ 3 titrateable groups are not present ➔ organic sulfhydryls are not present	1035	36	0.50	0.97
Species with net charges ≥ 3 are not present and organic sulhydryls are not present ➔ organic chelators are not present	1467	53	0.72	0.97
Species with net charges ≤ -2 are not present ➔ organic sulfhydryls are not present	729	23	0.36	0.97
Species with net charges ≤ -3 are not present ➔ organic sulfhydryls are not present	1212	43	0.59	0.97
Species with neutral titrateable groups are not present ➔ organic sulfhydryls are not present	901	29	0.44	0.97
Species with strongly acidic PKAs are not present ➔ organic sulfhydryls are not present	1018	32	0.50	0.97
Species with strongly acidic PKAs are not present ➔ additives with perceived roles of reducing agent are not present	1018	32	0.50	0.97
Species with neutral titrateable groups are not present ➔ additives with perceived roles of reducing agent are not present	900	30	0.44	0.97
Species with net charges of ≤ -3 are not present ➔ additives with perceived roles of reducing agent are not present	1212	43	0.59	0.97
Species with net charges of ≤ -2 are not present ➔ additives with perceived roles of reducing agent are not present	729	23	0.36	0.97
Additives with perceived roles of reducing agent	1955	93	0.95	0.96

are not present ➔ EDTA is not present				
Additives with perceived roles of anti microbial are not present ➔ organic chelators are not present	1607	72	0.79	0.96
Species with ≥ 4 titrateable groups are not present ➔ organic sulfhydryls are not present	1675	62	0.82	0.96
Species with ≥ 4 titrateable groups are not present ➔ additives with perceived roles of reducing agent are not present	1675	62	0.82	0.96
Additives with perceived roles of chelator are not present ➔ additives with perceived roles of reducing agent are not present	1947	98	0.95	0.95
Additives with perceived roles of reducing agent are not present ➔ additives with perceived roles of chelator are not present	1947	101	0.95	0.95
Additives with perceived roles of reducing agent are not present ➔ organic chelators are not present	1946	102	0.95	0.95
EDTA is not present ➔ organic sulfhydryls are not present	1958	99	0.95	0.95
The species EDTA is not present ➔ organic sylfhydryls are not present	1956	96	0.95	0.95
EDTA is not present ➔ additives with perceived roles of reducing agent are not present	1955	102	0.95	0.95
The species EDTA is not present ➔ additives with perceived roles of reducing agent are not present	1953	99	0.95	0.95
Organic chelators are not present ➔ organic sulfhydryls are not present	1948	96	0.95	0.95
Organic sulfhydryls are not present ➔ EDTA is not present	1958	95	0.95	0.95
Organic sulfydryls are not present ➔ organic chelators are not present.	1948	105	0.95	0.95
Organic sulfhydryls are not present ➔ additives with perceived roles of chelator are not present	1949	104	0.95	0.95
Additives with perceived roles of reducing agent are not present ➔ the species EDTA is not present	1953	95	0.95	0.95
Organic sulfhydryls are not present ➔ the species EDTA is not present	1956	97	0.95	0.95
Species with dipole charges are not present ➔ organic sulfhydryls are not present	1599	80	0.78	0.95
Species with somewhat mixed charges are not present ➔ additives with perceived roles of chelator are not present	1740	93	0.85	0.95
Species with dipole charges are not present ➔	1599	80	0.78	0.95

additives with perceived roles of reducing agent are not present				
Species with dipole charges are not present ➔ organic chelators are not present	1842	109	0.90	0.94
Organic sulfhydryls are present with a total concentration between 0.0003M and 0.001M ➔ the species EDTA is present with a concentration ≤ 0.001M	26	52	0.29	0.33
Organic sulfhydryls are present with a total concentration ≤ 0.001M ➔ organic chelators are present with a total concentration between 0.0003M and 0.001M	26	52	0.28	0.33
Organic sulfhydryls are present with a total concentration ≤ 0.001M ➔ additives with perceived roles of chelator are present with a total concentration between 0.0003M and 0.001M	26	52	0.28	0.33
Additives with perceived roles of reducing agent are present with a total concentration between 0.0003M and 0.001M ➔ the species EDTA is present with a concentration ≤ 0.001M	25	56	0.27	0.31
Additives with perceived roles of reducing agent are present with a total concentration ≤ 0.001M ➔ additives with perceived roles of chelator are present with a total concentration between 0.0003M and 0.001M	25	56	0.27	0.31
The species EDTA is present with a concentration between 0.0003M and 0.001M ➔ organic sulfhydryls are present with a total concentration ≤ 0.001M	26	65	0.33	0.29
Organic chelators are present with a total concentration between 0.0003M and 0.001M ➔ organic sulfhydryls are present with a total concentration ≤ 0.001M	26	66	0.33	0.28
Additives with perceived roles of chelator are present with a total concentration between 0.0003M and 0.001M ➔ additives with the perceived roles of reducing agent are present with a total concentration ≤ 0.001M	25	67	0.31	0.27
The species EDTA is present with a concentration between 0.0003M and 0.001M ➔ additives with perceived roles of reducing agent are present with a total concentration ≤ 0.001M	25	66	0.31	0.27
EDTA is present with a concentration between 0.0003M and 0.001M ➔ additives with perceived roles of reducing agent are present with	23	65	0.28	0.26

a total concentration ≤ 0.001M				

Spermine, cacodylate, nucleic acids, and magnesium chloride

Rule	TP	FP	SEN	PPV
Inorganic divalent metals are not present ➔ spermine is not present	1802	1	0.82	1.00
Sodium cacodylate is not present ➔ spermine tetrahydrochloride is not present	2105	9	0.96	1.00
Species with ≥ 2 titrateable groups are not present ➔ metabolite nucleotides are not present	974	0	0.45	1.00
Species with acidic PKAs are not present ➔ metabolite nucleotides are not present	952	0	0.44	1.00
Species with acidic PKAs are present and the macromolecule is a very small peptide ➔ additives with unknown perceived roles are present *** fix***	9	0	2193.00	1.00
Species with net charges of ≤ -1 are present with a total concentration > 2.23M and additives with perceived roles of precipitating agent are present with a total concentration between 0.7M and 0.701M ➔ the macromolecule is a ribonuclease	11	0	0.18	1.00
Species with net charges of ≥ 1 are present with a total concentration > 2.41M and additives with perceived roles of precipitating agent are present with a total concentration between 0.7M and 0.701M ➔ the macromolecule is a ribonuclease	11	0	0.18	1.00
Species with strongly acidic PKAs are not present ➔ metabolite nucleotide are not present	1048	2	0.49	1.00
Species with strongly acidic PKAs are not present ➔ the species maleate is not present	1050	0	0.48	1.00
Species with strongly basic PKAs are not present ➔ the species spermine is not present	727	0	0.34	1.00
The macromolecule is a protein ➔ spermine is not present	1921	1	0.87	1.00
The macromolecule is a protein ➔ spermine tetrahydrochloride is not present	1918	4	0.87	1.00
The macromolecule is an enzyme ➔ spermine is not present	809	0	0.38	1.00
The macromolecule is an enzyme ➔ the species spermine is not present	809	0	0.38	1.00
The species magnesium II is not present ➔ spermine is not present	1985	3	0.90	1.00
The species magnesium II is not present ➔ spermine tetrahydrochloride is not present	1981	7	0.90	1.00

The macromolecule is a little miscellaneous protein ➔ species with a basic PKA are not present	139	112	0.27	0.55
Species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M ➔ the macromolecule is a nucleic acid	40	36	0.35	0.53
The macromolecule is a deoxy oligonucleotide ➔ species with net charges of = 0 are present with a total concentration between 0.21M and 2.34M	34	36	0.17	0.49
The macromolecule is a little miscellaneous protein ➔ species with net charges of ≤ -1 are not present	108	143	0.32	0.43
Magnesium chloride is present with a concentration between 0.012M and 0.04M ➔ species with net charges of ≥ 3 are present with a total concentration between 0.002M and 0.01M	14	20	0.35	0.41
The species magnesium II is present with a concentration between 0.01M and 0.015M ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	9	13	0.30	0.41
Species with basic PKAs are not present ➔ the macromolecule is a miscellaneous little protein	78	121	0.34	0.39
The species spermine is present with a concentration between 0.0008M and 0.002M ➔ the species cacodylate is present with a total concentration ≤ 0.01M	10	16	0.29	0.38
Sodium cacodylate is present ➔ the species spermine is present with a concentration between 0.0025M and 0.006M	6	10	0.60	0.37
The macromolecule is a trna ➔ the species cacodylate is present with a total concentration ≤ 0.01M	12	21	0.34	0.36
The macromolecule is an oligonucleotide ➔ species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M	28	50	0.37	0.36
Sodium cacodylate is present with a concentration between 0.02M and 0.03M ➔ magnesium chloride is present with a concentration between 0.012M and 0.04M	9	17	0.26	0.35
The macromolecule is an nucleic acid ➔ species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M	40	74	0.53	0.35
Species with 0 titrateable groups are present and species with titrateable groups ≥ 2 are not present ➔ the macromolecule is a little soluble protein	62	118	0.25	0.34
Species with net charges of ≤ -2 are not present	124	244	0.49	0.34

and additives with perceived roles of buffer are not present ➔ the macromolecule is a little soluble protein****				
Species with ≥ 1 titrateable groups are not present ➔ the macromolecule is a little soluble protein	58	117	0.23	0.33
Species with net charges of ≤ -2 are not present and additives with perceived roles of buffer are present with a total concentration between 0.014M and 0.045M ➔ the macromolecule is a deoxy oligonucleotide****	38	76	0.54	0.33
The macromolecule is a deoxy oligonucleotide ➔ species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M	23	47	0.30	0.33
The species cacodylate is present with a concentration between 0.002M and 0.03M ➔ magnesium chloride is present with a concentration between 0.012M and 0.04M	10	20	0.29	0.33
The macromolecule is a nucleic acid ➔ inorganic divalent species are present with a total concentration between 0.01M and 0.04M**	36	78	0.50	0.32
The macromolecule is a deoxy oligonucleotide ➔ inorganic divalent species are present with a total concentration >0.01M and 0.04M**	21	49	0.29	0.30
The species cacodylate is present with a concentration between 0.002M and 0.03M ➔ the species magnesium II is present with a concentration between 0.001M and 0.015M	9	21	0.41	0.30
Magnesium chloride is present with a concentration between 0.012M and 0.04M ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	10	24	0.33	0.29
The species cacodylate is present with a concentration ≤ 0.01M ➔ the species spermine is present with a concentration between 0.0008M and 0.002M	10	25	0.38	0.29
The macromolecule is a nucleic acid ➔ the species magnesium II is present with a concentration between 0.004M and 0.01M	31	83	0.38	0.27
The macromolecule is an oligonucleotide ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	21	57	0.70	0.27
The macromolecule is a little miscellaneous protein ➔ species with net charges of = 0 are present	66	185	0.38	0.26
The macromolecule is a little miscellaneous protein ➔ species with net charges of = 0 are	60	191	0.32	0.24

present with a total concentration > 2.34m				
The macromolecule is a deoxy oligonucleotide ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	16	54	0.53	0.23
The macromolecule is a nucleic acid ➔ inorganic divalent species are present with a total concentration between 0.005M and 0.01M**	26	88	0.30	0.23
Species with net charges of ≥ 1 are not present ➔ the macromolecule is a very small peptide	73	253	0.90	0.22
The macromolecule is an oligonucleotide ➔ magnesium chloride is present with a concentration between 0.012M and 0.04M	17	61	0.50	0.22
Additives with perceived roles of buffer are not present ➔ the macromolecule being crystallized is a small soluble protein	163	617	0.65	0.21
The macromolecule is a deoxy oligonucleotide ➔ the species spermine is present with a concentration between 0.0008M and 0.002M	15	55	0.58	0.21
The macromolecule is a nucleic acid ➔ magnesium chloride is present with a concentration between 0.012M and 0.04M	24	90	0.71	0.21
The macromolecule is an nucleic acid ➔ species with net charges of ≥ 2 are present with a total concentration between 0.005M and 0.01M	24	90	0.32	0.21
Species with net charges of ≤ -2 are not present ➔ the macromolecule is a small soluble protein.	154	598	0.61	0.20
Species with net charges of ≤ -1 are not present ➔ the macromolecule is a very small peptide	65	272	0.80	0.19
The macromolecule is an oligonucleotide ➔ the species magnesium II is present with a concentration between 0.01M and 0.015M	15	63	0.68	0.19
Species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	14	62	0.47	0.18
Species with net charges of ≥ 2 are present with a total concentration between 0.01M and 0.03M ➔ the species cacodylate is present with a concentration between 0.02M and 0.03M	14	62	0.47	0.18
The macromolecule is a nucleic acid ➔ the species cacodylate is present with a total concentration between 0.02M and 0.03M	20	94	0.57	0.18
The macromolecule is a nucleic acid ➔ the species cacodylate is present with a total concentration ≤ 0.01M	21	93	0.70	0.18
The macromolecule is a nucleic acid ➔ inorganic	19	95	0.27	0.17

divalent species are present with a total concentration between 0.0025M and 0.005M**				
The macromolecule is a nucleic acid ➔ the species magnesium II is present with a concentration between 0.01M and 0.015M	19	95	0.86	0.17
The macromolecule is a nucleic acid ➔ magnesium chloride is present with a concentration between 0.005M and 0.01M	18	96	0.40	0.16
The macromolecule is a nucleic acid ➔ magnesium chloride is present with a concentration less than 0.005M	18	96	0.34	0.16
Additives with perceived roles of buffer are present with a total concentration between 0.014M and 0.045M ➔ the macromolecule is a deoxy oligonucleotide	42	233	0.60	0.15
Species with ≥ 1 titrateable groups are not present ➔ the macromolecule is a little miscellaneous protein	27	148	0.52	0.15
Species with basic PKAs are not present ➔ the macromolecule is a peptide	76	438	0.94	0.15
Additives with perceived roles of salt are present with a total concentration between 0.01M and 0.05M ➔ the macromolecule is an oligonucleotide	30	180	0.38	0.14
Species with 0 titrateable groups are present with a total concentration > 3.2M and species with titrateable groups ≥ 2 are not present ➔ the macromolecule is a ribonuclease	24	146	0.39	0.14
The macromolecule is a deoxy oligonucleotide ➔ the species spermine is present with a concentration ≤ 0.0008M	10	60	0.59	0.14
Species with ≥ 1 titrateable groups are present with a total concentration ≤ 0.03M ➔ the macromolecule is a deoxy oligonucleotide	32	265	0.46	0.11
Species with net charges of ≤ -2 ares not present ➔ the macromolecule is a very small peptide	77	675	0.95	0.10
Species with 0 titrateable groups are present with a total concentration > 3.2M ➔ the macromolecule is a ribonuclease	27	268	0.44	0.09
Species with net charges of ≤ -2 are not present ➔ the macromolecule is a deoxy oligonucleotide.	65	687	0.93	0.09
Species with basic PKAs are not present ➔ the macromolecule is a miscellaneous little protein	37	477	0.71	0.07
Species with strongly acidic PKAs are not present ➔ the macromolecule is a deoxy oligonucleotide	65	985	0.93	0.06
2-methyl-2/4 pentanediol is not present ➔ the	1836	183	0.92	0.91

species magnesium II is not present				
Additives with perceived roles of salt are not present ➔ the species cacodylate is not present	1183	36	0.57	0.97
Additives with perceived roles of salt are not present ➔ sodium cacodylate is not present	1204	15	0.57	0.99
Additives with perceived roles of salt are not present ➔ the macromolecule is a soluble protein	1038	181	0.60	0.85
Inorganic divalent metals are not present ➔ 2-methyl 2/4 pentanediol is not present	1684	119	0.83	0.93
Inorganic divalent metals are not present ➔ the species cacodylate is not present	1750	53	0.85	0.97
Inorganic divalent metals are not present ➔ sodium cacodylate is not present	1774	29	0.84	0.98
Inorganic halides are not present ➔ sodium cacodylate is not present	1352	25	0.64	0.98
Magnesium chloride is not present ➔ the species cacodylate is not present	1969	90	0.95	0.96
Magnesium chloride is not present ➔ species with net charges of ≥ 3 are not present	1538	521	0.95	0.75
Magnesium chloride is not present ➔ the species spermine is not present	2028	31	0.95	0.98
Metabolite nucleotides are not present ➔ species with two or more polymerization states are not present	2114	45	0.99	0.98
The species cacodylate is not present ➔ spermine is not present	2054	11	0.93	0.99
Sodium cacodylate is not present ➔ magnesium chloride is not present	2008	106	0.98	0.95
Sodium cacodylate is not present ➔ the species magnesium II is not present	1943	171	0.98	0.92
Sodium cacodylate is not present ➔ the species spermine is not present	2084	30	0.97	0.99
Species with ≥ 4 titrateable groups are not present ➔ metabolite nucleotides are not present	1710	27	0.79	0.98
Species with dipole charges are not present ➔ metabolite nucleotides are not present	1660	19	0.77	0.99
Species with somewhat mixed charges are not present ➔ metabolite nucleotides are not present	1810	23	0.84	0.99
Species with net charges of ≤ -1 are not present and species with a neutral PKA are present ➔ the macromolecule is a very small peptide	31	1	0.38	0.97
Species with net charges of ≥ 2 are not present ➔ the species cacodylate is not present	1347	42	0.65	0.97
Species with net charges of ≥ 2 are not present ➔ the macromolecule is a protein	1274	115	0.66	0.92
Species with net charges of ≥ 2 are not present ➔	1347	42	0.65	0.97

the species cacodylate is not present				
Spermine is not present ➔ the species magnesium II is not present	1985	212	1.00	0.90
Spermine tetrahydrochoride is not present ➔ the species magnesium II is not present	1981	213	1.00	0.90
The macromolecule is a protein ➔ 2-methyl 2/4 pentanediol is not present	1788	134	0.89	0.93
The macromolecule is a protein ➔ inorganic divalent species are not present	1660	262	0.92	0.86
The macromolecule is a protein ➔ magnesium chloride is not present	1862	60	0.90	0.97
The macromolecule is a protein ➔ the species cacodylate is not present	1857	65	0.90	0.97
The macromolecule is a protein ➔ species with net charges of ≥ 2 are not present	1274	648	0.92	0.66
The macromolecule is a soluble protein ➔ magnesium chloride is not present	1696	47	0.82	0.97
The macromolecule is a soluble protein ➔ the species cacodylate is not present	1686	57	0.82	0.97
The macromolecule is a soluble protein ➔ species with net charges of ≥ 2 are not present	1172	571	0.84	0.67
The macromolecule is a soluble protein ➔ the species magnesium is not present	1662	81	0.84	0.95
The macromolecule is an enzyme ➔ the species cacodylate is not present	783	26	0.38	0.97
The species cacodylate is not present ➔ magnesium chloride is not present	1969	96	0.96	0.95
The species cacodylate is not present ➔ spermine is not present	2042	23	0.95	0.99
The species cacodylate is not present ➔ the species magnesium II is not present	1907	158	0.96	0.92
The species chloride is not present ➔ sodium cacodylate is not present	1360	25	0.64	0.98
The species magnesium II is not present ➔ 2-methyl 2/4 pentanediol is not present	1836	152	0.91	0.92
The species magnesium II is not present ➔ the species cacodylate is not present	1907	81	0.92	0.96
The species spermine is not present ➔ 2-methyl 2/4 pentanediol is not present	1981	161	0.98	0.92
The species spermine is not present ➔ magnesium chloride is not present	2028	114	0.98	0.95
The species spermine is not present ➔ the species cacodylate is not present	2042	100	0.99	0.95
The species spermine is not present ➔ sodium cacodylate is not present	2084	58	0.99	0.97
The species spermine is present with a	6	4	0.37	0.60

concentration between 0.0025M and 0.006M ➔ sodium cacodylate is present with a concentration between 0.03M and 0.045M				
The species chloride is not present ➔ the macromolecule is a protein	1271	114	0.66	0.92
Inorganic halides are not present ➔ the macromolecule is a protein	1263	114	0.66	0.92

Rules involving heme porphyrin containing proteins

Rule	TP	FP	SEN	PPV
Species with strongly acidic PKAs are present ➔ the macromolecule is a heme porphyrin protein	68	31	0.30	0.69
Species with net charges of ≤ -2 are present ➔ the macromolecule is a heme porphyrin protein	71	73	0.31	0.49
Species with strongly acidic PKAs are present with a total concentration > 0.86M ➔ the macromolecule is a heme porphyrin protein	55	65	0.24	0.46
Species with net charges of ≥ 1 are present ➔ the macromolecule is a heme porphyrin protein	70	121	0.31	0.37
Species with net charges of ≤ -1 are present ➔ the macromolecule is a heme porphyrin protein	72	127	0.31	0.36
The macromolecule is a heme porphyrin protein ➔ species with a basic PKA are present.	78	151	0.39	0.34
Species with ≥ 1 titrateable groups are present ➔ the macromolecule is a heme porphyrin protein	79	179	0.34	0.31
The macromolecule is a heme porphyrin protein ➔ species with net charges of ≤ -1 are present	72	157	0.36	0.31
The macromolecule is a heme porphyrin protein ➔ species with net charges of ≥ 1 are present	70	159	0.37	0.31
The macromolecule is a heme porphyrin protein ➔ species with net charges of ≥ 1 are present with a total concentration of 2.42M	66	163	0.25	0.29
Species with 0 titrateable groups are present ➔ the macromolecule is a heme porphyrin protein	80	229	0.35	0.26
The macromolecule is a heme porphyrin protein ➔ inorganic divalent species are present.	25	204	0.62	0.11
The macromolecule is a heme porphyrin protein ➔ species with net charges of ≥ 2 are present	26	203	0.57	0.11

Miscellaneous Rules between macromolecules and additives

Rule	TP	FP	SEN	PPV
Organic simple hydroxyls are present and the macromolecule is a miscellaneous soluble protein	6	0	0.75	1.00

➔ sodium acetate is present				
The concentration of the macromolecule is greater than 25m and the molecular weight of the macromolecule is ≤ 10,500 and the PH is between 5.8 and 6.6 ➔ ammonium phosphate is present	21	0	0.42	1.00
The macromolecule has a molecular weight of 10,500 and the PH is between 5.8 and 6.6 and an additive with a perceived role of precipitating agent is present ➔ ammonium phosphate is present	25	1	0.50	0.96

Crystallization method versus macromolecule characteristics and concentration

Rule	TP	FP	SEN	PPV
Organic hydroxyls are not present and the molecular weight of the macromolecule is <= 10,500 ➔ the crystallization method is other*	117	177	0.63	0.40
The crystallization method scale is micro ➔ methanol is not present	1337	7	0.62	0.99
The crystallization method scale is micro ➔ simple organic hydroxyls are not present	1280	64	0.63	0.95
The crystallization method description is batch-method and additives with perceived roles of buffer are not present and species with ≥ 3 titrateable groups are present ➔ the species sulfate is present	33	2	0.31	0.94
The crystallization method scale is macro ➔ poly alcohol organic hydroxyls are not present***	380	70	0.28	0.84
The crystallization method is other ➔ species with net charges of ≥1 are not present	111	74	0.34	0.60
Species with basic PKAs are present ➔ the crystallization method type is batch	92	107	0.24	0.46
Species with net charges of ≤ -1 are present ➔ the crystallization method is batch method	91	108	0.25	0.46
Species with net charges of ≥1 are present ➔ the crystallization method is batch method	88	103	0.24	0.46
Species with ≥ 1 titrateable groups are present ➔ the crystallization method type is batch	105	153	0.27	0.41
Species with ≥ 1 titrateable groups are not present ➔ the crystallization method type is other	65	110	0.35	0.37
Species with net charges of 0 are present with a concentration > 234M***??? ➔ the crystallization method type is other	64	123	0.35	0.34
Species with net charges of 0 are not present ➔ the crystallization method type is batch	210	509	0.55	0.29
Species with basic PKAs are not present ➔ the crystallization method type is other	136	378	0.74	0.26
Species with net charges of 0 are present ➔ the crystallization method is concentration-by-evaporation	44	130	0.43	0.25
Species with net charges of ≥ 1 are not present ➔ the crystallization method is concentration by evaporation	75	251	0.73	0.23
The crystallization method type is batch ➔ species with net charges of ≥ 1 are present	89	295	0.47	0.23
Species with net charges of ≤ -1 are not present ➔ the crystallization method is concentration by evaporation	70	267	0.68	0.21
Species with basic PKAs are not present ➔ the crystallization method is crystallization-by-evaporation	90	424	0.87	0.18
Species with net charges of 0 are present with a concentration of 234M***??? ➔ the crystallization method is temperature-crystallization	21	166	0.70	0.11

Crystallization method versus species or additives or their properties

Additive versus Additive

EDTA and dithioethreitol

Rule	TP	FP	SEN	PPV
A strong base is not present ➔ dithiothreitol is not present	727	0	0.34	1.00
A species with a strongly basic PKA is not present ➔ the species dithiothreitol is not present	727	0	0.34	1.00
Species with ≥ 4 titrateable groups are not present ➔ dithiothreitol is not present	1697	40	0.80	0.98
Species with a charge of ≤ -3 are not present ➔ dithiothreitol is not present	1227	28	0.58	0.98
A strong base is not present ➔ EDTA is not present	713	14	0.35	0.98
A species with a strongly basic PKA is not present ➔ the species EDTA is not present	712	15	0.35	0.98
EDTA is not present ➔ dithiothreitol is not present	1998	59	0.94	0.97
Organic chelators are not present ➔ the species dithiothreitol is not present	1987	57	0.94	0.97
Species with a dipole charge are not present ➔ dithiothreitol is not present	1632	47	0.77	0.97
Species with a dipole charge are not present ➔ EDTA is not present	1612	67	0.78	0.96
Species with a somewhat mixed charge are not present ➔ EDTA is not present	1748	85	0.85	0.95
Dithiothreitol is not present ➔ EDTA is not present	1998	123	0.97	0.94
Dithiothreitol is not present ➔ an additive with perceived roles of chelator is not present	1988	133	0.97	0.94

Ammonium phosphate and iron citrate

Rule	TP	FP	SEN	PPV
Ammonium phosphate is not present ➔ iron II citrate is not present.	2171	4	0.99	1.00
The species phosphate is not present ➔ iron II citrate is not present.	1589	0	0.72	1.00
Iron II citrate is not present ➔ ammonium phosphate is not present*	2171	28	1.00	0.99
Iron II citrate is present ➔ ammonium phosphate is present*	22	4	0.44	0.85
Ammonium phosphate is present ➔ iron II citrate is present.	22	28	0.85	0.44
The species phosphate is present ➔ iron II citrate is present.	23	58	0.88	0.28

Additives versus Properties of Additives or of Species

Sodium Azide

Rule	TP	FP	SEN	PPV
Species with strongly acidic PKAs are not present ➔ sodium azide	1050	0	0.53	1.00

is not present				
Species with strongly basic PKAs are not present ➔ sodium azide is not present	727	0	0.37	1.00
Species with basic PKAs are not present ➔ sodium azide is not present	514	0	0.26	1.00
Species with net charges of $\leq$ -3 is not present ➔ sodium azide is not present	1255	0	0.64	1.00
Species with net charges of = 0 is not present ➔ sodium azide is not present	719	0	0.36	1.00
Species with net charges of $\geq$ 3 is not present ➔ sodium azide is not present	1620	0	0.82	1.00
Species with dipole charges are not present ➔ sodium azide is not present	1679	0	0.85	1.00
Species with somewhat mixed charges are not present ➔ sodium azide is not present	1833	0	0.93	1.00
Species with neutral PKAs are is not present ➔ sodium azide is not present	930	0	0.47	1.00
Species with more than 5 titrateable groups are not present ➔ sodium azide is not present	1847	0	0.94	1.00
Sodium azide is not present ➔ species with somewhat mixed charges are not present.	1826	24	1.00	0.99
Sodium azide is not present ➔ additives with perceived roles of chelator are not present.	1862	110	0.91	0.94
Species with many polymerization states are not present ➔ sodium azide is not present	1454	109	0.74	0.93

Ammonium Sulfate

Rule	TP	FP	SEN	PPV
Species with acidic PKAs are present and the macromolecule is a heme porphyrin protein and the molecular weight of the macromolecule is $\leq$ 10,500 ➔ ammonium sulfate is present	28	0	0.28	1.00
Species with strongly basic PKAs are not present ➔ ammonium sulfate is not present	727	0	0.50	1.00
The crystallization method description is batch-method and additives with perceived roles of salt are present and species with $\geq$ 2 titrateable groups are present ➔ ammonium sulfate is present	34	3	0.34	0.92
The crystallization method description is batch-method and additives with perceived roles of buffer are not present and species with strongly acidic PKAs are present ➔ inorganic sulfates are present	36	1	0.34	0.97
The crystallization method description is batch-method and species with neutral PKAs are present and additives with perceived roles of buffer are not present ➔ inorganic sulfates are present	36	1	0.34	0.97
The crystallization method description is batch-method and the molecular weight of the macromolecule is $\leq$ 10,500 and species with neutral PKAs are present ➔ ammonium sulfate is present	28	0	0.28	1.00
The crystallization method description is batch-method and the	26	1	0.25	0.96

molecular weight of the macromolecule is ≤ 10,500 and species with ≥ 2 titrateable groups are present ➔ inorganic sulfates are present				
The crystallization method scale is macro and additives with perceived roles of salt are present and species with ≥ 2 titrateable groups are present ➔ ammonium sulfate is present	35	3	0.35	0.92
The crystallization method scale is macro and additives with perceived roles of buffer are not present and species with ≥ 3 titrateable groups are present ➔ inorganic sulfates are present	36	1	0.34	0.97
The crystallization method type is batch and additives with perceived roles of salt are present and species with neutral PKA are present ➔ ammonium sulfate is present	34	3	0.34	0.92
The crystallization method type is batch-method and additives with perceived roles of buffer are not present and species with ≥ 3 titrateable groups are present ➔ inorganic sulfates are present	33	2	0.31	0.94
The crystallization scale is macro and the molecular weight of the macromolecule is ≤ 10,500 and species with ≥ 2 titrateable groups are present ➔ inorganic sulfates are present	26	1	0.25	0.96
The macromolecule is a heme porhpyrin protein and the molecular weight of the macromolecule is ≤ 10,500 and species with ≥ 2 titrateable groups are present ➔ inorganic sulfates are present	26	0	0.25	1.00
The macromolecule is a heme porphyrin protein and species with neutral PKAs are present and additives with perceived roles of buffer are not present ➔ the species sulfate is present	37	2	0.35	0.95
The macromolecule is a heme porphyrin protein and the molecular weight of the macromolecule is ≤ 10,500 and species with neutral PKAs are present ➔ ammonium sulfate is present	28	0	0.28	1.00
The macromolecule is a heme porphyrin protein and the molecular weight of the macromolecule is ≤ 10,500 and species with neutral PKAs are present ➔ inorganic sulfates are present	28	0	0.27	1.00

Properties of additives or of species versus other properties of additives or of species

Versus perceived role of anti microbial

Rule	TP	FP	SEN	PPV
Species with net charges of 0 are not present ➔ additives with perceived roles of anti microbial are not present	719	0	0.37	1.00
Species with ≥ 5 titrateable groups are not present ➔ additives with perceived roles of anti microbial are not present	1831	16	0.94	0.99
Species with ≥ 4 titrateable groups are not present ➔ additives with perceived roles of anti microbial are not present	1722	15	0.88	0.99
Species with ≥ 3 titrateable groups are not present ➔ additives with perceived roles of anti microbial are not present	1063	8	0.54	0.99
Species with ≥ 2 titrateable groups are not present ➔ additives with perceived roles of anti microbial are not present	966	8	0.50	0.99
Species with strongly basic PKAs are not present ➔ additives with	719	8	0.37	0.99

perceived roles of anti microbial are not present				
Species with strongly acidic PKAs are not present ➔ additives with perceived roles of anti microbial are not present	1042	8	0.53	0.99
Species with somewhat mixed charges are not present ➔ additive with perceived roles of anti microbial are not present	1816	17	0.93	0.99
Species with neutral PKAs are not present ➔ additives with perceived roles of anti microbial are not present	922	8	0.47	0.99
Species with dipole charges are not present ➔ additives with perceived roles of anti microbial are not present	1662	17	0.85	0.99
Species with net charges of ≥ 3 are not present ➔ additives with perceived roles of anti microbial are not present	1603	17	0.82	0.99
Species with net charges of ≥ 2 are not present ➔ additives with perceived roles of anti microbial are not present	1374	15	0.70	0.99
Species with net charges of ≤ -2 are not present ➔ additives with perceived roles of anti microbial are not present	745	7	0.38	0.99
Species with net charges of ≤ -3 are not present ➔ additives with perceived roles of anti microbial are not present	1247	8	0.64	0.99
Species with basic PKAs are not present ➔ additives with perceived roles of anti microbial are not present	506	8	0.26	0.98
Species with many polymerization states are not present ➔ additives with perceived roles of anti microbial are not present	1436	127	0.74	0.92

Versus perceived role of substrate analogue

Rule	TP	FP	SEN	PPV
Species with acidic PKAs are not present and species with dipole charges are not present ➔ additives with perceived roles of substrate analogue are not present	740	2	0.37	1.00
Species with ≥ 2 titrateable groups are not present ➔ additives with perceived roles of substrate analogue are not present	955	19	0.47	0.98
Species with net charges of ≤ -2 are not present ➔ additives with perceived roles of substrate analogue are not present	730	22	0.36	0.97
Species with an acidic PKA are not present ➔ additives with perceived roles of substrate analogue are not present**	928	24	0.46	0.97
Metabolite nucleotides are not present and species with neutral PKAs are not present ➔ additives with perceived roles of substrate analogue are not present	884	37	0.44	0.96
Species with dipole charges are not present ➔ additives with perceived roles of substrate analogue are not present	1605	74	0.80	0.96
Species with ≥ 4 titrateable groups are not present ➔ additives with perceived roles of substrate analogue are not present	1649	88	0.82	0.95
Species with ≥ 3 titrateable groups are not present ➔ additives with perceived roles of substrate analogue are not present	1022	49	0.51	0.95
Species with somewhat mixed charges are not present ➔ additives with perceived roles of substrate analogue are not present	1742	91	0.87	0.95
Species with net charges of ≤ -3 are not present ➔ additives with	1193	62	0.59	0.95

perceived roles of substrate analogue are not present				
Species with ≥ 5 titrateable groups are not present ➔ additives with perceived roles of substrate analogue are not present	1740	105	0.86	0.94
Species with net charges of ≥ 3 are not present ➔ additives with perceived roles of substrate analogue are not present	1519	101	0.75	0.94
Species with net charges of ≥ 2 are not present ➔ additives with perceived roles of substrate analogue are not present	1309	80	0.65	0.94
Species with ≥ polymerization states are not present ➔ additives with perceived roles of substrate analogue are not present	1981	165	0.98	0.92
Species with an acidic PKA are present ➔ additives with perceived roles of substrate analogue are present**	20	137	0.61	0.13

Versus perceived roles of precipitating agents

Rule	TP	FP	SEN	PPV
Species with dipole charges are not present and additives with unknown perceived roles are not present ➔ organic species are not present	1622	0	0.74	1.00
Species with ≥ 3 titrateable groups are present with a concentration > 0.5M ➔ additives with perceived roles of precipitating agent are not present	124	12	0.28	0.91
Species with strongly acidic PKAs are present with a concentration > 0.86M ➔ additives with perceived roles of precipitating agent are not present	109	11	0.24	0.91
Species with acidic PKAs are not present ➔ additives with perceived roles of precipitating agent are not present*	178	45	0.40	0.80
Species with acidic PKAs are present ➔ additives with perceived roles of precipitating agent are present*	55	102	0.27	0.35

Misc

Rule	TP	FP	SEN	PPV
Species with ≥ 5 titrateable groups are not present ➔ Species with somewhat mixed charges are not present	1821	24	0.99	0.99
Species with dipole charges are not present ➔ Species with somewhat mixed charges are not present	1679	0	0.92	1.00
Metabolite amino acids & their analogues are not present and species with a strongly acidic PKA are not present ➔ Species with somewhat mixed charges are not present	1031	10	0.56	0.99
Species with no titrateable groups are not present ➔ Species with somewhat mixed charges are not present	968	6	0.53	0.99

Global properties

Rule	TP	FP	SEN	PPV
Temperature is between 4 and 20 ➔ inorganic “other” species are not present	716	2	0.33	1.00
Species with acidic PKAs are present with a concentration of ≤	4	0	0.33	1.00

0.205M ***???*** and ionic strength is ≤ 0.025 ➔ additives with unknown perceived roles are present with a concentration of 0.3M				
Ionic strength is ≤ 0.025 and species with a neutral PKA are present ➔ the macromolecule is a very short peptide	31	1	0.38	0.97
The PH of the solution is between 5.8 and 6.6 and additives with perceived roles of precipitating agent are present and the space groups description of the resulting crystal is p2<1> ➔ the species phosphate is present	25	1	0.31	0.96
Ionic strength is ≤ 0.025 and species with ≥ 1 titrateable groups are present ➔ the macromolecule is a very short peptide	33	2	0.41	0.94
The macromolecule is a small protein ➔ ionic strength is ≤ 0.025	114	137	0.32	0.45
The macromolecule is a heme porphyrin protein ➔ ionic strength is > 5.975	90	139	0.28	0.39
Ionic strength is ≤ 0.025 ➔ the macromolecule is a very small miscellaneous protein	114	246	0.45	0.32
Ionic strength is > 5.975 ➔ the macromolecule is a heme porphyrin protein	90	235	0.39	0.28
The macromolecule is an enzyme ➔ ionic strength is between 2.207 and 5.975	171	638	0.53	0.21
Ionic strength is ≤ 0.025 ➔ the macromolecule is a very short peptide	65	295	0.80	0.18
The diffraction limit of the crystal is ≥ 3 and species with strongly acidic PKAs are not present ➔ the macromolecule is a tRNA	19	106	0.58	0.15
Ionic strength is between 0.025 and 0.157 ➔ the macromolecule is a deoxy oligonucleotide	41	248	0.59	0.14
The diffraction limit of the crystal is ≤ 2 and species with strongly acidic PKAs are not present ➔ the macromolecule is a deoxy oligonucleotide	35	207	0.50	0.14
Species with a net charge of > 3 are not present and the diffraction limit of the crystal is ≥ 3 ➔ the macromolecule is a tRNA	19	131	0.58	0.13
Species with a net charge of ≤ -3 are not present and the diffraction limit of the crystal is ≤ 2 ➔ the macromolecule is a deoxy oligonucleotide	35	260	0.50	0.12