Supporting Information for

Seeing is Believing: Experimental Spin States from Machine Learning Model Structure Predictions

Michael G. Taylor[1,#], Tzuhsiung Yang[1,#], Sean Lin[1,#], Aditya Nandy[1,2], Jon Paul Janet[1], Chenru Duan[1,2], and Heather J. Kulik[1,*]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

[2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

[#]These authors contributed equally.

Contents

Table S1. Characteristics of spectrochemical series homoleptic complexes that were obtained from the CSD UO set for Fe(II) and Fe(III) complexes: the bond length (in Å) of the CSD complex, the name of the complex assigned during deposit, the nominal ligand, the molecular weight used to search for the complex, and the CSD ID of the complex are all shown. In several cases, no available experimental complex could be found, as indicated by --.

	Fe(II)				Fe(III)		
	MW	Name	CSD ID	Value	Name	CSD ID	Value
H2O	163.935	hexakis(aqua)-iron(ii)	AMAVOB	2.12	hexaaqua-iron(iii)	MEWZAP	2.00
acac	356.196	--	--	--	--	--	--
NCS-	404.313	--	--	--	--	--	--
NH3	158.031	hexaammine-iron(ii)	CACDIW	2.23	--	--	--
phen	596.475	tris(1,10-Phenanthroline)-iron(ii)	AFIRAJ	1.98	tris(1,10-Phenanthroline-N,N')-iron(iii)	BIPGEN	1.98
CN-	211.953	hexacyano-iron(ii)	BAMBAV	1.91	hexacyano-iron(iii)	XESMEQ	1.93
CO	223.905	--	--	--	--	--	--

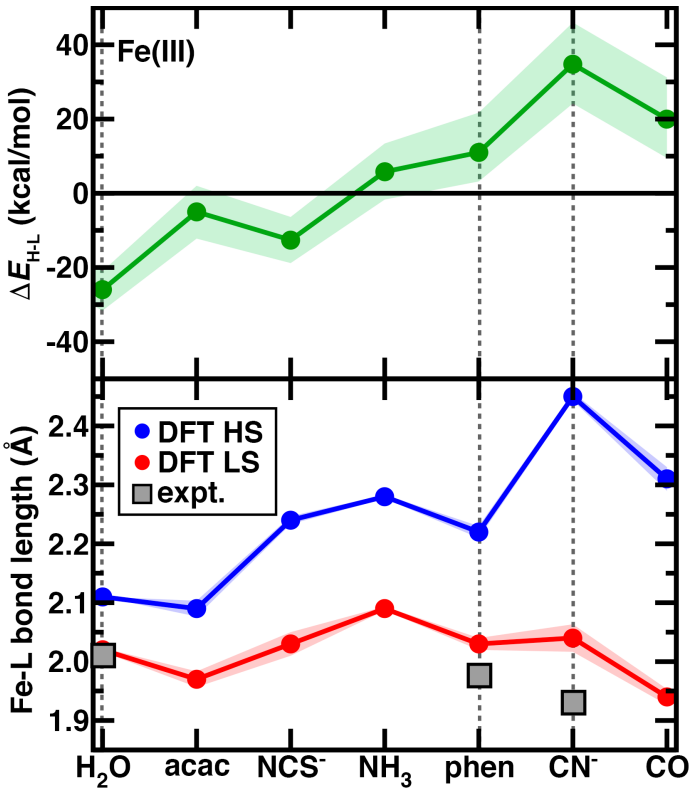

Figure S1. Properties of homoleptic, octahedral mononuclear Fe(III) transition metal complexes with ligands ordered by their field strength in the spectrochemical series. Properties shown are evaluated with hybrid DFT (B3LYP, a_{HF} = 0.2, sold lines and circle symbols) along with the range of properties evaluated at a_{HF} = 0.1-0.3 shown as a translucent shaded region. The high-spin to low-spin adiabatic spin splitting energy ($\Delta E_{H\text{-}L}$, in kcal/mol) is shown at top, and the Fe–L bond lengths (in Å) for the high spin (HS) and low spin (LS) states are shown at bottom. Representative Fe–L bond lengths from crystal structures are shown as gray squares as indicated in inset legend, and vertical dotted lines are shown to enable comparison of bond-length-derived spin-state assignment (bottom) and energetic assignment (top).

Table S2. Hyperparameters and details of the RAC-155/ANN bond length model. The RAC-155 ANN hyperparameters were selected with Hyperopt. The hidden ANN topology is the number of nodes in each layer (i.e., 3 hidden layers for the RAC-155 ANN).

Hyperparameter	**RAC-155 Bond length ANN**
Hidden ANN topology	300x300x200
Activation function	tanh
Output layer activation function	linear
Learning rate	0.002
Optimizer	adam
Batch size	64
Epochs	1200
L2 regularization	1e-15
Dropout % (all hidden nodes)	5
Training data	Spin splitting data as per Ref.[1] and Redox data set as per Ref.[2]
Hyperparameter optimization	Hyperopt
Data partitioning	80% train (10% used in Hyperopt for validation) / 20% test

Text S1. Details of equatorial plane assignment.

The following general rules were used to resolve the equatorial planes in a consistent, repeatable, and physically informed manner accounting for as many as possible combinations of different ligand identity and denticity. First, if the highest denticity ligand in a structure lay entirely in a defined plane (e.g. tetradentate ligand with two *trans*-monodentate ligands), the highest denticity ligand plane was assigned as the equatorial plane. If the highest denticity ligand did not lie entirely in a single plane (e.g. tetradentate ligand with two *cis*-monodentate ligands), but an equatorial plane could be unambiguously assigned by any set of the lower denticity ligands, the equatorial plane was assigned by this unambiguous plane (e.g. the two *cis*-mondentate ligands and the tetradentate metal-connecting opposite them). For cases where the lowest denticity ligands did not unambiguously form a plane (e.g. three *fac*-monodentate ligands with a tridentate ligand), the plane with the highest molecular weight was assigned the equatorial plane or, for pentadentate ligands, the four contacts of the pentadentate ligand that form a plane were assigned as the equatorial plane. For hexadentate ligands, the plane that contains the metal-connecting atoms with the highest total molecular weight was assigned as the equatorial plane. For complexes with only monodentate ligands, identical ligands were treated like a single ligand of the denticity equal to the number of each type of identical ligand and were treated following identical rules as above. For cases where monodentate ligands formed binding conformations not physical with higher denticities (e.g. three pairs of *trans*-monodentates, where triple bidentate cases can only form equivalent "*cis*" conformations) the plane with the highest molecular weight was assigned as the equatorial plane. If, following all of these rules, an identified equatorial plane did not contain metal-connecting atoms of the same elements, the equatorial plane was re-assigned to a plane containing only metal-connecting atoms of the same element.

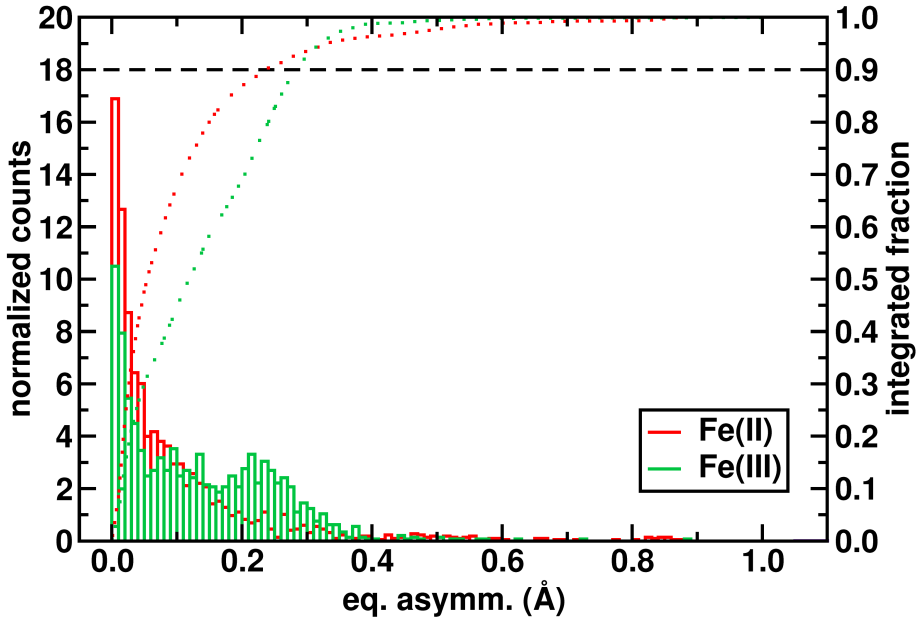

Figure S2. Degree of distortion (i.e., max-min distances) for an equatorial plane selected to minimize the overall equatorial distortion for all 3,627 unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

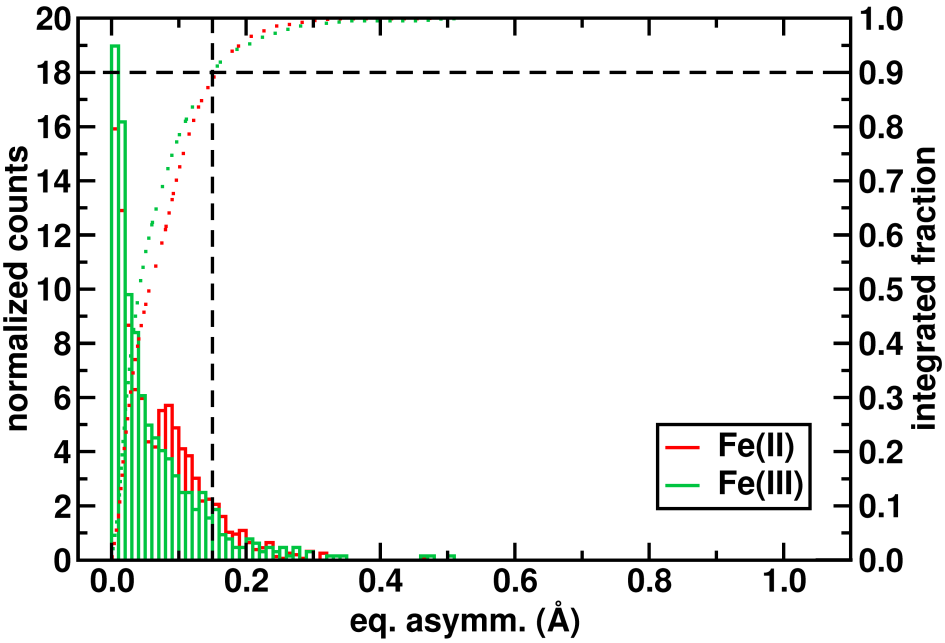

Figure S3. Degree of distortion (i.e., max-min distances) in the equatorial plane for 2,201 complexes with equatorial symmetry (i.e., same elements in the equatorial plane), unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

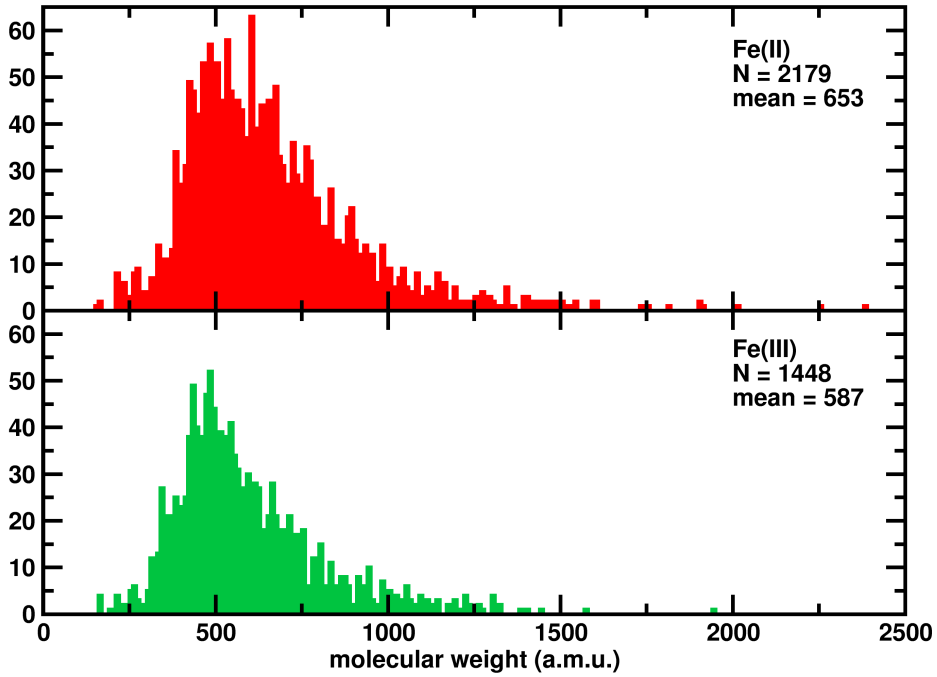

Figure S4. Molecular weight distribution of all 3,627 unique mononuclear Fe octahedral complexes with Fe(II) shown at top and Fe(III) shown at bottom. The number of data points as well as the mean molecular weight of the data set is shown in inset.

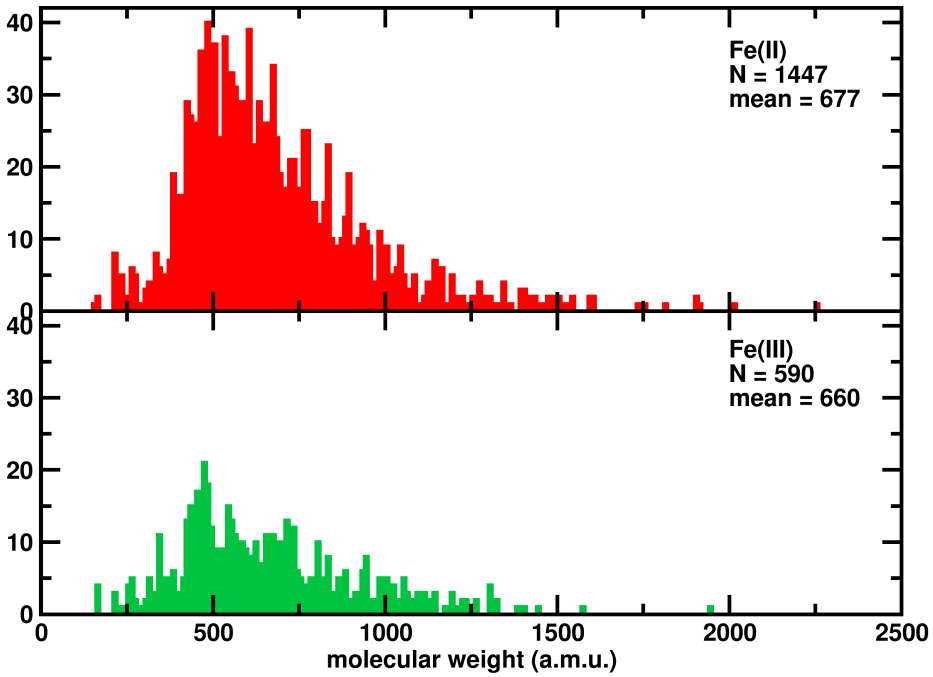

Figure S5. Molecular weight distribution of all 2,037 unique mononuclear Fe octahedral complexes that are ANN compatible with distortion outliers removed. Fe(II) is shown at top and Fe(III) shown at bottom. The number of data points as well as the mean molecular weight of the data set is shown in inset.

Table S3. Overall number of element types and frequency of each occurrence observed in 3,627 unique mononuclear iron octahedral complexes (2,179 Fe(II) and 1,448 Fe(III)). Variations are listed in increasing number of distinct element types, from 1 to a maximum of 5 (for Fe(II) only).

Fe(II)		Fe(III)		Fe(II)		Fe(III)		Fe(II)		Fe(III)		Fe(II)		Fe(III)	
One ligand elem type				**Two ligand elem types (cont'd)**				**Three ligand elem types (cont'd)**				**Four ligand elem types**			
NNNNNN	964	NNNNNN	184	IIPPPP	2			NNNNOS	2			CClClNPP	6	ClNNOOS	4
OOOOOO	48	OOOOOO	135	HHHPPP	2			CCCIPP	2			CHNPPP	4	ClClNOOS	3
CCCCCC	17	SSSSSS	29	HPPPPP	2			ClNNPPP	2			BrCCNNS	3	ClClNNOS	2
SSSSSS	7	CCCCCC	6	OOOOPP	2			CCClClPP	1			CCClNPP	2	BrNNOOS	1
PPPPPP	2	FFFFFF	1	FFOOOO	1			BrCCCNN	1			CHHNPP	2		
ClClClClClCl	1	ClClClClClCl	1	ClClClClOO	1			CCNNNP	1			CClNPPP	2		
AsAsAsAsAsAs	1			CCCCCS	1			CCCCIS	1			CNNPSS	2		
Two ligand elem types				BrBrBrCCC	1			CCCNSS	1			ClNPPP	2		
NNNNOO	304	NNNNOO	284	ClClClClNN	1			CCNNOO	1			CClNNPP	2		
NNOOOO	95	NNOOOO	115	ClClSSSS	1			NOSSSS	1			CCNOPP	1		
ClClNNNN	75	NNNOOO	105	CCSiSiSiSi	1			AsCCCHH	1			ClNNNOS	1		
NNNNNO	41	NNNNSS	68	CSSSSS	1			CCPPSeSe	1			CHNNPP	1		
ClNNNNN	31	ClClClNNN	60	CCCIII	1			CCCNNO	1			HNNPPSi	1		
NNNNPP	31	ClClNNNN	58	CCCCClCl	1			NNNOSS	1			BrCCNPP	1		
CCNNNN	27	NNNNNO	38	CCCCCP	1			CCPSSS	1			BrNNPSS	1		
NNPPPP	26	CCNNNN	21	BrBrOOOO	1			CNSSSS	1			BrNNOSS	1		
NNNOOO	24	CCCNNN	20	CCCCCCl	1			CNNNNS	1			BrCCOPP	1		
CNNNNN	23	NOOOOO	18	AsAsNNNN	1			BrBrNNNO	1			BrBrCCNP	1		
NNNNSS	22	ClOOOOO	9	AsAsAsAsII	1			CCCClNN	1			ClNNOOS	1		
CCPPPP	17	ClClOOOO	8	NNNNNP	1			ClNPPPP	1			CCCIPS	1		
ClClPPPP	17	CCCCNN	8	CCCCOO	1			CISSSS	1			CCCHPSi	1		
NNNPPP	12	CNNNNN	8	**Three ligand elem types**				CFPPPP	1			CCClPPS	1		
NNSSSS	11	CCCCCN	7	HNPPPP	14	ClNNNOO	32	BrBrCCCTe	1			CCCOPSn	1		
BrBrNNNN	11	ClNNNNN	6	ClClNNOO	11	ClClNNNO	29	HOPPPP	1			**Five ligand elem types**			
CCCCNN	9	BrBrNNNN	6	ClClNNNO	8	ClClClNNO	21	AsAsCCCP	1			CHNOPP	1		
OOPPPP	8	ClClClClNN	5	CHPPPP	8	NNOOSS	19	ClPPPP	1			CHNPPSi	1		
NOOOOO	7	NNNSSS	5	ClNNNNO	8	ClClNNOO	18	HPPPPSe	1			CClNOP	1		
BrNNNNN	7	ClClPPPP	5	CCNNSS	7	ClNNOOO	18	FNNNNS	1						
CCSSSS	7	OOOSSS	5	NNOOSS	6	ClNNNNO	9	CCPPPS	1						
CCCCCN	6	NNNNPP	4	ClNNNOO	6	NNNOOS	8	CCNNPP	1						
CCCCSS	6	NNSSSS	3	CCCNPP	5	ClClNOOO	5	CCCSeSeTe	1						
CCCCII	6	BrBrBrNNN	3	CClPPPP	5	CNNNOO	5	CCCSeTeTe	1						
BrBrPPPP	6	FFNNNN	2	CNNNNO	4	CNNNNO	4	CCCIISe	1						
CCCNNN	5	PPSSSS	2	CCPPSS	4	ClNOOOO	3	ClClNNPP	1						
CCCSSS	4	FFFFOO	1	CCNPPP	4	ClClClNNS	3	HNNPPP	1						
OOSSSS	4	FFFFFO	1	HPPPPS	4	BrNNNOO	3	CPPSSS	1						
ClPPPPP	4	ClClClOOO	1	NNPPSS	4	CCNNOO	2	ClNNNPP	1						
NNNNNS	3	ClClClClOO	1	ClHPPPP	4	NNNNOS	2	CCHHPP	1						
CCCCPP	3	ClClClClClO	1	CNPPPP	4	ClClClNOO	1	BrCPPPP	1						
PPSSSS	3	FFFNNN	1	CCCIITe	4	FNNNNO	1	CCCHNN	1						
NNNNSeSe	3	BrOOOOO	1	ClClNNSS	3	ClClNNNS	1	BrBrCCCN	1						
CCCCSnSn	3	NNNNNS	1	NPSSSS	3	ClClOOSS	1	CCCCHN	1						
BrCCCCC	2	ClClClSSS	1	CNNNPP	3	CNPPPP	1	CCIIPP	1						
OOOOSS	2	OOOSeSeSe	1	COPPPP	3	CCNNSS	1	CClClPPP	1						
ClClClNNN	2	FNNNNN	1	CNNSSS	2	BrBrNNNO	1	NNOOPP	1						
ClClOOOO	2	NNNNNP	1	HHHNPP	2	CNNNSS	1	CCNNSiSi	1						
NNNSSS	2	NNNNSeSe	1	CCCNNS	2	BrNNOOO	1	CCClPPP	1						
CCCCSiSi	2	OOOOSS	1	ClNNNNS	2	NNNOSS	1	NOOPPP	1						
CCCPPP	2	AsAsAsAsII	1	CNNNSS	2	NNOOPP	1	CCCClPP	1						
lNNNNN	2	OOOOSeSe	1	CPSSSS	2			CCCPPSn	1						
IISSSS	2	BrBrOOOO	1	HNNNPP	2										
PPPPSS	2			CCOOPP	2										

Table S4. Combinations of equatorial ligand element types and frequency of each occurrence observed after assignment of equatorial plane in 3,627 unique mononuclear iron octahedral complexes (2,179 Fe(II) and 1,448 Fe(III)). The types of equatorial ligand elements are grouped by the total unique number of elements and listed in increasing order, where applicable. Only the one-element type subset is ANN-compatible.

Fe(III)		Fe(III)		Fe(III)		Fe(III)	
One ligand element type				**Two ligand element types (cont'd)**			
N	1264	N	372	IPPP	1		
O	127	O	214	CCCH	1		
P	95	S	32	OOSS	1		
C	38	C	13	IIPP	1		
S	29	P	5	BrBrNN	1		
Cl	3	F	3	CCSiSi	1		
As	2	Cl	3	CSiSiSi	1		
		As	1	PPPSe	1		
Two ligand element types				CCCSn	1		
NNOO	124	NNNO	197	**Three ligand element types**			
NNNO	124	NNOO	176	CNPP	5	ClClNO	19
NNPP	46	ClClNN	108	ClNNO	5	NNOS	17
ClNNN	44	NOOO	71	CPPS	4	ClNNO	15
ClClNN	25	NNSS	62	CNNS	4	CNNO	6
CCNN	21	ClNNN	51	ClNPP	4	ClNOO	3
CPPP	18	CCNN	17	CClTe	4	BrNNO	2
NOOO	14	CCCN	11	NPSS	2	CCNS	1
NNSS	14	NNNS	7	ClClNO	2		
NPPP	13	ClClClN	7	NPPSi	2		
CCCN	10	OOSS	6	CCSeTe	2		
OOPP	9	ClClOO	6	CCNS	1		
CNNN	8	CNNN	3	CNNO	1		
NNNS	8	ClOOO	2	CPSS	1		
CCPP	8	BrBrBrN	2	CCOP	1		
CCCS	7	FFNN	1	CCNO	1		
CSSS	6	NSSS	1	CCPS	1		
ClPPP	6	CCOO	1	NNOS	1		
CCSS	5	OSSS	1	CNNP	1		
CCCl	5	CPPP	1	ClClNS	1		
ClClPP	5	PPSS	1	CPPSe	1		
PSSS	3	PSSS	1	BrNNO	1		
CCCP	3	OOOS	1	BrCCN	1		
ClCPP	3	OOPP	1	AsCCP	1		
BrNNN	3	ClClClO	1	CClSe	1		
NNSeSe	3	ClClSS	1	**Four ligand element types**			
NSSS	2	BrNNN	1	CNOP	1		
HHHN	2	OOSeSe	1	BrNOS	1		
CCHH	2	BrBrNN	1				
PPPS	2	NNSeSe	1				
PPSS	2	OOOSe	1				
CCOO	2						
HHPP	2						
CCCO	2						
BrBrCC	2						
ClClSS	2						
BrPPP	2						
OPPP	1						
CCIl	1						
ISSS	1						
FNNN	1						
IISS	1						

Table S5. Combinations of axial ligands and frequency of each occurrence observed after assignment of equatorial plane in 3,627 unique mononuclear iron octahedral complexes (2,179 Fe(II) and 1,448 Fe(III)). Symmetric axial ligand connecting atoms are listed first followed by two-ligand combinations.

Fe(II)		Fe(III)	
One ligand element type			
N	1192	N	380
O	197	O	251
C	51	Cl	53
Cl	47	S	37
P	39	C	24
S	37	Br	6
Cl	17	P	4
Br	13	F	3
I	4	I	1
Si	2		
H	2		
As	2		
Sn	2		
F	1		
Two ligand element types			
NO	197	NO	414
ClN	82	ClN	91
CN	70	ClO	60
CS	34	CN	44
NP	21	OS	39
HN	20	NS	14
NS	19	BrN	7
CP	17	CO	5
BrN	12	ClS	5
HP	12	FO	2
CCl	11	FN	2
ClO	10	NP	2
CO	9	OSe	2
BrC	8	PS	1
CH	8	BrO	1
ClP	6		
BrP	5		
OS	4		
ClH	4		
HS	3		
IP	3		
AsC	2		
CSe	2		
HO	2		
IN	2		
PS	2		
PSn	2		
CTe	2		
CSi	1		
CF	1		
PSi	1		
CSn	1		

Table S6. Overall combination of equatorial and axial ligand element types and frequency of each occurrence observed after assignment of equatorial plane in 3,627 unique mononuclear iron octahedral complexes (2,179 Fe(II) and 1,448 Fe(III)) with the 2,201 that are ANN-compatible indicated.

Fe(II)																		Fe(III)								
ANN compatible						ANN incompatible												ANN compat			ANN Incompatible					
As	I	1	P	ClH	4	AsCCP	AsC	1	CCSS	CS	4	CPSS	CS	1	NPSS	CN	2	As	I	1	BrBrBrN	N	2	NNNO	ClO	8
As	As	1	P	ClP	4	BrBrCC	BrC	1	CCSS	S	1	CSiSiSi	CSi	1	NSSS	CN	2	C	CN	7	BrBrNN	BrN	1	NNNO	O	8
C	C	17	P	Br	4	BrBrCC	CTe	1	ClClNN	N	22	CSSS	CS	6	OOPP	P	7	C	C	6	BrNNN	BrN	1	NNNO	OS	7
C	CN	6	P	HS	3	BrBrNN	N	1	ClClNN	ClN	1	FNNN	NS	1	OOPP	N	1	Cl	O	1	BrNNO	BrN	1	NNNO	NS	1
C	BrC	2	P	CCl	2	BrCCN	BrP	1	ClClNN	O	1	HHHN	P	2	OOPP	NP	1	Cl	ClO	1	BrNNO	OS	1	NNNO	ClN	1
C	N	2	P	CO	2	BrNNN	BrN	3	ClClNN	P	1	HHPP	HP	2	OOSS	N	1	Cl	Cl	1	CCCN	CN	8	NNNS	NS	6
C	P	2	P	CN	2	BrNNO	BrN	1	ClClNO	NO	1	IIPP	C	1	OPPP	CP	1	F	O	1	CCCN	N	3	NNNS	S	1
C	Sn	2	P	P	2	BrNOS	NS	1	ClClNO	N	1	IISS	S	1	PPPS	HP	1	F	FO	1	CCNN	CN	17	NNOO	NO	87
C	CS	1	P	HP	2	BrPPP	BrP	2	ClClNS	NS	1	IPPP	IP	1	PPPS	PS	1	F	F	1	CCNS	NS	1	NNOO	N	39
C	Si	1	P	O	1	CCCH	PSi	1	ClClPP	CN	5	ISSS	CS	1	PPPSe	HP	1	N	N	184	CCOO	N	1	NNOO	ClN	19
C	Cl	1	P	ClN	1	CCCl	Cl	5	ClClSS	N	2	NNNO	NO	116	PPSS	S	1	N	O	73	ClClClN	N	5	NNOO	O	15
C	CP	1	P	CF	1	CCCN	CN	7	ClCPP	CCl	1	NNNO	ClN	4	PPSS	CS	1	N	NO	38	ClClClN	NO	2	NNOO	ClO	11
C	CCl	1	P	HO	1	CCCN	P	1	ClCPP	CN	1	NNNO	S	1	PSSS	CS	2	N	C	18	ClClClO	O	1	NNOO	BrN	3
C	HN	1	P	Cl	1	CCCN	HN	1	ClCPP	CS	1	NNNO	Cl	1	PSSS	PS	1	N	Cl	16	ClClNN	ClN	32	NNOO	NS	1
C	I	1	P	S	1	CCCN	Br	1	ClNNN	ClN	35	NNNO	O	1				N	CN	8	ClClNN	N	24	NNOO	Cl	1
Cl	O	1	P	I	1	CCCO	PSn	1	ClNNN	ClO	5	NNNO	NS	1				N	ClN	6	ClClNN	NO	20	NNOS	OS	17
Cl	Cl	1	P	BrC	1	CCCO	CO	1	ClNNN	NS	2	NNNS	NS	8				N	S	5	ClClNN	ClO	17	NNSeSe	N	1
Cl	N	1	S	N	9	CCCP	CP	1	ClNNN	OS	1	NNOO	N	68				N	Br	5	ClClNN	O	5	NNSS	N	57
N	N	964	S	S	7	CCCP	IP	1	ClNNN	Cl	1	NNOO	O	20				N	ClO	4	ClClNN	Cl	5	NNSS	NS	3
N	O	120	S	O	4	CCCP	PSn	1	ClNNO	ClO	2	NNOO	NO	20				N	CO	4	ClClNN	ClS	3	NNSS	CN	1
N	NO	41	S	NP	3	CCCS	CS	6	ClNNO	ClN	1	NNOO	Cl	7				N	P	4	ClClNN	NS	1	NNSS	NO	1
N	ClN	31	S	Cl	1	CCCS	Cl	1	ClNNO	N	1	NNOO	ClN	6				N	F	2	ClClNN	OS	1	NOOO	NO	54
N	CN	23	S	NO	1	CCCSn	CSn	1	ClNNO	OS	1	NNOO	S	3				N	FO	1	ClClNO	NO	10	NOOO	N	10
N	Cl	18	S	CN	1	CCHH	AsC	1	ClNPP	CN	2	NNOS	OS	1				N	NS	1	ClClNO	O	3	NOOO	ClN	6
N	C	18	S	CS	1	CCHH	P	1	ClNPP	C	1	NNPP	N	19				N	FN	1	ClClNO	OS	3	NOOO	BrN	1
N	P	12	S	P	1	CCII	Cl	1	ClNPP	CCl	1	NNPP	NP	12				N	NP	1	ClClNO	ClN	2	NSSS	N	1
N	S	7	S	I	1	CClSe	Cl	1	ClPPP	CP	3	NNPP	P	5				N	OS	1	ClClNO	N	1	OOOS	OS	1
N	BrN	7				CClTe	Cl	4	ClPPP	ClP	2	NNPP	S	4				O	O	135	ClClOO	O	5	OOOSe	OSe	1
N	Br	7				CCNN	S	6	ClPPP	CCl	1	NNPP	HN	2				O	N	46	ClClOO	S	1	OOPP	N	1
N	CO	4				CCNN	CN	5	CNNN	CN	6	NNPP	CN	2				O	NO	18	ClClSS	ClS	1	OOSeSe	OSe	1
N	ClO	3				CCNN	N	3	CNNN	P	1	NNPP	CH	1				O	ClO	9	ClNNN	Cl	23	OOSS	OS	4
N	NS	3				CCNN	CS	2	CNNN	S	1	NNPP	ClN	1				O	ClN	3	ClNNN	ClN	18	OOSS	N	2
N	IN	2				CCNN	BrC	1	CNNO	CO	1	NNSeSe	N	3				O	BrO	1	ClNNN	ClO	4	OSSS	O	1
N	CS	1				CCNN	NP	1	CNNP	CP	1	NNSS	N	7				O	Cl	1	ClNNN	NO	4	PPSS	S	1
N	As	1				CCNN	CO	1	CNNS	BrC	3	NNSS	NS	2				O	Br	1	ClNNN	O	2	PSSS	PS	1
N	NP	1				CCNN	CCl	1	CNNS	CS	1	NNSS	S	2				P	Cl	5	ClNNO	ClN	4	N		
N	OS	1				CCNN	Si	1	CNOP	CP	1	NNSS	O	1				S	S	29	ClNNO	OS	4	N		
O	N	64				CCNO	IP	1	CNPP	CP	2	NNSS	CN	1				S	N	3	ClNNO	NO	3	N		
O	O	48				CCNS	CS	1	CNPP	H	2	NNSS	BrP	1							ClNNO	ClO	2	N		
O	NO	7				CCOO	P	2	CNPP	HO	1	NOOO	NO	11							ClNNO	ClS	1	N		
O	S	2				CCOP	BrP	1	CPPP	CP	6	NOOO	N	3							ClNNO	O	1	N		
O	Cl	2				CCPP	CN	4	CPPP	HN	4	NPPP	HP	4							ClNOO	ClO	2	N		
O	P	2				CCPP	NP	1	CPPP	C	2	NPPP	CCl	2							ClNOO	Cl	1	N		
O	F	1				CCPP	BrN	1	CPPP	HP	2	NPPP	Cl	2							ClOOO	ClO	2	N		
O	Br	1				CCPP	Cl	1	CPPP	CN	1	NPPP	ClN	2							CNNN	CN	3	N		
P	N	20				CCPP	CCl	1	CPPP	NP	1	NPPP	CP	1							CNNO	NO	5	N		
P	Cl	15				CCPS	Cl	1	CPPP	CS	1	NPPP	HN	1							CNNO	CO	1	N		
P	C	11				CCSeTe	CSe	1	CPPP	CCl	1	NPPP	NP	1							CPPP	NP	1	N		
P	HN	10				CCSeTe	CTe	1	CPPS	CS	4	NPPSi	CH	1							FFNN	FN	1	N		
P	CH	6				CCSiSi	C	1	CPPSe	CSe	1	NPPSi	HN	1							NNNO	NO	172	N		

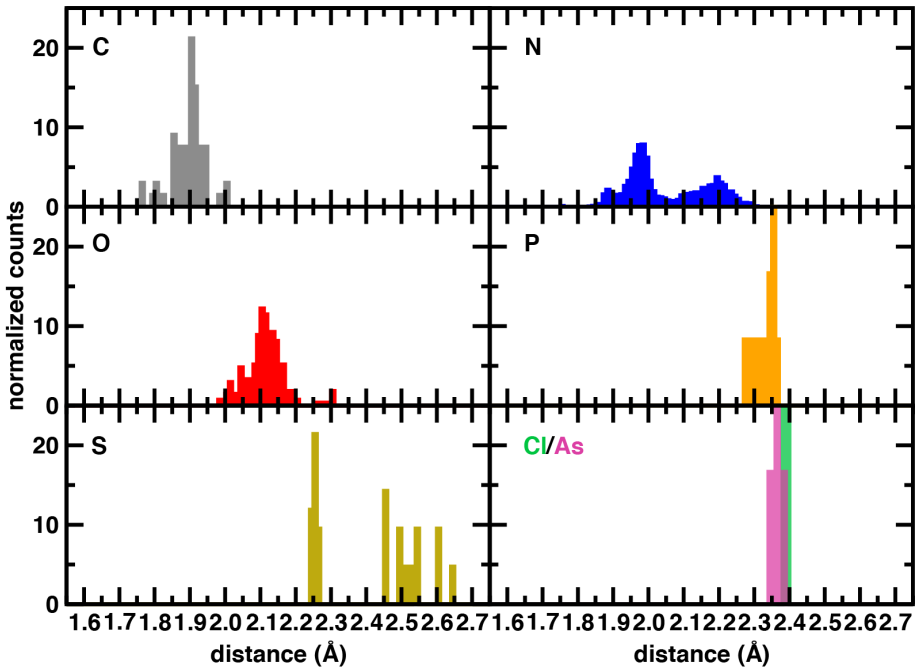

Figure S6. Normalized histograms of bond distances for 969 unique Fe(II) complexes with only a single element identity for the coordinating atom in the complex from the set of 1,447 complexes that have distortion outliers removed. All bond distances are shown in Å for the elements indicated in each inset.

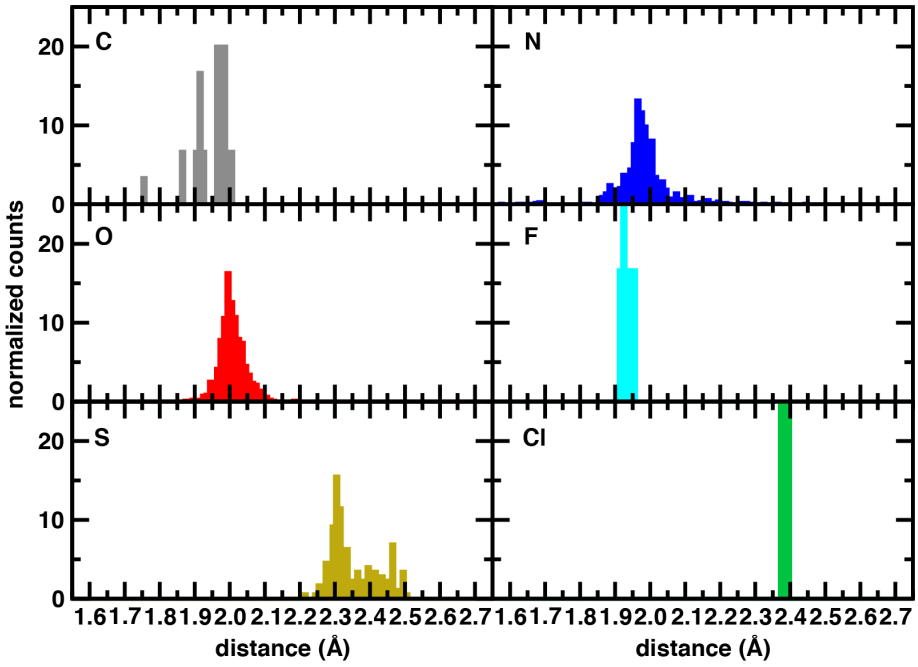

Figure S7. Normalized histograms of bond distances for 347 unique Fe(III) complexes with only a single element identity for the coordinating atom in the complex from the set of 590 complexes that have distortion outliers removed. All bond distances are shown in Å for the elements indicated in each inset.

Table S7. Covalent radii and sums of covalent radii used for definitions of relative distances of Fe–X bonds (in Å) as well as the low-spin and high-spin bounds expressed as both a distance and relative to the average covalent radii. The average Fe covalent radius of 1.42 Å is used and a low spin (LS) estimate of 1.32 Å and high spin (HS) estimate of 1.52 Å defines the low spin and high spin bounds. To draw HS and LS borders on relative bond distances, we select 0.95 and 1.05 based on typical fractions for different elements.

Fe-X	X radius (Å)	$r_{Fe}+r_X$			LS frac	HS frac
		Avg. (Å)	LS (Å)	HS (Å)		
Fe-As	1.19	2.61	2.51	2.71	0.96	1.04
Fe-Br	1.20	2.62	2.52	2.72	0.96	1.04
Fe-C	0.73	2.15	2.05	2.25	0.95	1.05
Fe-Cl	1.02	2.44	2.34	2.54	0.96	1.04
Fe-F	0.57	1.99	1.89	2.09	0.95	1.05
Fe-H	0.31	1.73	1.63	1.83	0.94	1.06
Fe-I	1.39	2.81	2.71	2.91	0.96	1.04
Fe-N	0.71	2.13	2.03	2.23	0.95	1.05
Fe-O	0.66	2.08	1.98	2.18	0.95	1.05
Fe-P	1.07	2.49	2.39	2.59	0.96	1.04
Fe-S	1.05	2.47	2.37	2.57	0.96	1.04
Fe-Se	1.20	2.62	2.52	2.72	0.96	1.04
Fe-Si	1.11	2.53	2.43	2.63	0.96	1.04
Fe-Sn	1.39	2.81	2.71	2.91	0.96	1.04
Fe-Te	1.38	2.80	2.70	2.90	0.96	1.04

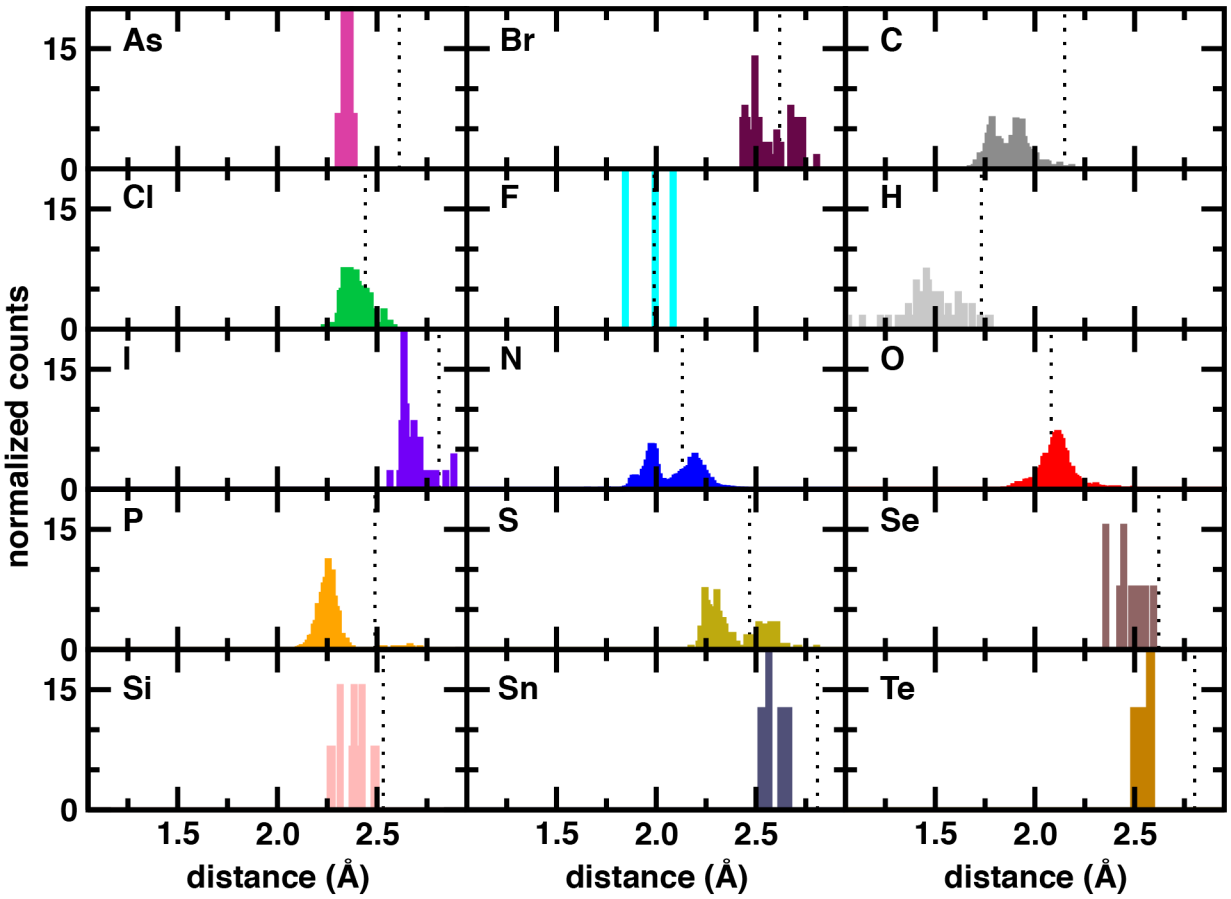

Figure S8. Normalized histograms of bond distances (in Å) for all unique Fe(II) complexes. The x-range of all graphs is fixed to 1.05 Å to 2.95 Å and the y-range between 0 and 20, causing truncation for some elements with narrow histograms. The averaged sum of covalent radii is indicated as a dotted line for each element. All bond distances are collected over all complexes in the data set, and data is shown for any element that occurs at least once as a coordinating atom in any complex in the data set.

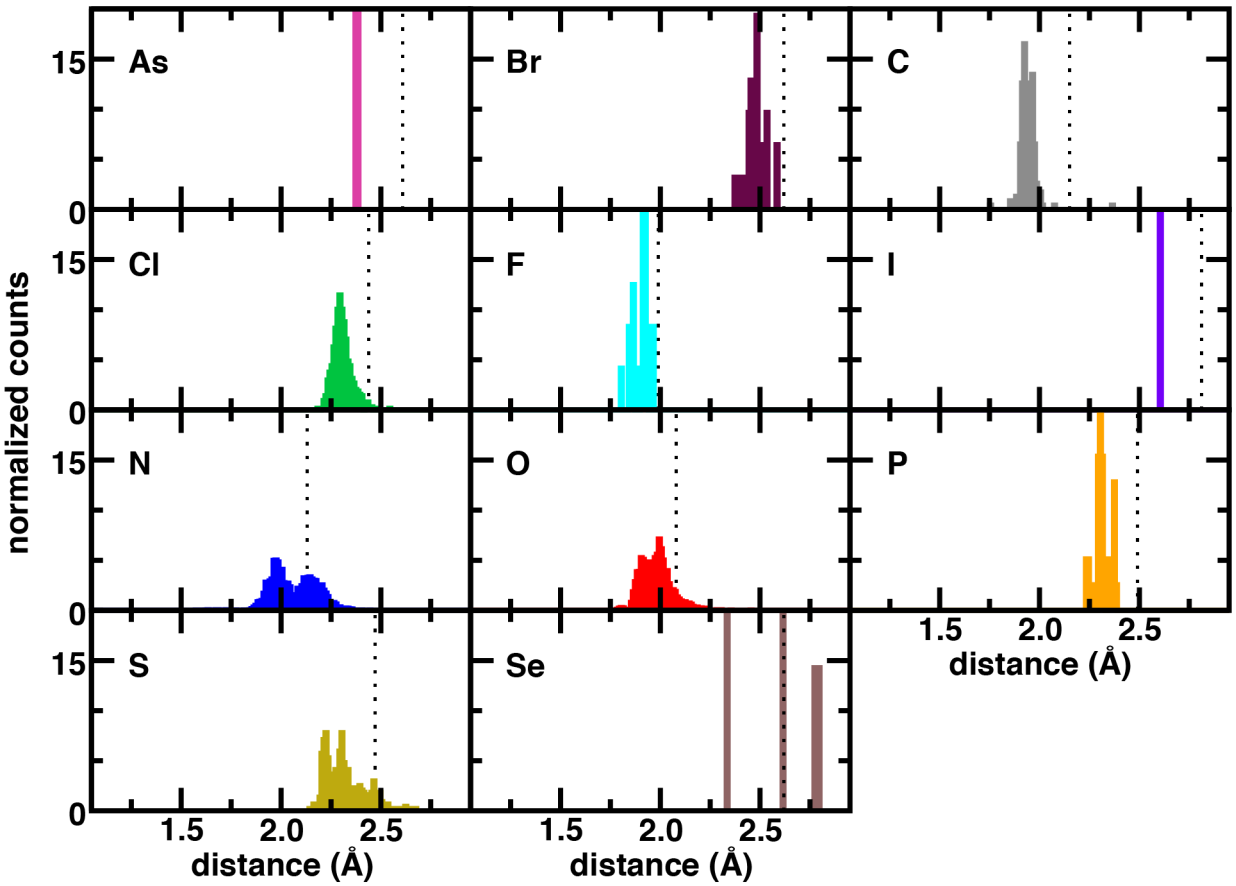

Figure S9. Normalized histograms of bond distances (in Å) for all unique Fe(III) complexes. The x-range of all graphs is fixed to 1.05 Å to 2.95 Å and the y-range between 0 and 20, causing truncation for some elements with narrow histograms. The averaged sum of covalent is indicated as a dotted line for each element. All bond distances are collected over all complexes in the data set, and data is shown for any element that occurs at least once as a coordinating atom in any complex in the data set.

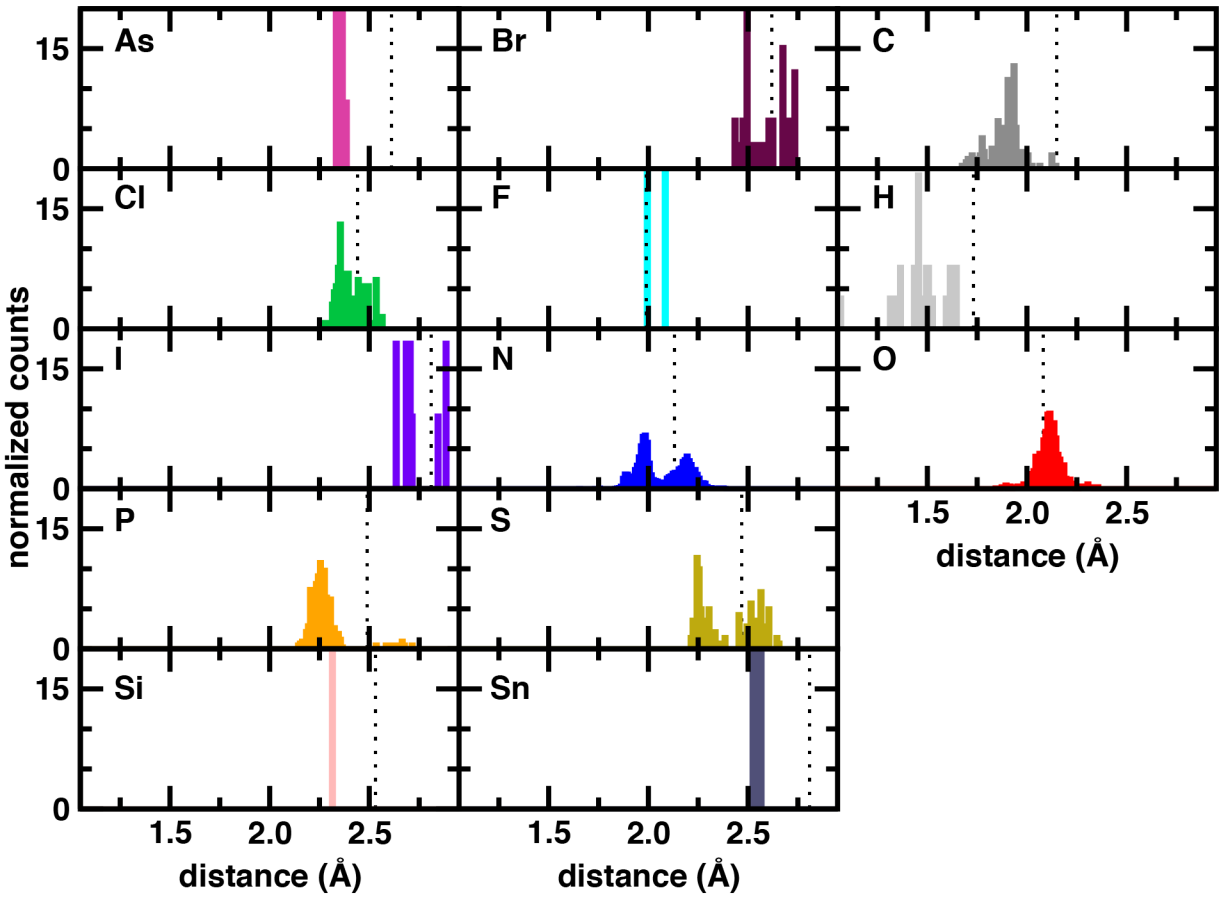

Figure S10. Normalized histograms of bond distances (in Å) for all ANN-compatible, low distortion, unique Fe(II) complexes. The x-range of all graphs is fixed to 1.05 Å to 2.95 Å and the y-range between 0 and 20, causing truncation for some elements with narrow histograms. The averaged sum of covalent radii is indicated as a dotted line for each element. All bond distances are collected over all complexes in the data set, and data is shown for any element that occurs at least once as a coordinating atom in any complex in the data set.

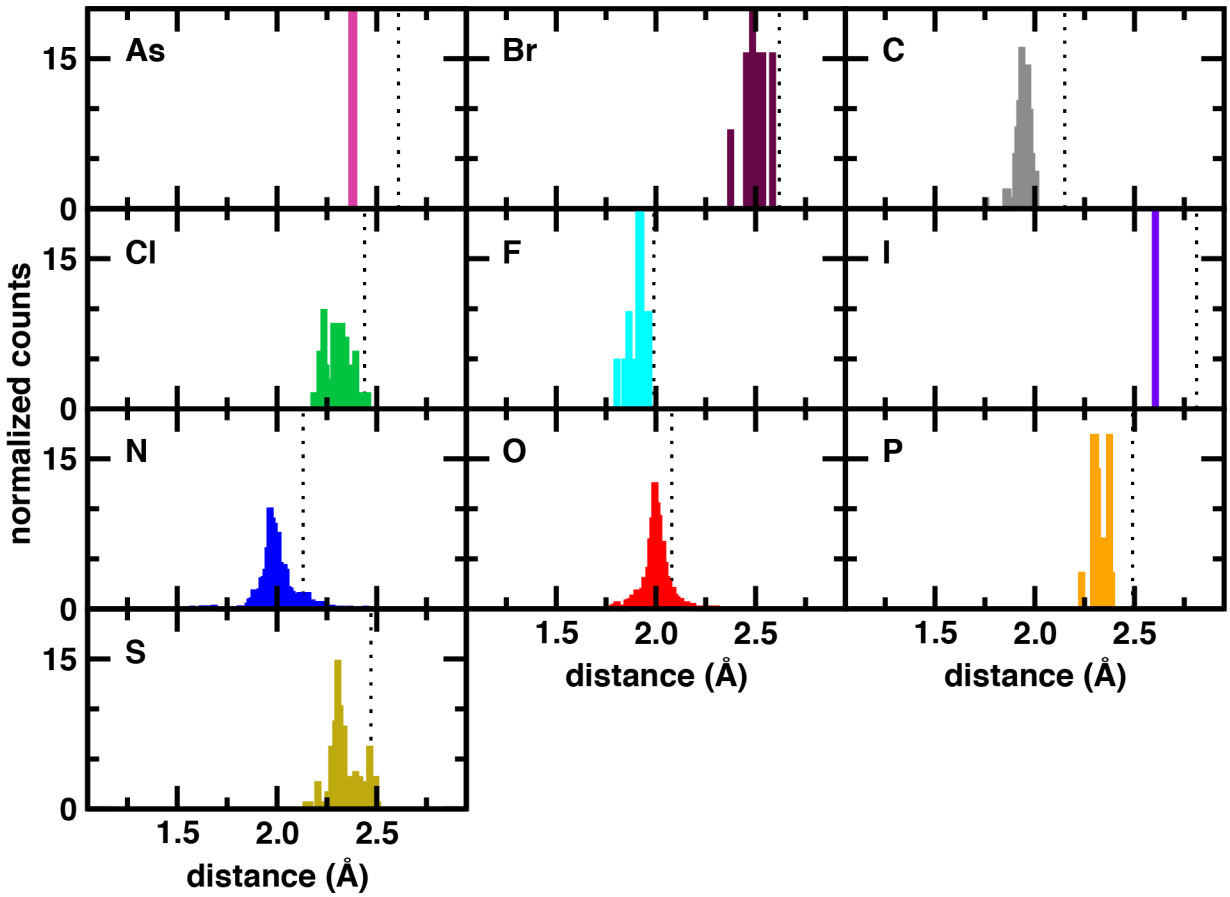

Figure S11. Normalized histograms of bond distances (in Å) for all ANN-compatible, low distortion, unique Fe(III) complexes. The x-range of all graphs is fixed to 1.05 Å to 2.95 Å and the y-range between 0 and 20, causing truncation for some elements with narrow histograms. The averaged sum of covalent radii is indicated as a dotted line for each element. All bond distances are collected over all complexes in the data set, and data is shown for any element that occurs at least once as a coordinating atom in any complex in the data set.

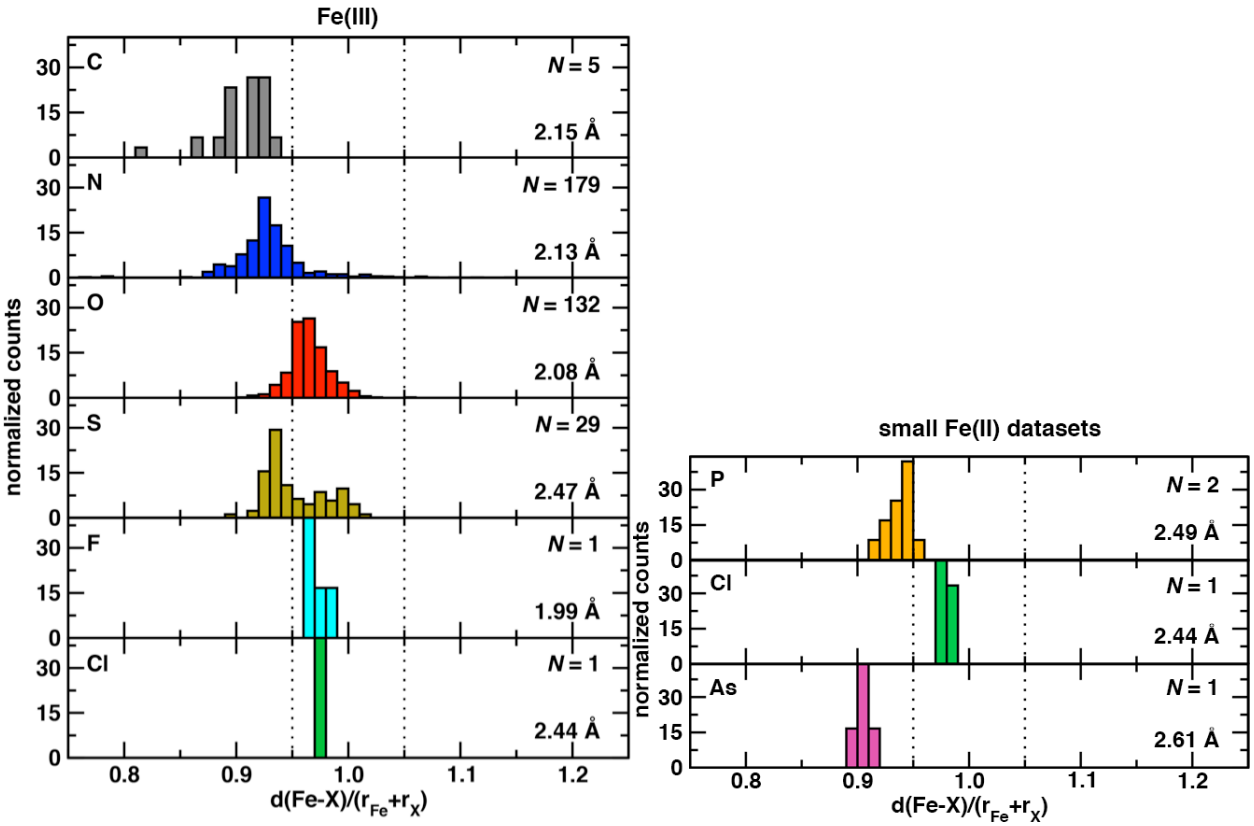

Figure S12. Normalized histograms of relative metal-ligand bond lengths for 347 mononuclear octahedral Fe(III) complexes (left) and P, Cl, or As Fe(II) complexes (right) in which all coordinating atoms are a single element, which is indicated in the upper left corner of each panel. The 347 Fe(III) complexes are obtained from the set of 590 complexes that have distortion outliers removed. Each relative metal-ligand bond length is obtained with respect to the sum of covalent radii of Fe and the ligand atom, and the value for each element is indicated in the bottom right corner of each panel. The total number of complexes used to compute each histogram is indicated in the top right corner of each panel, and each of the six bond lengths in the complex are used to construct the histogram. Vertical dotted lines indicate 0.95 and 1.05 relative metal-ligand bond length thresholds to nominally indicate low-spin or high-spin character, respectively. The 4 Fe(II) complexes are obtained from the Fe(II) dataset. Since there are only 1 or 2 complexes in these cases, it is difficult to reach conclusions about trends in bond lengths.

Table S8. Overall combination of equatorial and axial ligand element types and frequency of each occurrence observed after assignment of equatorial plane in 2,037 ANN-compatible, low-distortion, unique mononuclear iron octahedral complexes (1,447 Fe(II) and 590 Fe(III)). They are sorted by decreasing elemental symmetry – with homoelemental cases at top, followed by two ligand element types, followed by three ligand element types.

Fe(II)								Fe(III)			
eq.	ax.	num.	# typs.	eq.	ax.	num.	# typs.	eq.	ax.	num.	# typs.
N	N	902	1	P	Cl	1	2	N	N	179	1
O	O	45	1	P	I	1	2	O	O	132	1
C	C	11	1	S	P	1	2	S	S	29	1
S	S	7	1	S	I	1	2	C	C	5	1
P	P	2	1	S	CS	1	2	Cl	Cl	1	1
As	As	1	1	S	Cl	1	2	F	F	1	1
Cl	Cl	1	1	P	HN	10	3	N	O	70	2
N	O	111	2	P	CH	6	3	N	NO	29	2
O	N	57	2	N	CO	4	3	O	N	28	2
N	NO	36	2	P	ClH	4	3	N	C	18	2
N	ClN	29	2	P	HS	3	3	N	Cl	13	2
N	CN	23	2	S	NP	3	3	O	NO	10	2
P	N	20	2	N	ClO	2	3	N	CN	8	2
N	Cl	18	2	P	CCl	2	3	C	CN	7	2
N	C	16	2	P	CO	2	3	N	ClN	6	2
P	Cl	15	2	P	CN	2	3	O	ClO	6	2
N	P	12	2	N	OS	1	3	N	S	5	2
P	C	11	2	N	CS	1	3	N	Br	5	2
N	S	7	2	P	CF	1	3	P	Cl	5	2
N	Br	7	2	P	BrC	1	3	N	P	4	2
N	BrN	7	2	P	HO	1	3	S	N	3	2
O	NO	6	2	P	ClN	1	3	N	F	2	2
C	CN	5	2	S	CN	1	3	As	I	1	2
S	N	5	2					Cl	O	1	2
P	Br	4	2					Cl	ClO	1	2
S	O	4	2					F	O	1	2
N	NS	3	2					F	FO	1	2
P	ClP	3	2					N	NP	1	2
C	Sn	2	2					N	FN	1	2
C	N	2	2					N	NS	1	2
N	IN	2	2					O	Cl	1	2
O	Cl	2	2					O	BrO	1	2
O	S	2	2					O	Br	1	2
P	HP	2	2					N	ClO	4	3
As	I	1	2					N	CO	4	3
C	Cl	1	2					O	ClN	3	3
C	Si	1	2					N	OS	1	3
C	I	1	2					N	FO	1	3
C	CCl	1	2								
C	BrC	1	2								
C	CS	1	2								
Cl	N	1	2								
Cl	O	1	2								
N	As	1	2								
N	NP	1	2								
O	Br	1	2								
O	F	1	2								
O	P	1	2								
P	S	1	2								
P	O	1	2								

Table S9. Matrix of coordinating atom element pairs (or single element type on the diagonal) present in 1,979 of the ANN-compatible complex subset for Fe(II) (1,402 complexes, top) and Fe(III) (577 complexes, bottom). From the 2,037 ANN-compatible, low-distortion unique mononuclear iron octahedral complexes, there are 45 cases (25 involving nitrogen) with 3 element types for Fe(II) and 13 with 3 element types (all involving nitrogen) for Fe(III) which are combinations not used in bond pair analysis. N-containing pairs are bolded.

Fe(II)													
	C	N	O	Si	P	S	As	Sn	H	F	Cl	Br	I
C	11	**46**	0	1	11	2	0	2	0	0	2	1	1
N	**46**	**902**	**210**	**0**	**33**	**15**	**1**	**0**	**0**	**0**	**48**	**14**	**2**
O	0	**210**	45	0	2	6	0	0	0	1	3	1	0
Si	1	**0**	0	0	0	0	0	0	0	0	0	0	0
P	11	**33**	2	0	2	2	0	0	2	0	19	4	1
S	2	**15**	6	0	2	7	0	0	0	0	1	0	1
As	0	**1**	0	0	0	0	1	0	0	0	0	0	1
Sn	2	**0**	0	0	0	0	0	0	0	0	0	0	0
H	0	**0**	0	0	2	0	0	0	0	0	0	0	0
F	0	**0**	1	0	0	0	0	0	0	0	0	0	0
Cl	2	**48**	3	0	19	1	0	0	0	0	1	0	0
Br	1	**14**	1	0	4	0	0	0	0	0	0	0	0
I	1	**2**	0	0	1	1	1	0	0	0	0	0	0
Fe(III)													
	C	N	O	Si	P	S	As	Sn	H	F	Cl	Br	I
C	5	**33**	0	0	0	0	0		0	0	0	0	0
N	**33**	**179**	**137**	**0**	**5**	**9**	**0**	**0**	**0**	**3**	**19**	**5**	**0**
O	0	**137**	132	0	0	0	0	0	0	2	9	2	0
Si	0	**0**	0	0	0	0	0	0	0	0	0	0	0
P	0	**5**	0	0	0	0	0	0	0	0	5	0	0
S	0	**9**	0	0	0	29	0	0	0	0	0	0	0
As	0	**0**	0	0	0	0	0	0	0	0	0	0	1
Sn	0	**0**	0	0	0	0	0	0	0	0	0	0	1
H	0	**0**	0	0	0	0	0	0	0	0	0	0	0
F	0	**3**	2	0	0	0	0	0	0	1	0	0	0
Cl	0	**19**	9	0	5	0	0	0	0	0	1	0	0
Br	0	**5**	2	0	0	0	0	0	0	0	0	0	0
I	0	**0**	0	0	0	0	1	0	0	0	0	0	0

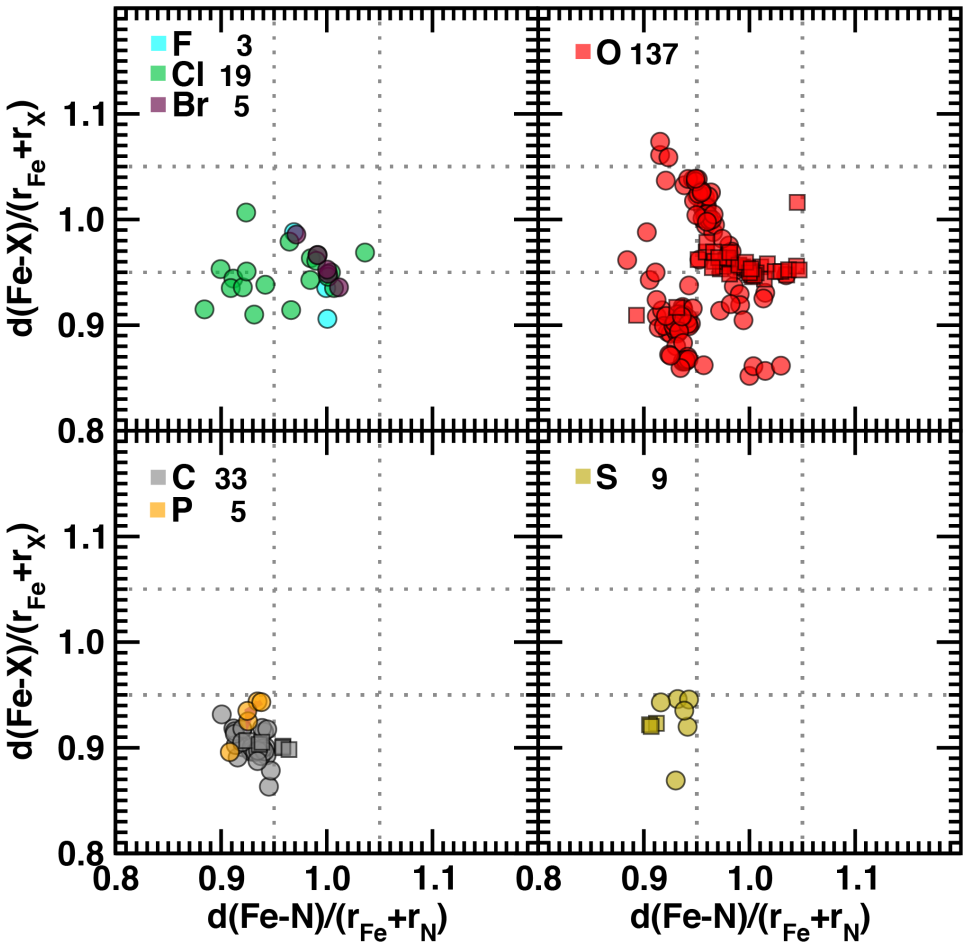

Figure S13. Fe–N vs. Fe–X (X indicated according to inset legend) bond length ratios computed relative to sums of covalent radii for mononuclear octahedral Fe(III) complexes: F, Cl, or Br halides (top left), O (top right), C or P (bottom left), and S (bottom right). Ratios of 0.95 and 1.05 are indicated by grey dashed lines. Circle symbols indicate cases where N is the majority coordinating element (i.e., equatorial plane and up to one of the axial positions), whereas square symbols reflect the reverse case. The Fe–N, Fe–X pair is computed from the average of all bonds of that type in the complex. Each element is colored with translucent fill to show high density regions according to each panel's inset legend, and the total number of cases is indicated in the legend.

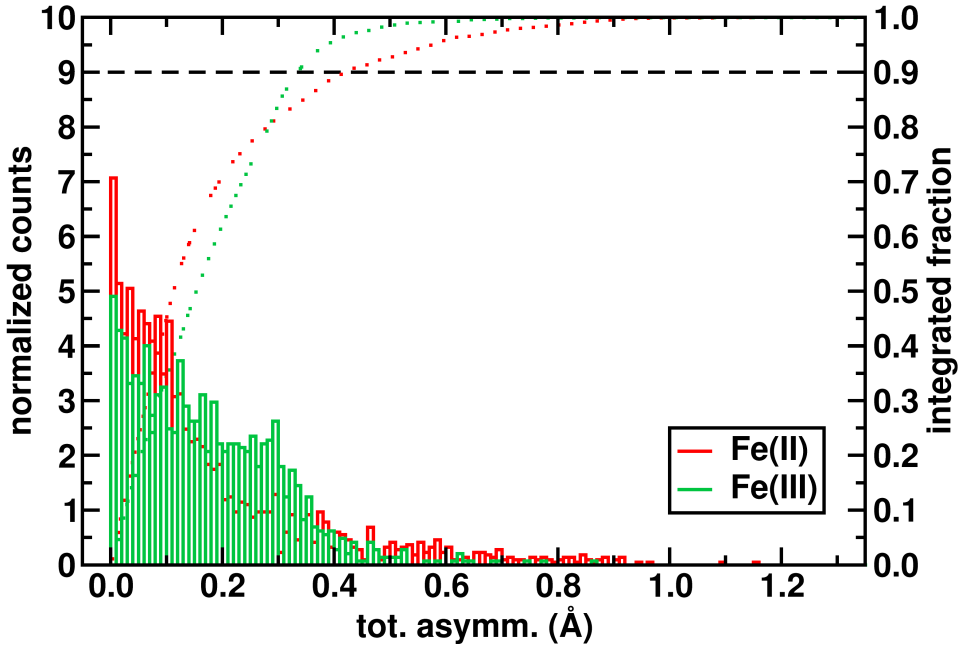

Figure S14. Degree of distortion (i.e., max-min distances) overall for all unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

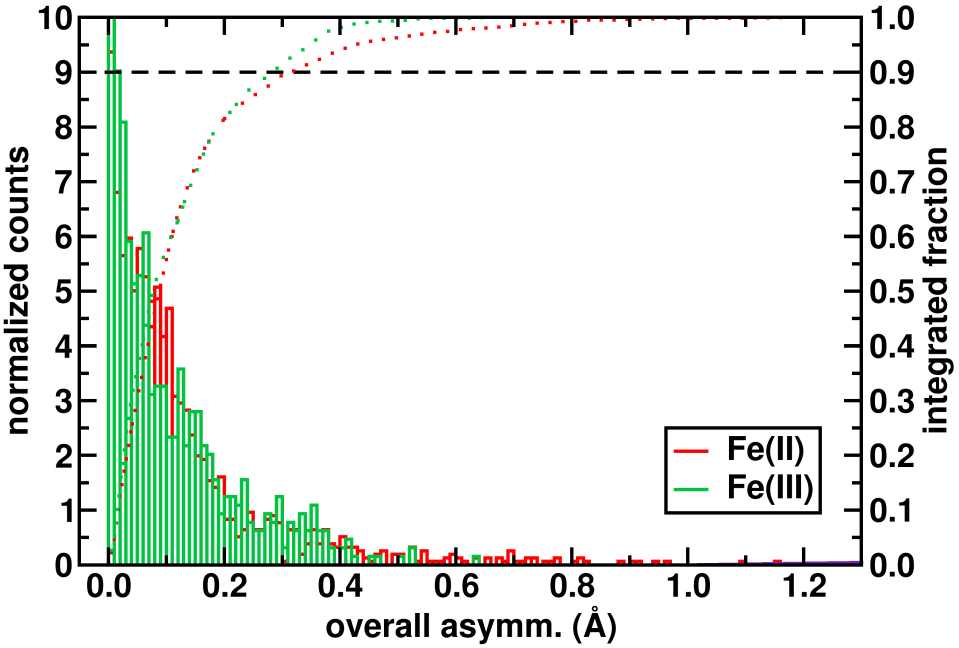

Figure S15. Degree of distortion (i.e., max-min distances) overall for ANN-compatible (i.e., same elements in the equatorial plane), unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

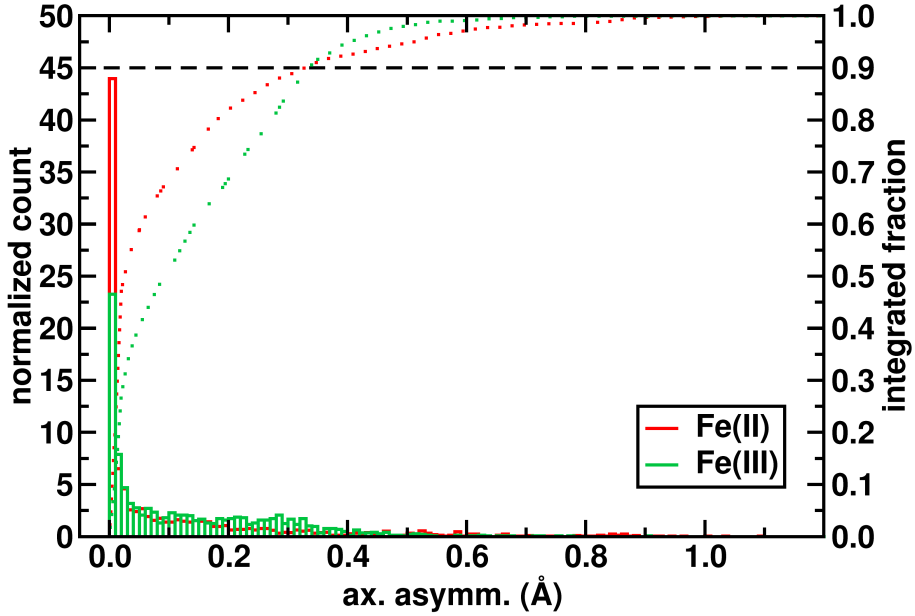

Figure S16. Degree of distortion (i.e., max-min distances) between axial metal–ligand distances based on equatorial plane selected to minimize the overall equatorial distortion for all 3,627 unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

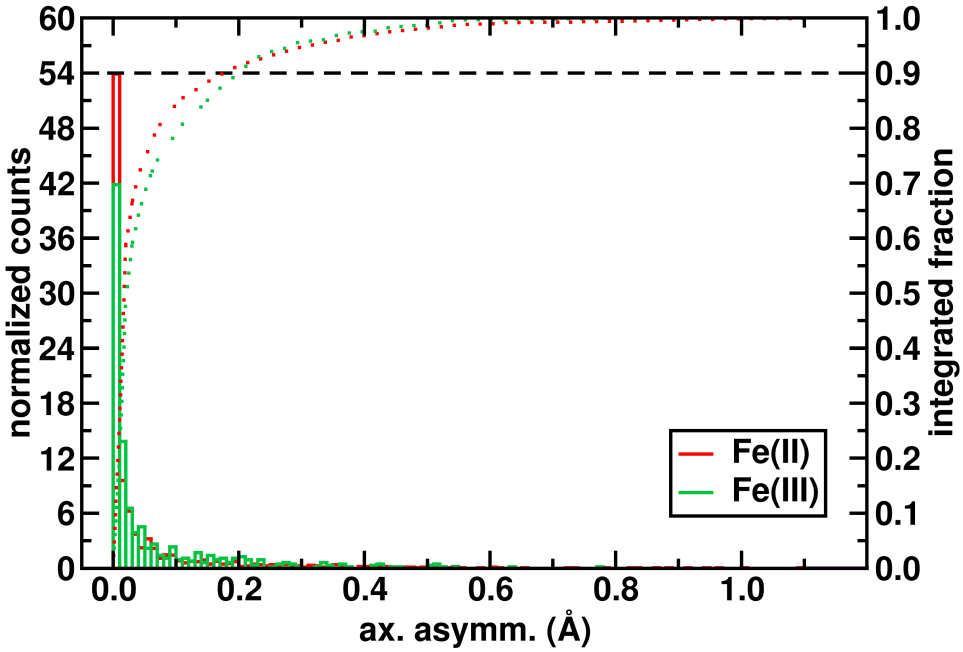

Figure S17. Degree of distortion (i.e., max-min distances) between axial metal–ligand distances for ANN-compatible (i.e., same elements in the equatorial plane), unique Fe(II) (shown in red) or Fe(III) (shown in green) complexes. A normalized histogram with 0.01 Å bins is shown with axis at left as well as an integrated fraction shown as a dotted line with the same coloring, axis shown at right. A black dashed horizontal line indicates where the fraction reaches 0.9.

Table S10. Results of spin state classification of the entire AC set: cases of each spin state classification type for HS and LS states along with total number are indicated.

Detailed category	Fe(II)	Fe(III)	Composite category	Fe(II)	Fe(III)
LS below, good Uncertainty	207	101	LS definite	288	180
LS between, good uncertainty	81	79			
LS below, bad uncertainty	490	178	LS lean	490	178
LS between, bad uncertainty	49	70			
Ambiguous	22	11	ambiguous	150	129
HS between, bad uncertainty	79	48			
HS above, bad uncertainty	200	28	HS lean	200	28
HS between, good uncertainty	180	67			
HS above, good uncertainty	139	8	HS definite	319	75
Total	**1447**	**590**		**1447**	**590**

Table S11. Number of overall HS or LS complexes classified in the HE set grouped by oxidation state (Fe(II) or Fe(III)) and ligand coordinating element, where relevant (P and As are shown only for Fe(II), whereas F is shown only for Fe(III)), as well as the subset for which the uncertainty score is less than or equal to 0.5 ('high confidence').

	Fe(II) overall						
	N	C	O	S	Cl	P	As
LS	327	11	15	4	1	1	1
HS	563	0	30	3	0	1	0
Unk.	12	0	0	0	0	0	0
Total	902	11	45	7	1	2	1
	Fe(II) high confidence						
	N	C	O	S	Cl	P	As
LS	194	11	10	2	0	1	1
HS	199	0	21	1	0	1	0
Total	393	11	31	3	0	2	1
	Fe(III) overall						
	N	C	O	S	Cl	F	
LS	160	5	77	17	1	1	
HS	19	0	51	12	0	0	
Unk.	0	0	4	0	0	0	
Total	179	5	132	29	1	1	
	Fe(III) high confidence						
	N	C	O	S	Cl	F	
LS	67	3	42	10	0	0	
HS	7	0	37	1	0	0	
Total	74	3	79	11	0	0	

Table S12. Number of confident spin state assignments (uncertainty score less than or equal to 0.5) in the NX set for X=Cl, O, S in Fe(II) or Fe(III) complexes as well as the total number of each type and the percentage of the total number assigned.

	Fe(II)					Fe(III)				
		Assigned, high confidence					Assigned, high confidence			
	All	LS	HS	LS+HS	% Assign	All	LS	HS	LS+HS	% Assign
N/Cl	48	4	6	10	21%	19	3	2	5	26%
N/O	210	11	78	89	42%	137	22	24	46	34%
N/S	15	7	1	8	53%	9	5	0	5	56%

Table S13. Summary of properties of 46 Fe(II) SCO complexes: LS and HS ANN equatorial bond length predictions (in Å) compared to the equatorially averaged low T and high T XRD structures. Classification based on the closest corresponding ANN prediction to the structure is shown in the class column, and the differences used for this calculation are also shown. Key minimum, maximum, and average values are also reported.

	refcode	max dent	LS ANN eq	HS ANN eq	low T eq av	high T eq av	low T class	high T class	LS ANN-low T	HS ANN-low T	HS ANN-high T	LS ANN-high T
1	PEJQIF	1	1.9547	2.3356	2.0010	2.1811	LS	HS	0.0463	0.3346	0.1544	0.2264
2	YAGYIP	1	2.0300	2.3162	2.0887	2.2208	LS	HS	0.0587	0.2275	0.0954	0.1908
3	BEXXOT	2	2.0031	2.2696	1.9947	2.1607	LS	HS	0.0084	0.2750	0.1089	0.1576
4	DOQRAC	3	2.0184	2.2779	2.0190	2.1343	LS	LS	0.0006	0.2589	0.1436	0.1158
5	BAKGUR	2	2.0032	2.2271	2.0082	2.2052	LS	HS	0.0050	0.2189	0.0219	0.2020
6	ECODOS	2	2.0074	2.2271	1.9873	2.1839	LS	HS	0.0201	0.2398	0.0432	0.1765
7	OLAYAC	2	2.0441	2.2629	2.0023	2.1658	LS	HS	0.0418	0.2606	0.0971	0.1217
8	AKENAF	2	1.9680	2.1841	1.9491	2.0929	LS	HS	0.0190	0.2350	0.0913	0.1248
9	ECODIM	2	1.9954	2.1958	2.0335	2.1835	LS	HS	0.0381	0.1624	0.0123	0.1881
10	ZERDOS	2	1.9857	2.1854	1.9962	2.1469	LS	HS	0.0106	0.1891	0.0384	0.1613
11	WUCLAJ	2	1.9666	2.1663	1.9946	2.1923	LS	HS	0.0280	0.1717	0.0260	0.2256
12	IDUHAT	2	1.9495	2.1488	1.9790	2.1753	LS	HS	0.0295	0.1698	0.0266	0.2259
13	SEYDAD	2	1.9941	2.1892	1.9832	2.2006	LS	HS	0.0109	0.2061	0.0114	0.2065
14	FIWXOZ	3	1.9575	2.1461	2.0056	2.1747	LS	HS	0.0481	0.1405	0.0286	0.2172
15	MALJIT	3	2.0005	2.1856	1.9833	2.1595	LS	HS	0.0172	0.2023	0.0261	0.1590
16	BAYZEH	2	1.9147	2.0992	1.9781	2.1443	LS	HS	0.0634	0.1211	0.0450	0.2296
17	MUCRUZ	3	1.9426	2.1236	1.9349	2.1279	LS	HS	0.0077	0.1887	0.0043	0.1853
18	OTUJUJ	2	1.9680	2.1471	1.9946	2.1959	LS	HS	0.0266	0.1525	0.0488	0.2279
19	QUNYEE	2	1.9367	2.1116	1.9588	2.1581	LS	HS	0.0221	0.1528	0.0465	0.2215
20	IMAHIO	2	2.0459	2.2182	1.9817	2.1198	LS	LS	0.0642	0.2365	0.0984	0.0739
21	MAVVUD	3	1.9727	2.1419	1.9547	2.1616	LS	HS	0.0180	0.1872	0.0197	0.1889
22	RESWIW	3	1.9736	2.1389	1.9933	2.1631	LS	HS	0.0197	0.1456	0.0242	0.1895
23	IGERIX	3	1.9200	2.0773	1.9672	2.1781	LS	HS	0.0473	0.1101	0.1008	0.2581
24	CEVXOT	2	2.0029	2.1589	1.9880	2.1610	LS	HS	0.0149	0.1709	0.0021	0.1581
25	FEYWIR	2	2.0270	2.1818	1.9857	2.1958	LS	HS	0.0413	0.1961	0.0140	0.1688
26	EKAKIM	3	1.9261	2.0662	1.9388	2.1538	LS	HS	0.0128	0.1274	0.0876	0.2277
27	UFIPIJ	2	1.9282	2.0632	1.9601	2.1619	LS	HS	0.0319	0.1031	0.0987	0.2337
28	MUCREJ	3	1.9269	2.0584	1.9316	2.1350	LS	HS	0.0047	0.1267	0.0766	0.2081
29	SEYDUX	2	2.0034	2.1336	1.9863	2.2092	LS	HS	0.0171	0.1473	0.0757	0.2058
30	RIPZAS	3	1.9643	2.0942	2.0487	2.1451	HS	HS	0.0844	0.0455	0.0509	0.1808
31	QALYOU	2	1.9650	2.0948	1.9741	2.1422	LS	HS	0.0091	0.1207	0.0474	0.1772
32	NETGEA	2	2.0810	2.2064	1.9901	2.2124	LS	HS	0.0909	0.2164	0.0059	0.1314
33	AFANEB	3	1.9610	2.0778	1.9519	2.1529	LS	HS	0.0090	0.1259	0.0751	0.1920
34	PEJPOK	2	1.9975	2.1131	1.9969	2.2109	LS	HS	0.0006	0.1162	0.0978	0.2134
35	ABISOV	4	1.9827	2.0969	1.9616	2.1153	LS	HS	0.0211	0.1353	0.0184	0.1327
36	EXOMOU	3	1.9712	2.0696	1.9673	2.1504	LS	HS	0.0039	0.1023	0.0808	0.1792
37	GEFCAY	3	1.9622	2.0601	1.9642	2.1596	LS	HS	0.0019	0.0959	0.0995	0.1973
38	IXAQEF	5	1.9722	2.0614	2.0141	2.1215	LS	HS	0.0419	0.0473	0.0601	0.1494
39	QERDIB	3	1.9723	2.0580	1.9957	2.1767	LS	HS	0.0233	0.0623	0.1187	0.2044
40	JEJROI	4	1.9925	2.0681	1.9901	2.1504	LS	HS	0.0023	0.0779	0.0823	0.1579
41	JEJREY	4	1.9982	2.0403	1.9964	2.1563	LS	HS	0.0018	0.0439	0.1160	0.1581
42	PASDOD	6	1.9798	2.0184	1.9767	2.1943	LS	HS	0.0031	0.0416	0.1760	0.2145
43	TIHTAF	4	2.0506	2.0839	2.0097	2.1865	LS	HS	0.0408	0.0742	0.1026	0.1359
44	DIHCUV	4	2.0486	2.0657	2.0290	2.1964	LS	HS	0.0195	0.0367	0.1308	0.1479
45	VASYUL	4	1.9968	2.0137	1.9924	2.1381	LS	HS	0.0045	0.0214	0.1244	0.1413
46	IMANIT	6	1.9833	1.9914	2.0810	2.1773	HS	HS	0.0977	0.0896	0.1860	0.1941
		Avg	**1.9836**	**2.1359**	**1.9896**	**2.1659**			**0.0267**		**0.0704**	
		Min	**1.9147**	**1.9914**	**1.9316**	**2.0929**			**0.0006**		**0.0021**	
		Max	**2.0810**	**2.3356**	**2.0887**	**2.2208**			**0.0977**		**0.1860**	

References

1 Janet, J. P. & Kulik, H. J. Predicting Electronic Structure Properties of Transition Metal Complexes with Neural Networks. *Chem. Sci.* **8**, 5137-5152 (2017).

2 Janet, J. P. & Kulik, H. J. Resolving transition metal chemical space: feature selection for machine learning and structure-property relationships. *Journal of Physical Chemistry A* **121**, 8939-8954 (2017).