Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2016

Supplementary Information

Predictions of the physicochemical properties of amino acid side chain analogs using molecular simulation

Alauddin Ahmed‡ and Stanley I. Sandler*

Center for Molecular and Engineering Thermodynamics, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, USA

‡ Current Address: Mechanical Engineering Department, University of Michigan, Ann Arbor, Michigan 48109, USA

**Corresponding Author: sandler@udel.edu*

(Version 21 January 2016)

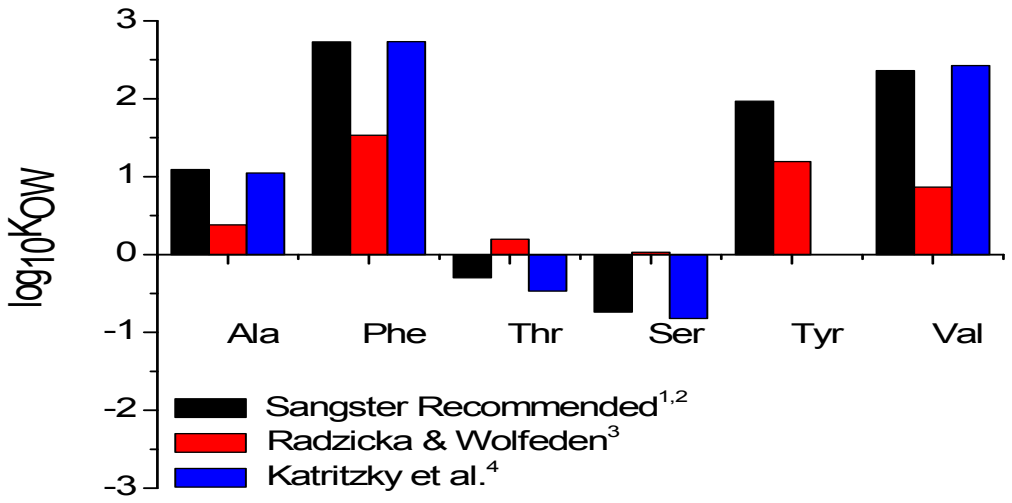


Figure S1: Octanol-water partition coefficients calculated and compared with recommended values (data shown in the Table S1).

Table S1: Octanol-water partition coefficients from different sources.[1-4]

Amino Acid	Side Chain	Sangster[1,2]Recommended	Radzicka and Wolfenden[3]	Katrizky et al.[4]
Ala (Alanine)	Methane	1.09	0.38	1.05
Phe (Phenylalanine)	Toluene	2.73	1.53	2.73
Thr (Threonine)	Ethanol	-0.3	0.20	-0.47
Ser (Serine)	Methanol	-0.74	0.03	-0.82
Tyr (Tyrosine)	p-Cresol	1.97	1.19	---
Val (Valine)	Propane	2.36	0.86	2.43

Table S2a: Self-solvation free energies (in kcal/mol) of 1-octanol.

Force Field	Experiment	ΔG_{self}
CGenFF/ParamChem (charge)	-8.41	-10.02 ± 0.67
GAFF/AM1-BCC (charge)		-8.31 ± 0.24
OPLS-AA (default charge)		-7.05 ± 0.47

Table S2b: Self-solvation free energies (in kcal/mol) of amino acid side chain analogs (ΔG_{self}).

Amino acid	Side chain analog/other	Experimental	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)	OPLS-AA
Ala (Alanine)	Methane	---	---	---	---
Phe (Phenylalanine)	Toluene	-5.17	-4.45 ± 0.21	-3.98 ± 0.21	-4.86 ± 0.18
Thr (Threonine)	Ethanol	-5.08	-5.16 ± 0.24	-4.60 ± 0.17	-4.58 ± 0.21
Ser (Serine)	Methanol	-4.90	-4.50 ± 0.17	-3.79 ± 0.20	-4.65 ± 0.14
Tyr (Tyrosine)	p-Cresol	-8.46	-7.86 ± 0.36	-7.69 ± 0.33	-8.26 ± 0.51
Val (Valine)	Propane	---	---	---	---
	AUE (kcal/mol)		0.45	0.90	0.33
	RMSE (kcal/mol)		0.51	0.94	0.34

Table S3: Hydration free energies (in kcal/mol) of amino acid side chain analogs (ΔG_{wat}).

Amino acid	Side chain analog	Experimental	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)[5]	OPLS-AA[6]
Ala (Alanine)	Methane	1.95	2.44 ± 0.03	2.54 ± 0.01	2.31 ± 0.02
Phe (Phenylalanine)	Toluene	-0.78	0.24 ± 0.05	-0.71 ± 0.02	-0.54 ± 0.04
Thr (Threonine)	Ethanol	-5.03	-4.52 ± 0.04	-3.45 ± 0.03	-4.11 ± 0.03
Ser (Serine)	Methanol	-5.03	-4.58 ± 0.03	-3.48 ± 0.01	-4.36 ± 0.02
Tyr (Tyrosine)	p-Cresol	-6.11	-4.02 ± 0.05	-5.36 ± 0.02	-5.25 ± 0.04
Val (Valine)	Propane	1.99	2.42 ± 0.03	2.56 ± 0.02	2.59 ± 0.03
	AUE (kcal/mol)		0.83	0.85	0.61
	RMSE (kcal/mol)		1.02	1.01	0.66

Table S4: Solvation free energy (in kcal/mol) of amino acid side chain analogs in 1-octanol (ΔG_{oct}).

Amino acid	Side chain analog	Experiment	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)	OPLS-AA
Ala (Alanine)	Methane	0.52	0.76 ± 0.18	1.03 ± 0.19	0.68 ± 0.12
Phe (Phenylalanine)	Toluene	-4.51	-3.98 ± 0.20	-3.27 ± 0.26	-3.88 ± 0.35
Thr (Threonine)	Ethanol	-4.39	-4.15 ± 0.26	-3.65 ± 0.33	-4.31± 0.35
Ser (Serine)	Methanol	-3.91	-3.24 ± 0.29	-3.07 ± 0.24	-2.86 ± 0.29
Tyr (Tyrosine)	p-Cresol	-8.84	-8.07 ± 0.35	-7.44 ± 0.24	-8.17 ± 0.40
Val (Valine)	Propane	-1.32	-0.53 ± 0.29	-0.62 ± 0.18	-0.30 ± 0.21
AUE (kcal/mol)			0.54	0.90	0.60
RMSE (kcal/mol)			0.58	0.96	0.71

Table S5a: Logarithm of vapor pressures in Pa unit ($log_{10}P_{vap}^{liq}$) for 1-octanol models.

Force Field	Molecular Weight (g/mol)	Density, ρ (g/L)	Experiment	$log_{10}P_{vap}^{liq/sol}$
CGenFF/ParamChem (charge)	130.23	827	1.03	-0.15
GAFF/AM1-BCC (charge)				1.10
OPLS-AA				2.03

Table S5b: Logarithm of vapor pressures in Pa unit ($log_{10}P_{vap}^{liq}$) for amino acid side chain analogs.

Amino acid	Side chain analog	Molecular Weight (g/mol)	Density, ρ (g/L)	Experiment	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)	OPLS-AA
Ala (Alanine)	Methane	16.04		7.79	---	---	---
Phe (Phenylalanine)	Toluene	92.14	852	3.58	4.10	4.44	3.79
Thr (Threonine)	Ethanol	46.07	799	3.90	3.85	4.26	4.26
Ser (Serine)	Methanol	32.04	759	4.23	4.48	4.18	4.36

Tyr (Tyrosine)	p-Cresol	108.13	1033	---	---	---	---
Val (Valine)	Propane	44.1		5.98	---	---	---
	AUE (log unit)				0.27	0.43	0.24
	RMSE (log unit)				0.33	0.54	0.25

Table S6: Logarithms of solubilities in mol/L units $(log_{10}S_{wat})$ of amino acid side chain analogs in water calculated from vapor pressures (P_{vap}^{liq}) and solvation free energies in water (ΔG_{wat}) using eqn. (3).

Amino acid	Side chain analog	Molecular Weight (g/mol)	Density, ρ (g/L)	Experiment	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)	OPLS-AA
Ala (Alanine)	Methane	16.04		-2.86	---	---	---
Phe (Phenylalanine)	Toluene	92.14	852	-2.24	-2.47	-1.43	-2.21
Thr (Threonine)	Ethanol	46.07	799	1.34	0.77	0.40	0.88
Ser (Serine)	Methanol	32.04	759	1.49	1.44	0.34	1.16
Tyr (Tyrosine)	p-Cresol	108.13	1033	---	---	---	---
Val (Valine)	Propane	44.1		-2.85	---	---	---
	AUE (log unit)				0.28	0.97	0.28
	RMSE (log unit)				0.35	0.98	0.33

Table S7: Logarithms of solubilities in mol/L units $(log_{10}S_{oct})$ of amino acid side chain analogs in 1-octanol calculated from vapor pressures (P_{vap}^{liq}) and solvation free energies in 1-octanol (ΔG_{oct}) using eqn. (3).

Amino acid	Side chain analog/other	Molecular Weight (g/mol)	Density, ρ (g/L)	Experiment	CGenFF/ParamChem (charge)	GAFF/AM1-BCC (charge)	OPLS-AA
Ala (Alanine)	Methane	16.04		2.22	---	---	---
Phe (Phenylalanine)	Toluene	92.14	852	0.49	0.62	0.44	0.25
Thr (Threonine)	Ethanol	46.07	799	0.72	0.50	0.55	1.03
Ser (Serine)	Methanol	32.04	759	0.70	0.46	0.04	0.06
Tyr (Tyrosine)	*p*-Cresol	108.13	1033	---	---	---	---
Val (Valine)	Propane	44.1		---	---	---	---
	AUE (log unit)				0.20	0.30	0.40
	RMSE (log unit)				0.20	0.40	0.43

Table S8: Logarithms of octanol-water ($log_{10}K_{OW}$) partition coefficientof amino acid side chain analogs.

		CGenFF	GAFF	OPLS-AA
Amino Acid Analog	Experiment	This work	This work + (Mobley et al. 2009)	This work + (Shirts et al. 2003)
Ala (Alanine)	1.09	1.23 ± 0.13	1.10 ± 0.14	1.19 ± 0.09
Phe (Phenylalanine)	2.73	3.09 ± 0.15	1.87 ± 0.19	2.45 ± 0.26
Thr (Threonine)	-0.3	-0.27 ± 0.19	0.15 ± 0.24	0.15 ± 0.26
Ser (Serine)	-0.74	-0.98 ± 0.21	-0.30 ± 0.17	-1.10 ± 0.21
Tyr (Tyrosine)	1.97	2.97 ± 0.26	1.53 ± 0.17	2.14 ± 0.30
Val (Valine)	2.36	2.17 ± 0.22	2.33 ± 0.13	2.12 ± 0.17
AUE (kcal/mol)		0.33	0.37	0.27
RMSE (kcal/mol)		0.45	0.47	0.29

Table S9: Logarithms of air-water partition coefficient ($log_{10}K_{AW}$) of amino acid side chain analogs calculated from solvation free energies in water.

Amino acid	Side chain analog/other	Experiment	CGenFF	GAFF	OPLS-AA
Ala (Alanine)	Methane	-1.45	-1.79	-1.86	-1.69
Phe (Phenylalanine)	Toluene	0.58	-0.17	0.52	0.40
Thr (Threonine)	Ethanol	3.67	3.32	2.53	3.01
Ser (Serine)	Methanol	3.74	3.36	2.55	3.20
Tyr (Tyrosine)	p-Cresol	4.5	2.95	3.93	3.85
Val (Valine)	Propane	-1.44	-1.78	-1.88	-1.90
AUE (log unit)			0.62	0.63	0.46
RMSE (log unit)			0.76	0.75	0.49

Table S10: Logarithms of air-octanol partition coefficient ($log_{10}K_{AO}$) of amino acid side chain analogs calculated from solvation free energies in 1-octanol.

Amino acid	Side chain analog/other	Experimental	CGenFF	GAFF	OPLS-AA
Ala (Alanine)	Methane	-0.38	-0.56	-0.76	-0.50
Phe (Phenylalanine)	Toluene	3.31	2.91	2.39	2.85
Thr (Threonine)	Ethanol	3.25	3.04	2.68	3.16
Ser (Serine)	Methanol	2.88	2.37	2.25	2.10
Tyr (Tyrosine)	p-Cresol	6.48	5.91	5.46	5.99
Val (Valine)	Propane	0.97	0.39	0.45	0.22
AUE (log unit)			0.41	0.67	0.45
RMSE (log unit)			0.44	0.71	0.52

Table S11a: Comparison of self-solvation free energies calculated using solvation models and force field-based molecular simulations.

Method	AUE (kcal/mol)	RMSE (kcal/mol)
SM5.4/AM1	0.21	0.26
SM5.4/PM3	0.19	0.20
SM5.2R/AM1//MM3	1.63	2.51
SM5.2R/PM3//MM3	1.17	1.75
SM5.2R/AM1//MM2	0.30	0.35
SM5.2R/PM3//MM2	0.29	0.29
SM5.0R/MM3	0.74	0.96
SM5.0R/MM2	0.26	0.36
Simulation: CGenFF/ParamChem Charge	0.40	0.48
Simulation: GAFF/AM1-BCC charge	0.93	0.98
Simulation: OPLS-AA/Default Charge	**0.35**	**0.37**

Table S11b: Comparison of self-solvation free energy of *p*-cresol calculated using solvation models and force field-based molecular simulations.

Method	*p*-cresol		
	ΔG_{self}^{expt} (kcal/mol)	$\Delta G_{self}^{sim/mod}$ (kcal/mol)	Unsigned Error (kcal/mol)
SM5.42R/HF	-8.46	-9.68	1.22
SM5.42R/HB3LYP		-9.77	1.31
SM5.42R/AM1		-9.80	1.34
CGenFF		-7.86	0.60
GAFF		-7.69	0.77
OPLS-AA		**-8.26**	**0.20**

Table S12a: Comparison of logarithm of vapor pressures (in Pa) calculated using solvation models and force field-based molecular simulations.

Method	AUE (log unit)	RMSE (log unit)
SM5.4: AM1[7-9]	0.14	0.16
SM5.4: PM3[7-9]	0.12	0.13
SM5.2R: AM1//MM3[7-9]	1.17	1.84
SM5.2R: PM3//MM3[7-9]	0.84	1.27
SM5.2R: AM1//MM2[7-9]	0.21	0.25
SM5.2R: PM3//MM2[7-9]	0.19	0.21
SM5.0R: MM3[7-9]	0.53	0.69
SM5.0R: MM2[7-9]	0.18	0.24
Simulation: CGenFF/ParamChem (charge) (this work)	0.27	0.33
Simulation: GAFF/AM1-BCC (charge) (this work)	0.43	0.54
Simulation: OPLS-AA/ (default charge) (this work)	0.24	0.25

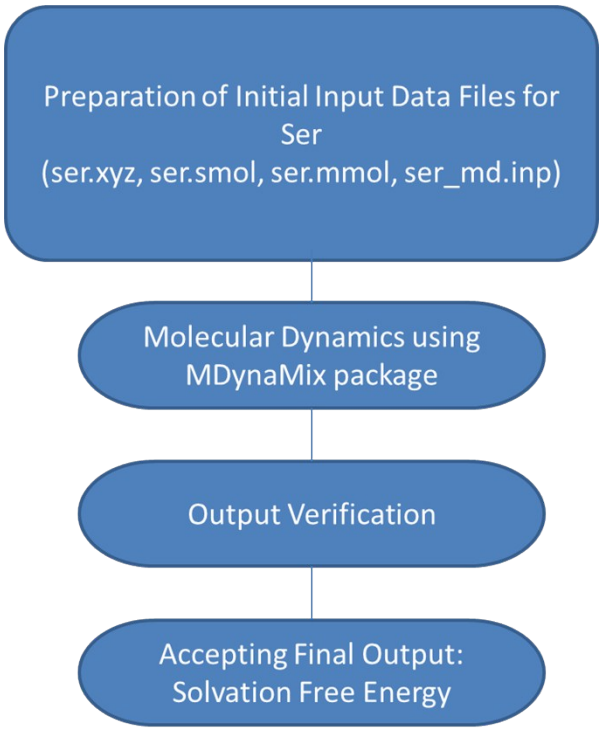
Table S12b: Comparison of logarithm of vapor pressures (in Pa) of tyr (*p*-cresol) calculated using solvation models and force-field based molecular simulations.

Method	Tyr (*p*-cresol)		
	Expt. (log unit)	$\log(P^{vap} in\ Pa)$	UE (log unit)
SM5.42R/HF	1.13	2.01	0.88
SM5.42R/HB3LYP		1.99	0.86
SM5.42R/AM1		2.25	1.12
CGenFF		1.60	0.47
GAFF		1.73	0.60
OPLS-AA		**1.59**	**0.46**

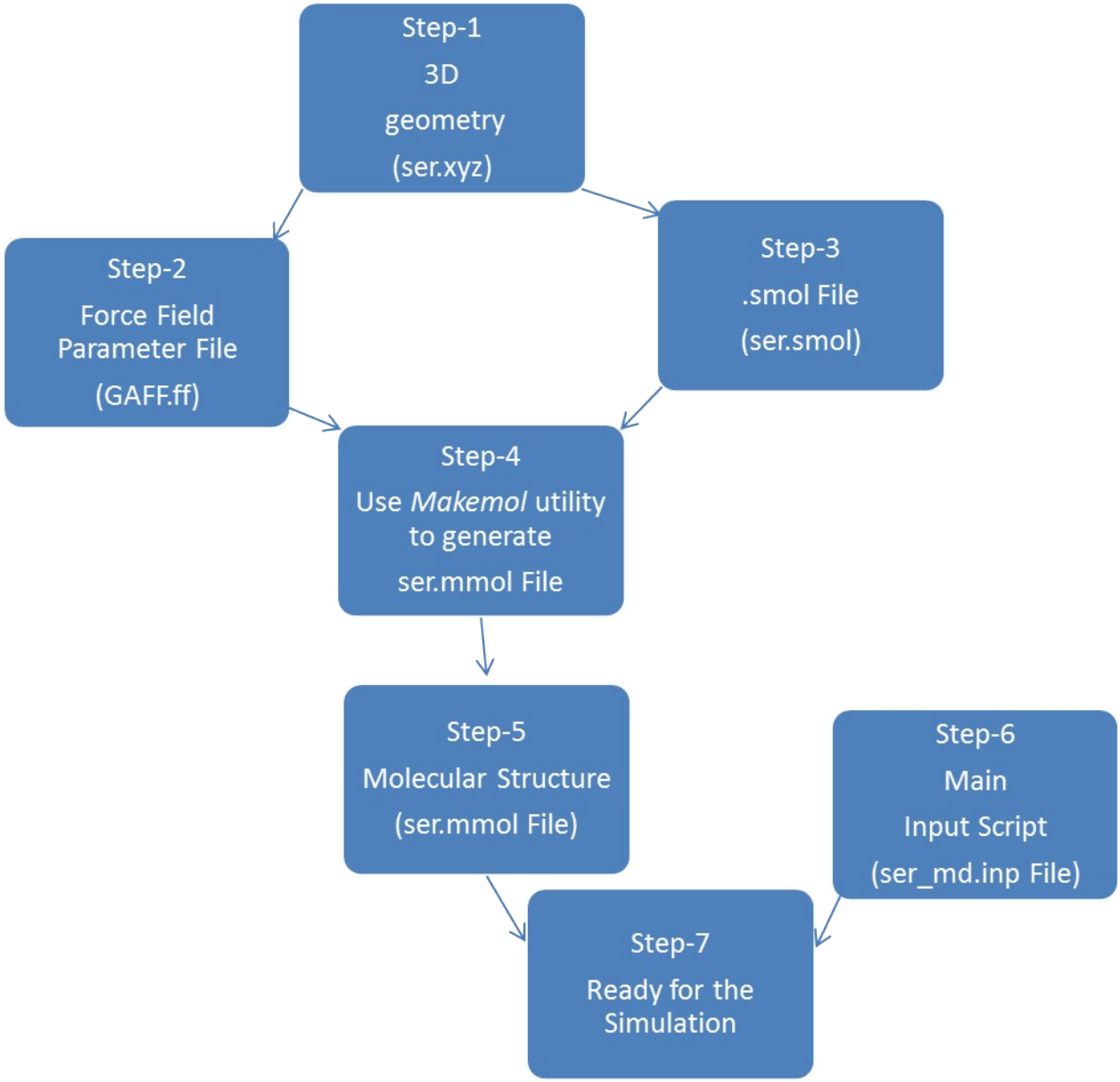
Table S13: Comparison of logarithms of sub-cooled solubility (in mol/L unit) calculated using solvation models and force field-based molecular simulations.

Method	Phe (Toluene)			Tyr (p-cresol)		
	Expt. (log unit)	log(S)	UE (log unit)	Expt.	log(S)	UE (log unit)
SM5.42R/HF	-2.24	-2.00	0.24	-0.70	-0.52	0.18
SM5.42R/HB3LYP		-1.88	0.36		-0.48	0.22
SM5.42R/AM1		-1.99	0.25		-0.45	0.25
CGenFF		-2.47	0.23		-1.92	1.22
GAFF		-1.43	0.81		-0.81	0.11
OPLS-AA		-2.21	0.03		-1.03	0.33

NPT-EE Work Flow with an Example



Preparation of Input Data Files



Ser Topology and Parameter File (ser.mmol)

```
#CGenFF Force Field Parameters
#Prepared by Alauddin, Sandler group, UD
# April 12, 2012
```

```
# Ser (Methanol)
#  Number of atoms
 6
C      -1.954     .117    -.419   12.011   -.0390   3.6527    .3266 CG331        1
O      -1.538    1.221   -1.193   15.999   -.6510   3.1449    .8043 OG311        2
H      -3.023    -.146    -.591    1.008    .0900   2.3876    .1005 HGA3         3
H      -1.301    -.749    -.694    1.008    .0900   2.3876    .1005 HGA3         4
H      -2.292    1.515   -1.711    1.008    .4200    .4000    .1926 HGP1         5
H      -1.797     .384     .656    1.008    .0900   2.3876    .1005 HGA3         6
 3
 File created by utility MAKEMOL, package MDynaMix
 Coordinates taken from file methanol.smol
 Force field from file CGEN.ff
#  Number of bonds
 8
     0     1     2  1.42000   1792.00                    #  CG331   -OG311
     0     1     3  1.11100   1348.20                    #  CG331   -HGA3
     0     1     4  1.11100   1348.20                    #  CG331   -HGA3
     0     1     6  1.11100   1348.20                    #  CG331   -HGA3
     0     2     5   .96000   2281.90                    #  OG311   -HGP1
#    Urey-Bradley terms (for 1-3 neigbours)
     2     3     4  1.80200     22.61                    #  HGA3    -HGA3
     2     3     6  1.80200     22.61                    #  HGA3    -HGA3
     2     4     6  1.80200     22.61                    #  HGA3    -HGA3
#  Number of angles
 7
     2     1     3  108.8900    192.20     #  OG311   -CG331   -HGA3
     2     1     4  108.8900    192.20     #  OG311   -CG331   -HGA3
     2     1     6  108.8900    192.20     #  OG311   -CG331   -HGA3
     3     1     4  108.4000    148.60     #  HGA3    -CG331   -HGA3
     3     1     6  108.4000    148.60     #  HGA3    -CG331   -HGA3
     4     1     6  108.4000    148.60     #  HGA3    -CG331   -HGA3
     1     2     5  106.0000    240.80     #  CG331   -OG311   -HGP1
#  Number of torsions
 3
     3     1     2     5     .000     .800   3  #HGA3    -CG331   -OG311   -HGP1
     4     1     2     5     .000     .800   3  #HGA3    -CG331   -OG311   -HGP1
     6     1     2     5     .000     .800   3  #HGA3    -CG331   -OG311   -HGP1
#   Special 1-4 interactions
special
 1
#  I   Sig14   Eps14
     1     3.3854      .0419
#
```

REFERENCES

1 J. Sangster, Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry, John Wiley & Son Ltd, New York, 1997.

2 Sangster Research Laboratories, http://logkow.cisti.nrc.ca/logkow/index.jsp, (accessed Oct 3, 2012).
3 A. Radzicka and R. Wolfenden, *Biochemistry*, 1988, **27**, 1664–1670.
4 A. R. Katritzky, A. A. Oliferenko, P. V. Oliferenko, R. Petrukhin and D. B. Tatham, *J. Chem. Inf. Comput. Sci.,* 2003, **43**, 1794–1805.
5 D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts and K. A. Dill, *J. Chem. Theory. Comput.,* 2009, **5**, 350–358.
6 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *J. Chem. Phys.*2003, **119**, 5740–5761.
7 P. Winget, G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys. B,* 2000, **104**, 4726–4734.
8 J. D. Thompson, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys,* 2003, **119**, 1661–1670.
9 A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer and D. G. Truhlar, *Minnesota Solvation Database Version 2007*, University of Minnesota: Minneapolis, MN, 2009.