Available online at www.sciencedirect.com
ScienceDirect
Organic Geochemistry 39 (2008) 1189–1196
Organic Geochemistry
www.elsevier.com/locate/orggeochem

Restoration of Circum-Arctic Upper Jurassic source rock paleolatitude based on crude oil geochemistry

K.E. Peters [a,*], L.S. Ramos [b], J.E. Zumberge [c], Z.C. Valin [a], C.R. Scotese [d]

[a] *U.S. Geological Survey, 345 Middlefield Road, MS 969, Menlo Park, CA 94025, USA*
[b] *Infometrix Inc., 10634 East Riverside Drive, Suite 250, Bothell, WA 98011, USA*
[c] *GeoMark Research Ltd., 9748 Whithorn Drive, Houston, TX 77095, USA*
[d] *University of Texas at Arlington, PALEOMAP Project, Box 19049, Arlington, TX 76012, USA*

Received 16 July 2007; received in revised form 14 January 2008; accepted 23 January 2008
Available online 9 February 2008

Abstract

Tectonic geochemical paleolatitude (TGP) models were developed to predict the paleolatitude of petroleum source rock from the geochemical composition of crude oil. The results validate studies designed to reconstruct ancient source rock depositional environments using oil chemistry and tectonic reconstruction of paleogeography from coordinates of the present day collection site. TGP models can also be used to corroborate tectonic paleolatitude in cases where the predicted paleogeography conflicts with the depositional setting predicted by the oil chemistry, or to predict paleolatitude when the present day collection locality is far removed from the source rock, as might occur due to long distance subsurface migration or transport of tarballs by ocean currents. Biomarker and stable carbon isotope ratios were measured for 496 crude oil samples inferred to originate from Upper Jurassic source rock in West Siberia, the North Sea and offshore Labrador. First, a unique, multi-tiered chemometric (multivariate statistics) decision tree was used to classify these samples into seven oil families and infer the type of organic matter, lithology and depositional environment of each organofacies of source rock [Peters, K.E., Ramos, L.S., Zumberge, J.E., Valin, Z.C., Scotese, C.R., Gautier, D.L., 2007. Circum-Arctic petroleum systems identified using decision-tree chemometrics. American Association of Petroleum Geologists Bulletin 91, 877–913]. Second, present day geographic locations for each sample were used to restore the tectonic paleolatitude of the source rock during Late Jurassic time (~150 Ma). Third, partial least squares regression (PLSR) was used to construct linear TGP models that relate tectonic and geochemical paleolatitude, where the latter is based on 19 source-related biomarker and isotope ratios for each oil family. The TGP models were calibrated using 70% of the samples in each family and the remaining 30% of samples were used for model validation. Positive relationships exist between tectonic and geochemical paleolatitude for each family. Standard error of prediction for geochemical paleolatitude ranges from 0.9° to 2.6° of tectonic paleolatitude, which translates to a relative standard error of prediction in the range 1.5–4.8%. The results suggest that the observed effect of source rock paleolatitude on crude oil composition is caused by (i) stable carbon isotope fractionation during photosynthetic fixation of carbon and (ii) species diversity at different latitudes during Late Jurassic time.

* Corresponding author. Tel.: +1 650 329 5171; fax: +1 650 329 4975.
E-mail address: kpeters@usgs.gov (K.E. Peters).

0146-6380/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.orggeochem.2008.01.016

1. Introduction

Because crude oil commonly migrates from an effective source rock through faults and sedimentary strata to the trap, geologic observations alone may be insufficient to clearly identify the source rock of an oil sample. The most important contribution of petroleum geochemistry to exploration is that it can establish the chemical similarity between migrated oil and bitumen remaining in the source rock (oil-source rock correlation), thus providing part of the information needed to map petroleum systems, and a key competitive exploration advantage (Magoon et al., 1994; Peters and Fowler, 2002). However, samples of the effective source rock may be unavailable for comparison with produced crude oil because source rocks are generally more deeply buried than reservoir rocks, so may not be sampled by drilling. Fortunately, the geochemical composition of crude oil can be used to infer the identity of the source rock (e.g., Peters et al., 2005).

Here, we expand on a new application of the geochemical data for crude oil samples. Because of plate tectonics, oil accumulations and their source rocks may be displaced by hundreds or even thousands of kilometers from the original site of source rock deposition (e.g., Mello et al., 1991; Peters et al., 1995). The question is: Can the geochemistry of crude oil be used to infer the paleogeographic setting of the source rock?

The study examines a collection of chemically similar crude oil samples that originated from different organofacies (Jones, 1987) of Upper Jurassic source rock in the Circum-Arctic region north of ~55°N. The samples were distinguished using decision-tree chemometrics as described by Peters et al. (2007). Applications similar to our decision-tree method were used to identify different species of mycobacteria (Ramos, 1994) and to classify crude oil (Zumberge and Ramos, 1996). The purpose of the present study was to apply chemometrics to crude oil geochemical data in order to predict paleolatitude of the corresponding source rock at the time of deposition. The approach provides a valuable tool for better determining source rock depositional settings, past climate and the present-day worldwide distribution and quality of petroleum.

2. Methods

To demonstrate the application of TGP models, geochemical data for 496 crude oil samples from Upper Jurassic source rocks in the Circum-Arctic area north of ~55°N were selected from a larger worldwide database (>10,000 samples; Reservoir Fluid Database or RFDbase; GeoMark Research) using several criteria (Peters et al., 2007). Heavily biodegraded or highly thermally mature oil samples were omitted because these secondary processes can alter source-related biomarker and isotope ratios. The study used 19 of the 20 sterane, terpane and stable carbon isotope ratios described by Peters et al. (2007). The oleanane/hopane ratio was omitted because the Upper Jurassic oil samples lack oleanane. Crude oil samples from Upper Jurassic source rocks were selected because many samples showing widespread distribution were available. Improved understanding of Upper Jurassic source rocks is important because they contain oil-prone type II and IIS kerogen that accounts for ~25% of worldwide petroleum reserves (Klemme and Ulmishek, 1991). According to these authors, most Upper Jurassic petroleum occurs in the Arabian–Iranian (46%), West Siberian (22%), Gulf of Mexico (13%) and North Sea (11%) basins.

Paleogeography of the source rock at the time of deposition was restored from present-day latitudes and longitudes of the collection site for each oil sample using AutoPointTracker© (PALEOMAP Project, C. Scotese). For a given geologic time slice, the program backtracks the tectonic movement from a present day location to a paleo-location. We used the available 150 Ma time slice provided by AutoPointTracker to represent Late Jurassic time, recognizing that this is an approximation. The restored paleogeographic location for each sample was mapped using Earth System History-GIS© (PALEOMAP Project, C. Scotese).

Chemometric analysis of the geochemical data was completed using Pirouette® (version 3.11, Infometrix Inc.). PLSR (Geladi and Kowalski, 1986) finds a linear model that describes some predicted variable, such as paleolatitude, in terms of observed variables, such as biomarker and stable carbon isotope ratios. It was completed using autoscale preprocessing, where all values for each variable are normalized to the standard deviation for that variable among the samples. This results in equal weight for each biomarker or stable isotope ratio in the computation.

Validation is an important step when fitting PLSR models to be used for prediction of new observations. This requires that one applies the current results to a new set of observations that were

Table 1
Comparison of PLSR calibration and validation results to predict geochemical paleolatitude from tectonic paleolatitude for Circum-Arctic Upper Jurassic oil samples[a]

Oil family	Calibration (70% of samples)				Validation (30% of samples)			
	# of Samples	Mean paleolatitude[b]	SEV[c] (degrees)	RSEP[d] (%)	# of Samples	Mean paleolatitude	SEP[c] (degrees)	RSEP[d] (%)
All	347	57.46	2.93	4.38	149	58.14	2.10	3.60
231	148	59.68	1.15	1.92	64	59.62	0.89	1.49
2322	6	54.49	13.34	26.25	3	52.39	2.64	4.84
2323	32	60.87	3.06	5.00	15	61.66	1.50	2.48
3211	28	65.09	1.90	2.92	13	63.41	1.20	1.90
3212	11	54.45	2.48	4.61	5	53.46	1.66	3.06
3213	72	51.89	1.78	3.43	31	51.68	1.42	2.76
3214	47	54.72	3.02	5.46	21	55.84	2.55	4.45
3214_no isotopes	47	54.80	3.46	6.25	21	55.25	3.20	5.59

[a] Seven families, including Family 3214, with and without isotope data (Peters et al. (2007)).
[b] Mean paleolatitude: geochemical paleolatitude.
[c] SEV, SEP: standard error of cross validation and prediction, respectively (degrees latitude).
[d] Relative standard error of prediction (%, Faber et al. (2004)).

not used to compute the initial results. We used two validation methods: (i) PLSR for all crude oil samples for each family was cross validated using the method in Pirouette where each sample is treated as an unknown for prediction using a model based on the remaining samples (leave-one-out cross validation). (ii) A rigorous calibration and validation was completed where predictive PLSR models were calibrated using 70% of the samples in each family (70%-PLSR model) selected with the method of Kennard and Stone (1969). The remaining 30% were treated as unknowns for prediction (validation) of geochemical paleolatitude using the 70%-PLSR models. Because standard correlation coefficients (r^2) are inappropriate for PLSR correlation statistics (Davies and Fearn, 2006), the quality of the calibrations was measured using relative standard error of prediction (RSEP, %; Table 1). RSEP for each oil family is the square root of the ratio of the standard deviations of the differences between tectonic (measured) and geochemical (predicted) paleolatitude to that of the tectonic values.

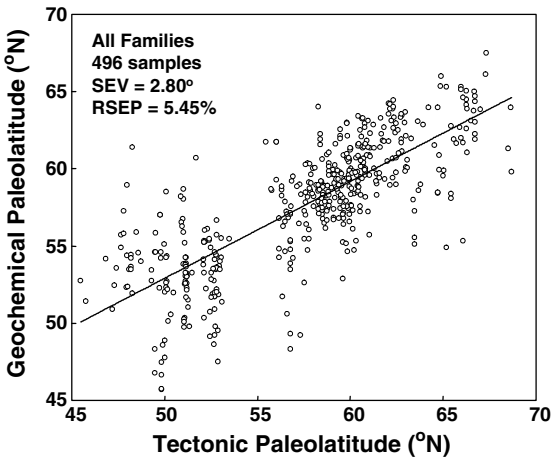

Fig. 1. Tectonic–geochemical paleolatitude model for 496 Circum-Arctic crude oil samples generated from Upper Jurassic source rock. Tectonic paleolatitude was restored from present-day latitude and longitude using AutoPointTracker© (PALEOMAP Project, C. Scotese). Geochemical paleolatitude was determined by partial least squares regression and leave-one-out cross validation of 19 source-related biomarker and isotopic ratios. SEV: standard error of cross validation, RSEP: relative standard error of prediction.

3. Results and discussion

Tectonic paleolatitude, determined by restoration from present day geographic information, shows positive correlation with geochemical paleolatitude, determined by leave-one-out cross validation PLSR of 19 biomarker and isotope ratios for 496 crude oil samples that originated from Upper Jurassic source rock in the Circum-Arctic area (Fig. 1). The TGP correlation suggests that crude oil geochemical data contain information that can be used to restore the paleolatitude of the source rock at the time of deposition.

Chemometric analysis of the data for the 496 oil samples identified seven families generated from different organofacies of Upper Jurassic distal marine shale source rock (families 231, 2322, 2323, 3211, 3212, 3213, 3214; Peters et al., 2007). We examined the correlation of tectonic and geochemical

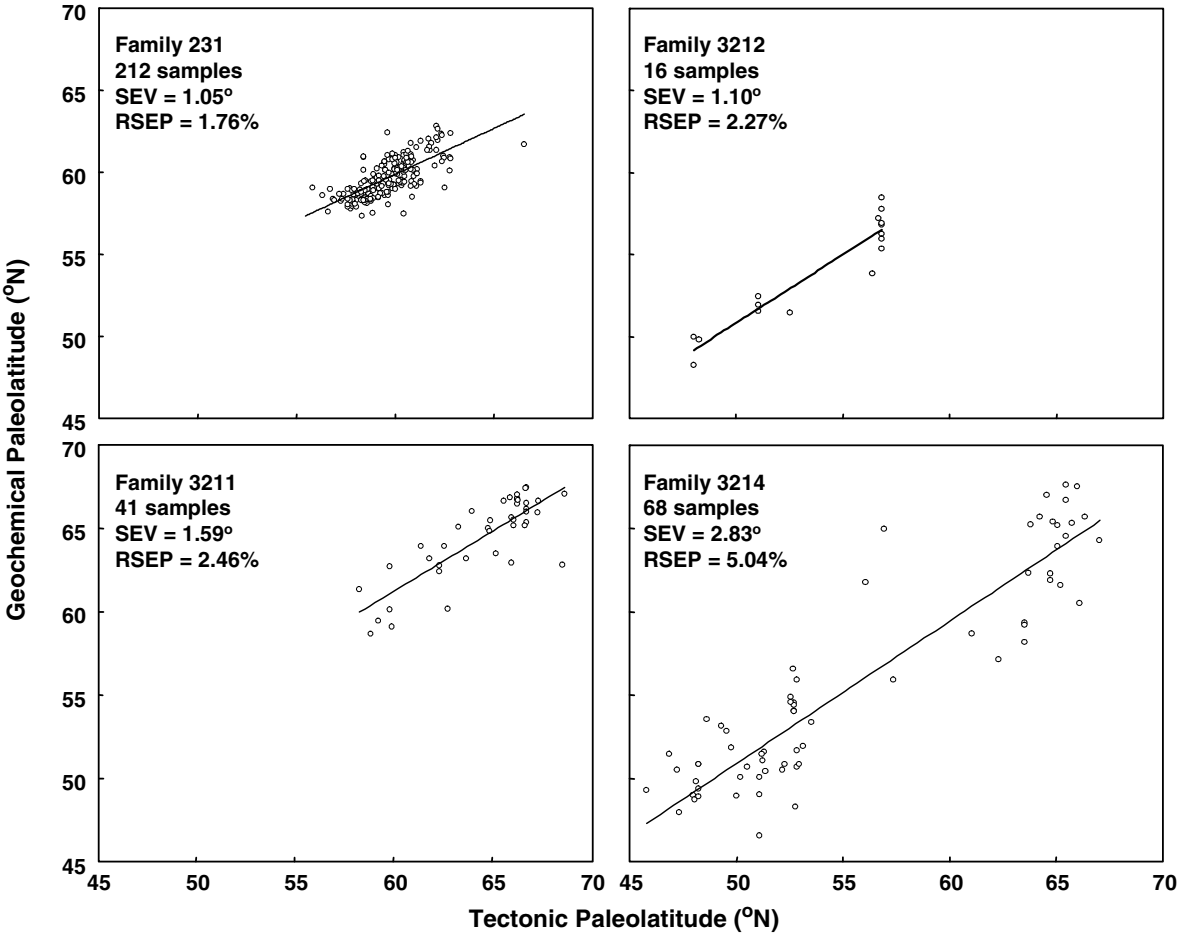

Fig. 2. Tectonic-geochemical paleolatitude models for four Circum-Arctic oil families derived from different organofacies of Upper Jurassic source rock based on oil–oil and oil–source rock correlations and geologic evidence (Peters et al., 2007). Geochemical paleolatitude was determined by partial least squares regression and leave-one-out cross validation of 19 source-related biomarker and isotopic ratios. SEV: standard error of cross validation, RSEP: relative standard error of prediction.

paleolatitude for each family separately. Fig. 2 shows leave-one-out cross validated PLSR models for four of the seven oil families and associated RSEP values. Families 231 and 3211 are from West Siberia, Family 3212 is from the North Sea and Family 3214 samples are from West Siberia, the North Sea and offshore Labrador. Family 231 consists of 212 samples generated from a highly reducing to anoxic marine organofacies of the Bazhenov Formation in the central part of the West Siberian Basin (Fig. 7 in Peters et al., 2007). These samples show a limited range of tectonic paleolatitude (7.3° with the exception of one sample), resulting in a dense distribution of data points that nonetheless show a positive relationship between tectonic and geochemical paleolatitude (RSEP 1.76%, Fig. 2). Family 3211 samples originated from a more oxic, terrigenous organofacies than Family 231 and occur between the center and periphery of the basin. These samples show a slightly greater range of tectonic paleolatitude (10.4° than Family 231 and similar correction of tectonic and geochemical paleolatitude (RSEP 2.46%). Family 3212 consists of only 16 samples from the North Sea (8.8° paleolatitude range), yet the relationship between tectonic and geochemical paleolatitude is positive (RSEP 2.27%). Finally, the 68 samples from Family 3214 show the broadest range of tectonic paleolatitude (21.3°) and RSEP is 5.04%. The 25 Family 3214 samples from West Siberia originated from a more oxic, terrigenous organofacies of the Bazhenov formation than Family 3211 and occur near the periphery of the basin (Fig. 7 in Peters et al., 2007). Families 2323 and 3213 are omitted from

Fig. 2 for brevity, but they afford results similar to those for the other Upper Jurassic oil families in the figure. Results for Family 2322 are omitted from Fig. 2 because this family contains only nine samples, thus limiting confidence in the TGP model, as confirmed by a high RSEP (6.79%).

To more rigorously validate each prediction of geochemical paleolatitude, TGP models were calibrated using 70% of the samples in each family (70%-PLSR), where the remaining 30% of the samples (validation set) were treated as unknowns (Table 1; Figs. 3 and 4). Fig. 3 shows the predictions for the Family 231 validation set from the central West Siberian Basin. Despite the limited range of tectonic paleolatitude in the validation set (6.2°), the positive relationship between tectonic and geochemical paleolatitude suggests that geochemical paleolatitude can be estimated to within ±1° (SEP 0.89°, RSEP 1.49%, Table 1). Similar conclusions apply to TGP validation for Families 3211 and 3212 (Fig. 4, left). Family 3214 validation set samples cover a broader range of tectonic paleolatitude (18.1°) than the other three families, but again shows a positive relationship between tectonic and geochemical paleolatitude (Fig. 4, upper right). The significant difference in inferred paleolatitude among the Family 3214 validation set samples from North Sea and West Siberia that occurs between samples N81UJ_DMS (Norway) and R214J_DMS (West Siberia) is probably due to error introduced by way of different ages for the Upper Jurassic source rock. The tectonic reconstruction was completed using a 150 Ma time, but Kimmeridge Clay in the North Sea and Bazhenov Formation in West Siberia are Upper Jurassic source rocks deposited ~151–156 Ma (av. ~154 Ma) and ~144–152 Ma (av. ~148 Ma), respectively.

Both stable carbon isotope and biomarker compositions influence the relationship between tectonic and geochemical paleolatitude. Fig. 4 (lower right) compares tectonic to geochemical paleolatitude for Family 3214 validation set samples, where the 70%-PLSR model was constructed using 17 biomarker ratios, but without using stable carbon isotope ratios for saturate and aromatic hydrocarbons. The fit between tectonic and geochemical paleolatitude for the data without isotopes is less favorable than with isotopes (RSEP 5.59% and 4.45%, Table 1; lower and upper right in Fig. 4, respectively), but the results show that

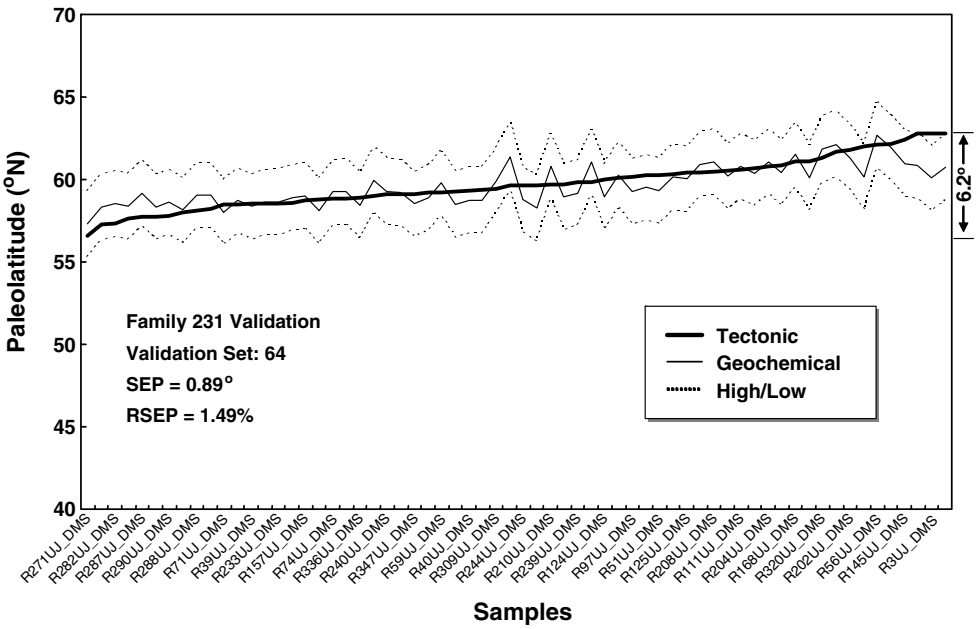

Fig. 3. Comparison of tectonic and geochemical paleolatitude (thick and thin lines, respectively) for Family 231 oil samples based on partial least squares regression of 70% of the samples (see text). Validation set consists of the remaining 30% of samples (Table 1). Dotted lines show high and low sample-specific confidence limits based on error in the model, sample leverage (its influence with respect to the model), and Student's t, where the t is 95% probability. Only every other sample is labeled on the x-axis. SEP: standard error of prediction, RSEP: relative standard error of prediction.

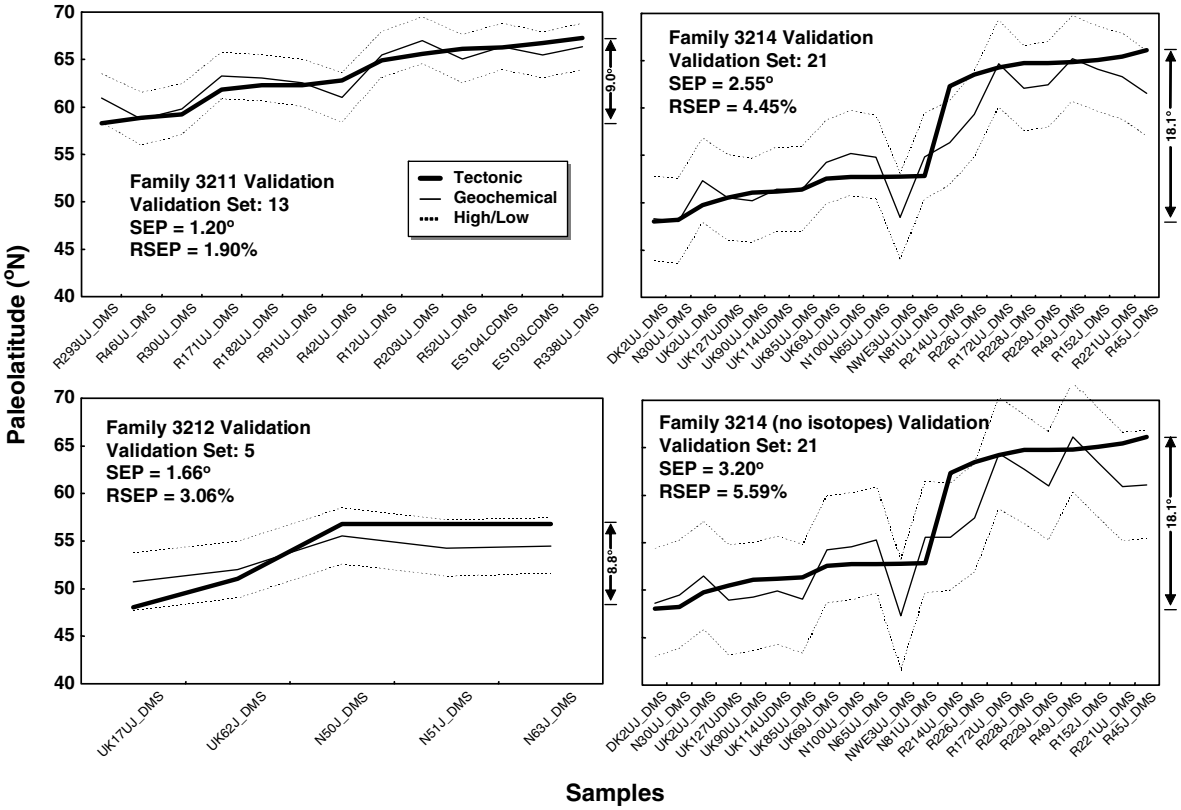

Fig. 4. Comparison of tectonic and geochemical paleolatitude (thick and thin lines, respectively) for Families 3211, 3212, 3214, and 3214 without isotopes based on partial least squares regression of 70% of the samples in each family (see text). Validation sets consist of the remaining 30% of samples in each family (Table 1). Dotted lines show high and low sample-specific confidence limits based on error in the model, the samples leverage (its influence with respect to the model), and Student's t, where the t is the 95% probability level. SEP: standard error of prediction, RSEP: relative standard error of prediction.

paleolatitude of the source rock affects both biomarkers and isotopes.

Various factors might explain the observed correlation between tectonic paleolatitude, based strictly on restoration from present day geographic location, and geochemical paleolatitude, based on biomarker and isotope ratios in crude oil. The composition of oil is determined by: (i) properties of the source rock depositional environment, including surface water temperature and its effect on the stable carbon isotope composition of organic matter, (ii) differences in species diversity at different latitudes and (iii) secondary processes, such as biodegradation and thermal maturation. Heavily biodegraded or highly mature oil samples were omitted from the study (see Methods section). Surface water temperature and dissolved CO_2 are primary controls on the stable carbon isotopic composition of marine biomass (Andrusevich et al., 2000). Dissolution of atmospheric CO_2 in water and the isotope fractionation during photosynthetic fixation of carbon increase with decreasing temperature. As CO_2 concentration decreases due to increased water temperature or phytoplankton growth, the proportion of ^{13}C compared to ^{12}C incorporated into biomass from bicarbonate during photosynthesis increases due to exchange and kinetic isotope effects. This may explain the difference in stable carbon isotope compositions of saturate and aromatic hydrocarbons for samples from Family 3214 above and below 55°N tectonic paleolatitude (Fig. 5, top). However, as discussed above (Fig. 4, right), the correlation between tectonic and geochemical paleolatitude is only partly controlled by stable isotope ratio values.

Biomarker data support the conclusion that species diversity supplements isotope exchange and kinetic isotope effects of temperature as a control on crude oil geochemical composition (Fig. 5, bottom). Family 3214 samples from lower tectonic

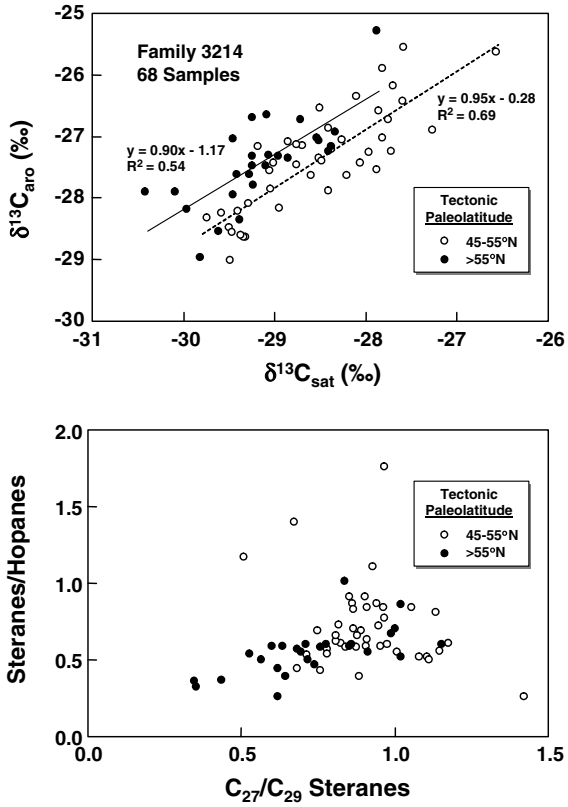

Fig. 5. Stable carbon isotope ratios for saturate vs. aromatic hydrocarbons (top) and C_{27}/C_{29} steranes vs. steranes/hopanes (bottom) for Family 3214 samples. Although overlap occurs, these parameters show significant differences for samples above and below 55°N tectonic paleolatitude. The isotopic differences (top) are highlighted by best fit lines with correlation coefficients; solid and dashed lines for greater than and less than 55°N, respectively. Samples from >55°N generally have higher steranes/hopanes and C_{27}/C_{29} steranes than samples from <55°N.

paleolatitudes (<55°N) have generally higher C_{27}/C_{29} sterane and sterane/hopane ratios, suggesting more marine character and relatively more algal than bacterial input to the source rock than those from higher tectonic paleolatitudes (>55°N). Based on analysis of many crude oil samples, Andrusevich et al. (2000) determined that organic matter in Upper Jurassic carbonate source rock from equatorial Arabian and Gulf of Mexico basins differs from that in high latitude siliciclastic source rock of about the same age in West Siberia and the North Sea. Equatorial carbonate source rock contains relatively more bacterial input than the high latitude siliciclastic source rock based on lower sterane/hopane ratios. Different C_{27}/C_{29} sterane values between the high and low latitude oil suggest differences in precursor algal input. Crude oil samples from high latitude Upper Jurassic source rock (e.g., Neuquén, North Sea and West Siberian basins) are depleted in ^{13}C compared to those from equatorial areas (e.g., Campeche-Salina, Central Arabian and Gotnia basins). These observations led Andrusevich et al. (2000) to conclude that the paleolatitude of source rock deposition affects both the isotopic and biomarker composition of subsequently generated crude oil.

Our study differs from that of Andrusevich et al. (2000) because we focussed on Circum-Arctic crude oils derived from Upper Jurassic distal marine shale (siliciclastic) source rocks. More importantly, rather than making observations about the geochemical compositions of crude oil from source rocks deposited at various paleolatitudes, we reversed the approach by developing TGP models that predict paleolatitude of the source rock from the geochemical composition of genetically related crude oil.

4. Conclusions

This work proves that crude oil contains information about both the depositional environment and the paleogeography of the source rock. It thus validates studies designed to reconstruct ancient source rock depositional environments using oil chemistry and tectonic reconstruction of source rock paleolatitude from the coordinates of the present day collection site. Tectonic reconstructions from present day locations of oil samples do not always yield source rock paleogeographic settings that match those predicted from oil chemistry. For example, consider crude oil having biomarker ratios that indicate anoxic marine source rock. If tectonic reconstruction of the paleogeography of the oil places it on an ancient landmass, then either the paleogeographic map or the interpretation of the chemistry is incorrect. In this case, interpretation of paleogeography from the oil chemistry might be used to corroborate the tectonic reconstruction. As other examples, consider (i) oil that underwent long distance subsurface migration from the source rock, or (ii) a beached tarball that originated from a tanker spill or a seep many hundreds of kilometers away. Tectonic reconstruction using the collection locality would provide spurious paleogeographic information, but the chemistry of the oil might allow useful interpretation of the character and paleogeography of the source rock.

Our TGP model results show that it is possible to predict the paleolatitude of petroleum source rock solely from the geochemical composition of the expelled crude oil. TGP requires training sets of crude oil samples that originated from each organofacies of the source rock, their present day geographic locations and careful chemometric analysis of source-related biomarker and isotope data that were quantified using the same procedures and instrumentation. For the seven Upper Jurassic oil families, standard error of prediction for geochemical paleolatitude ranges from 0.9° to 2.6° of tectonic paleolatitude, which translates to a relative standard error of prediction in the range 1.5–4.8% (Table 1). The approach offers the potential for improving our understanding of source rock depositional settings, paleoclimate and crude oil distributions. For example, the geochemical paleolatitude of a sample can help identify the original site of source rock deposition from paleogeographic maps. The same source-related biomarkers and isotopes used to determine geochemical paleolatitude can help to describe depositional conditions, such as redox potential, aridity, water salinity, mineralogy, marine vs. terrigenous organic matter input, and age (e.g., Peters et al., 2005). Given sufficient crude oil samples, the approach can be applied to other important time intervals for source rock deposition.

Acknowledgements

The research benefited from support by Stephen W. Brown (GeoMark Research Ltd.) and Brian G. Rohrback (Infometrix Inc.). We thank Leslie B. Magoon, Robert J. Rosenbauer and Kenneth J. Bird (U.S. Geological Survey) for helpful reviews of an early draft. We also acknowledge useful comments by J.A. Curiale and an anonymous reviewer. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Guest Associate Editor—**J.A. Curiale**

References

Andrusevich, V.E., Engel, M.H., Zumberge, J.E., 2000. Effects of paleolatitude on the stable carbon isotope composition of crude oils. Geology 28, 847–850.

Davies, A.M.C., Fearn, T., 2006. Back to basics: calibration statistics. Spectroscopy Europe 18, 3–31.

Faber, N.M., Schreutelkamp, F.H., Vedder, H.W., 2004. Estimation of prediction uncertainty for a multivariate calibration model. Spectroscopy Europe 10, 17–20.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185, 1–17.

Jones, R.W., 1987. Organic facies. In: Brooks, J., Welte, D. (Eds.), Advances in Petroleum Geochemistry. Academic Press, New York, pp. 1–90.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.

Klemme, H.D., Ulmishek, G.F., 1991. Effective petroleum source rocks of the world; stratigraphic distribution and controlling depositional factors. American Association of Petroleum Geologists Bulletin 75, 1809–1851.

Magoon, L.B., Dow, W.G., 1994. The petroleum system. In: Magoon, L.B., Dow, W.G. (Eds.), The Petroleum System—From Source to Trap, vol. 60. American Association of Petroleum Geologists Memoir, pp. 3–24.

Mello, M.R., Mohriak, U., Koutsoukos, A.M., Figueira, J.C.A., 1991. Brazilian and West African oils: generation, migration, accumulation and correlation. In: Proceedings of the Thirteenth World Petroleum Congress. John Wiley, New York, pp. 153–164.

Peters, K.E., Fowler, M.G., 2002. Applications of petroleum geochemistry to exploration and reservoir management. Organic Geochemistry 33, 5–36.

Peters, K.E., Das Gupta, U., Clark, M.E., McCaffrey, M.A., Lee, C.Y., 1995. Recognition of a tectonically displaced Infra-Cambrian source rock based on biomarkers in the Baghewala-1 oil, India. American Association of Petroleum Geologists Bulletin 79, 1481–1494.

Peters, K.E., Walters, C.C., Moldowan, J.M., 2005. The Biomarker Guide, second ed. Cambridge University Press, Cambridge, 1155 pp.

Peters, K.E., Ramos, L.S., Zumberge, J.E., Valin, Z.C., Scotese, C.R., Gautier, D.L., 2007. Circum-Arctic petroleum systems identified using decision-tree chemometrics. American Association of Petroleum Geologists Bulletin 91, 877–913.

Ramos, L.S., 1994. Characterization of mycobacteria species by HPLC and pattern recognition. Journal of Chromatographic Science 32, 219–227.

Zumberge, J.E., Ramos, L.S., 1996. Classification of crude oils based on genetic origin using multivariate modeling techniques (abs.). In: 13th Australian Geological Convention, Canberra, 19–25 February, 1996.